The Impact of Heteroskedasticity in Observational Studies of Causal Effects

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

December 2023

Abstract

There is a large and rapidly growing causal inference literature, yet little is known about the impact of heteroskedasticity in popular causal settings. In observational studies where the presence of heteroskedasticity can not be ruled out with certainty, its effects on treatment assignment and response generation must be studied not only because they can be of interest in their own right, but also because omitted heteroskedasticity can interact with nonlinearities in each case and impact the bias and consistency properties of estimators which can not be corrected by standard error adjustments. Our approach is Bayesian and involves specific modeling whose practical adequacy is addressed through model comparisons. We extend the methodology underlying well-known settings such as the sharp and fuzzy regression discontinuity designs, the Rubin causal (Roy-type) model, propensity score matching and inverse probability weighting. Key features of our approach include flexible modeling, the development of customized computationally efficient estimation algorithms, the ability to recover various functions of the treatment parameters, and improved efficiency of estimation. Simulation studies demonstrate the effects of omitted heteroskedasticity and gauge the adequacy of our proposed modeling and estimation methods, while their practical applicability is studied in three applications. In particular, we examine the effect of academic probation on subsequent academic performance, the influence of Medigap on healthcare expenditures, and the impact of COVID-19 vaccination on mental well-being. These applications illustrate the consequences of misspecification and provide strong evidence that the dangers of omitted heteroskedasticity should not be ignored.

`Keywords:` Bayesian estimation; 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 crucial role in observational studies of causal effects. This importance is underscored by challenges arising from non-random treatment assignment, unobserved confounders, or the inherent missingess of counterfactual outcomes at the unit level. A

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

variety of parametric, semi-parametric and nonparametric approaches have been proposed in the literature that have dealt with the effects of continuous, binary, and categorical treatments in non-experimental settings in both classical and Bayesian contexts. Regression discontinuity, Rubin causal (a.k.a. Roy-type) models, and matching estimators, among others, have been proposed and implemented in applications. Classical approaches have tended to center around estimators that are robust to potential misspecifications of the data-generating process (DGP), which is often not explicitly stated, and inference is asymptotic. The Bayesian methods adopted in this paper, consider the DGP directly and explicitly, leading to finite-sample inferences; possible misspecification is handled by allowing for flexible modeling and conducting formal model comparisons and specification searches.

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

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

Following the seminal work of Rosenbaum and Rubin (1983) and Rosenbaum (1987), substantial attention has also been directed toward the development and application of methods centered around the conditional probability of receiving treatment, known as the propensity score (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; Chesnaye et al., 2022; Duan et al., 2023). A recent review of these methodologies is offered in Rosenbaum and Rubin (2022). Propensity score matching (PSM) and inverse probability of treatment weighting (IPTW) estimators have found application across a broad range of settings. The framework is elegant and theoretically powerful; yet, in practice, results from its implementation have often been mixed. Because propensity score misspecification can compromise the efficacy of PSM and IPTW methods, formally addressing model uncertainty is a pivotal challenge that warrants careful consideration in empirical practice.

There has also been a recent surge in interest in causal analysis within the regression discontinuity design (RDD) framework. The RDD approach, introduced in Thistlethwaite and Campbell (1960) aims to address causal inference in a quasi-experimental settings where treatment assignments are based on an ancillary variable crossing a known cutoff, with a discontinuous treatment assignment rule at this cutoff point. Sharp RDD has a strict rule for treatment assignment based on the cutoff, shereas in fuzzy RDD the assignment is probabilistic. There are many different applications and extensions in the literature (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). RDD methods continue to evolve rapidly, offering new perspectives and analytical tools in the study of causal relationships.

Despite the large and rapidly expanding body of causal methodology, the ramifications of heteroskedasticity in many popular treatment models remain poorly understood. One recent exception is the work of Ferman and Pinto (2019) who show that the presence of heteroskedasticity can severely impede the performance of standard methods, even in straightforward linear specifications such as difference-in-differences, especially when confronted with small data sets. In non-linear contexts, including those mentioned earlier, the detrimental effects of heteroskedasticity in treat-

ment assignment and response generation are expected to be amplified by any non-linearity and affect not only the efficiency, but also the bias and consistency properties of traditional estimators. This challenge underscores the need for a deeper study of the effects of heteroskedasticity, the development of new methodologies, and their careful implementation in empirical practice.

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

The rest of the 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 Rubin causal model, while Section 4 focuses on PSM and IPTW estimation under heteroskedasticity. Each section presents MCMC estimation algorithms, assesses performance and impact of heteroskedasticity through simulations, and applies them in practical scenarios. Section 5 offers concluding remarks.

2 Regression Discontinuity Design

This section considers heteroskedastic variants of the sharp and fuzzy RDD framework and provides the necessary estimation and model comparison techniques. Key parts of the methodology, e.g., the sampling of heteroskedasticity parameters and the approach for estimating the marginal likelihood, will continue to play a pivotal role in subsequent sections. Nonparametric functions are employed

to safeguard against misspecification while improving efficiency by capitalizing on the entirety of available data (cf. Branson et al., 2019; Chib et al., 2023), as opposed to limiting the analysis to only a subset of observations around the cutoff point. Techniques are developed for both continuous and binary outcomes.

2.1 Sharp Regression Discontinuity Design

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

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

where we observe $y_i = (1 - T_i)\, y_{i0} + T_i y_{i1}$. Heterogeneity is allowed to depend on some collection of variables z_i that could include, but is not necessarily limited to, the variables in $\{x_i, w_i, T_i\}$ and their interactions. The model specified in equation (1) is one of structural change between the treated and untreated samples; thus, intuitively, estimation can simply be performed separately within each sub-sample. However, estimation has typically been performed under the assumption $\beta_0 = \beta_1$, which emphasizes the discontinuity in the running variable and, if confirmed by the data, makes inference more precise (the assumption will be examined in the application in Section 2.2). In the literature, the RD average treatment effect (RD ATE) is defined as

$$\begin{aligned} \tau_{SRD} &\equiv \lim_{w \downarrow w^{*+}} E\left(Y_1 | w, x_i\right) - \lim_{w \uparrow w^{*-}} E\left(Y_0 | w, x_i\right) \\ &= \lim_{w \downarrow w^{*+}} E\left(g_1(w) + x_i'\beta_1\right) - \lim_{w \uparrow w^{*-}} E\left(g_0(w) + x_i'\beta_0\right), \end{aligned} \tag{2}$$

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

$$\tau_{SRD} = \lim_{w \downarrow w^{*+}} g_1(w) - \lim_{w \uparrow w^{*-}} g_0(w). \tag{3}$$

The function $g_j(\cdot)$ plays a crucial role in this setting and is modeled nonparametrically with only local penalties for smoothness in order to mitigate the potential for undue influence of values of w far from w^* on the estimated values of $g_j(\cdot)$ close to w^* (Gelman and Imbens, 2019). Flexible functional modeling can be implemented through a variety of approaches including B-splines, regression splines, natural splines, truncated power series, or wavelets, among others (for a review,

see, e.g., Ruppert et al., 2003; Ahamada and Flachaire, 2010). Here we focus on flexible modeling through Gaussian random fields Poirier (1973); Shiller (1984); Williams (1998); Fahrmeir and Lang (2001); Koop and Poirier (2004); Koop et al. (2005); Rue and Held (2005); Chib and Jeliazkov (2006); Chan and Jeliazkov (2009a); Jeliazkov (2013); Branson et al. (2019) because it allows for smoothing at every observed value of the running variable (instead of a coarser set of knots), while retaining key desirable computational properties. Extensions of the model in (1) to an additive nonparametric mean structure for the covariates x_i can be implemented as in Koop et al. (2005), Panagiotelis and Smith (2008), or Jeliazkov (2013), whereas extensions beyond normality can be pursued by employing scale mixtures of normals (Andrews and Mallows, 1974) or Dirichlet process mixtures (Ferguson, 1973; Antoniak, 1974). Chib and Greenberg (2010) combine nonparametric mean and error distribution modeling.

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

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

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

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

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

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

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

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

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

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

$$g_j|\tau_j^2 \sim N\left(g_{j0}, \tau_j^2 K_j^{-1}\right), \quad \tau_j^2 \sim IG\left(t_{\nu 0}/2, t_{d0}/2\right), \quad \beta_j \sim N\left(b_{j0}, B_{j0}\right), \quad \gamma_j \sim N\left(\gamma_{j0}, \Gamma_{j0}\right), \quad (4)$$

which, combined with the sampling density

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

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

$$\begin{aligned} \eta_{ij} &= z_i'\gamma_j + (e_{ij}^2 - \sigma_{ij}^2)/\sigma_{ij}^2, \quad \eta_j = \left(\eta_{1j}, \ldots, \eta_{n_j}\right)', \\ V_j &= \left(\Gamma_{0j}^{-1} + \frac{1}{2}Z_j'Z_j\right)^{-1}, \quad \hat{\gamma}_j = V_j\left(\Gamma_{j0}^{-1}\gamma_{j0} + \frac{1}{2}Z_j'\eta_j\right). \end{aligned} \quad (5)$$

The treatment effects in equations (2) and (3) can be computed by averaging over the output of the MCMC sampler and, if needed, the empirical distribution of the covariates $\{x_i, w_i, z_i\}$ in the neighborhood of w^*. The homoskedastic model results in the special case when $z_i = 1$. The overall approach is also easily adaptable to binary outcomes y_i as discussed next.

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

$$y_{ij}^* = g_j(w_i) + x_i'\beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2\right), \quad \ln(\sigma_{ij}^2) = z_i'\gamma_j, \quad \text{for} \quad j \in \{0, 1\}. \quad (6)$$

Algorithm 1 (Semi-parametric Sharp RDD)

---

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

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

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

(4) Sample $[\gamma_j|y_j,g_j,\beta_j]$, $j = 0, 1$, using an MH step by drawing a proposed $\gamma_j^\dagger \sim q\left(\gamma_j|\hat{\gamma}_j,V_j\right)$, where $\hat{\gamma}_j$ and V_j are computed in (5) using the current value of γ_j. Also use $\gamma_j^\dagger$ in equation (5) to produce $\hat{\gamma}_j^\dagger$. Accept the proposed $\gamma_j^\dagger$ with probability

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

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

For identification purposes, the vector z_i, which plays a role in determining the variance, does not include a constant term (Gu et al., 2009). The variance in the homoskedastic version of the model is fixed at 1 and is not estimated.

The complete data likelihood can be expressed as

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

which, combined with the priors in (4) produces the joint posterior that is sampled in Algorithm 2. Note that Algorithms 1 and 2 are closely related, but the latter makes use of the suitably generated

latent $\{y_i^*\}$ instead of the observed $\{y_i\}$; both algorithms also trivially handle homoskedasticity. Finally, for computing RD ATE in equations (2)-(3) in the case of a binary outcome, it is helpful to recognize that $E\left(Y_j|w,x_i,z_i\right)$ is given by $\Phi\left((g_j(w)+x_i'\beta_j)/\sqrt{\exp\left(z_i'\gamma_j\right)}\right)$, which is averaged over the MCMC draws and covariates. Accounting for both covariate variability and estimation uncertainty is essential in this context because of the inherent nonlinearity of the estimand (see, e.g., Verlinda, 2006; Jeliazkov and Vossmeyer, 2018).

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

---

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

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

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

(4) Sample $\left[\gamma_j|y_j^*,g_j,\beta_j\right]$, $j=0,1$, using an MH step by drawing a proposed $\gamma_j^\dagger \sim q\left(\gamma_j|\hat{\gamma}_j,V_j\right)$, where $e_i=y_i^*-g_j(w_i)-x_i'\beta_j$ and $\hat{\gamma}_j$ and V_j are computed in (5) using the current value of γ_j and y_j^*. Also use $\gamma_j^\dagger$ in equation (5) to produce $\hat{\gamma}_j^\dagger$. Accept the proposed $\gamma_j^\dagger$ with probability

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

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

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

2.1.1 Bayesian Model Comparison and Marginal Likelihood Estimation

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

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

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

$$m(y|\mathcal{M}_s) = \frac{f(y|\theta_s, \mathcal{M}_s)\, \pi(\theta_s|\mathcal{M}_s)}{\pi(\theta_s|y, \mathcal{M}_s)}, \tag{7}$$

which holds for any θ_s in the parameter space. The terms in the numerator of equation (7) are often available directly, so the primary challenge is in estimating the posterior density $\pi(\theta_s|y, \mathcal{M}_s)$ in the denominator of equation (7). In practice, the right-hand side of equation (7) is evaluated at some appropriate point θ_s^*, typically taken to be the posterior mean or mode.

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

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

Thus, model adequacy, as indicated by the marginal likelihood, corresponds to cumulative out-of-sample predictive performance. This assessment involves evaluating the fit of y_i based on the

posterior density, utilizing data $\{y_j\}_{j<i}$ up to the ith data point. Unlike in-sample measures conditioned on the entire dataset y or split-sample comparisons sensitive to sample selection, the marginal likelihood remains unaffected by permutations in the order of data.

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

To simplify the notation in the remainder of our discussion, we suppress the model indicator $\mathcal{M}_s$ and focus on the case where the parameter vector consists of several blocks $\theta = (\theta_1', \ldots, \theta_B')'$. To handle this case, the posterior density in the denominator of (7), evaluated at the point θ^*, can be decomposed as

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

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

$$\pi(\theta_b^*|y, \{\theta_s^*\}_{(s<b)}) = E\left\{\pi\left(\theta_b^*|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\right\} \tag{8}$$

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

$$\pi(\theta_b^*|y, \{\theta_s^*\}_{(s<b)}) = \frac{E\left\{\alpha\left(\theta_b, \theta_b^*|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\, q\left(\theta_b, \theta_b^*|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\right\}}{E\left\{\alpha\left(\theta_b^*, \theta_b|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\right\}} \tag{9}$$

when the full-conditional density is non-standard and sampling requires the MH algorithm (Chib and Jeliazkov, 2001). The expectation in equation (8) is with respect to $\pi\left(\{\theta_s\}_{(s>b)}|y, \{\theta_s^*\}_{(s<b)}\right)$, whereas the expectations in the numerator and denominator of equation (9) are evaluated using $\pi\left(\{\theta_s\}_{(s\geq b)}|y, \{\theta_s^*\}_{(s<b)}\right)$ and $q\left(\theta_b^*, \theta_b|y, \{\theta_s^*\}_{(s<b)}, \{\theta_s\}_{(s>b)}\right)\pi\left(\{\theta_s\}_{(s>b)}|y, \{\theta_s^*\}_{(s<b)}\right)$, respectively. Estimation of the marginal likelihood could, therefore, become computationally intensive as it requires additional simulation of $\{\theta_s\}_{(s\geq b)}$ in reduced runs where $\{\theta_s^*\}_{(s<b)}$ are held fixed.

To deal with this problem and improve computational efficiency, we group all parameter blocks that are sampled from known densities into the set $\psi = \{\psi_1, \ldots, \psi_R\}$, whereas latent data and

parameters that are sampled from non-standard distributions are denoted by ξ. We propose estimation of the joint ordinate of the blocks in ψ based on the invariance condition of Markov chains (Ritter and Tanner, 1992; Jeliazkov and Lee, 2010) as

$$\pi\left(\psi^{*}|y\right)=E\left\{K\left(\psi,\psi^{*}|y,\xi\right)\right\}, \tag{10}$$

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

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

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

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

where $\pi(\psi^{*}|y)$ can be estimated using equation (10) with draws from the main MCMC run, and $\pi(\gamma^{*}|y,\psi^{*})$ is obtained by equation (9), which requires a single reduced run. The computation in the binary case is done analogously, but integration is also done over the latent $\{y_{ij}^{*}\}$.

2.1.2 Simulation Study

In this section, we conduct targeted simulations to assess the influence of ignored heteroskedasticity, evaluate the performance of the MCMC algorithm, and examine the effectiveness of the proposed model comparison approach. We simulate the data from

$$g_0\left(w\right)=1-\sin\left(w+1\right)+\left(w+1\right)^2,\quad g_1\left(w\right)=-1-\sin\left(w\right)+w^2,\quad w\sim U\left[-1,1\right],$$
$$\gamma_j\sim N\left(0,I\right),\quad \beta_j\sim N\left(0,I\right),\quad X\sim N\left(0,I\right),\quad Z=\left(1,w,X_1\right),$$

where X_1 is the first column of the generated covariates X. We report means, standard deviations, and 95% credible intervals of the posterior distribution for the treatment effect in each model. Additionally, we present marginal likelihood estimates to facilitate model comparisons. We also report the RD ATE estimates, standard errors and 95% confidence intervals provided by RDRobust (Calonico et al., 2017).

The marginal likelihood and the estimated treatment effect is presented in Table 1. The heteroskedastic model is supported by the data in all scenarios. As the sample size increases, both homoskedastic and heteroskedastic models yield point estimates that approach the true treatment effect. However, as the sample size increases, the evidence in favor of the heteroskedastic model grows based on the marginal likelihood estimates reported in Table 1. The Bayesian models provide more precise estimates than RDRobust, because the latter relies only on data around the cutoff (sample sizes are reported in Table 1). Sensitivity analysis confirmed that the reduction in variability is not driven by the priors. Table 17 in Appendix A presents details on these results.

A second set of simulation was conducted under the restriction $\beta_0 = \beta_1$, with all other parameters sampled as before. The results, presented in Table 2, support the aforementioned conclusions, namely that the heteroskedastic model yields more efficient estimates RDRobust and the homoskedastic specification.

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

	Model	**True ATE**	**RD ATE**	**SD**	**95% CI**	**Marg. Like.**	**Obs.**
n = 500	Homoskedastic	-2.2357	-2.3638	0.2514	(-2.8657, -1.8774)	-765.22	(230, 270)
	Heteroskedastic	-2.2357	-2.2616	0.1774	(-2.6127, -1.9173)	-617.77	(230, 270)
	RDRobust	-2.2357	-2.6305	0.6780	(-4.1775, -1.0734)		(49, 68)
n = 5000	Homoskedastic	-2.2195	-2.0420	0.1289	(-2.2936, -1.7872)	-8711.93	(2480, 2520)
	Heteroskedastic	-2.2195	-2.0996	0.1062	(-2.3055, -1.8885)	-7929.93	(2480, 2520)
	RDRobust	-2.2195	-2.1892	0.2142	(-2.7119, -1.7242)		(955, 994)
n = 50000	Homoskedastic	-2.2092	-2.1987	0.0409	(-2.2793, -2.1191)	-68715.40	(24873, 25127)
	Heteroskedastic	-2.2092	-2.1993	0.0283	(-2.2550,-2.1439)	-59044.64	(24873, 25127)
	RDRobust	-2.2092	-2.0633	0.0690	(-2.2046, -1.8831)		(7983, 8124)

SD: Standard deviation for the Bayesian methods; Standard Error for RDRobust.
CI: Credible Interval for the Bayesian methods; Confidence Interval fro RDRobust.
Obs: Number of observations used in the analysis.

Table 2: ATE with Continuous Outcome Variable ($\beta_1 = \beta_0$)

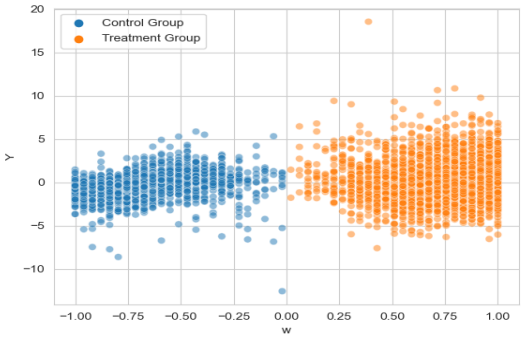
	Model	**True ATE**	**RD ATE**	**SD**	**95% CI**	**Marg. Like.**	**Obs.**
n = 500	Homoskedastic	-2.1585	-2.3144	0.2276	(-2.7632, -1.8710)	-784.62	(238, 262)
	Heteroskedastic	-2.1585	-2.3964	0.1764	(-2.7449, -2.0534)	-680.02	(238, 262)
	RDRobust	-2.1585	-2.3896	0.2250	(-2.8610, -1.8196)		(72, 80)
n = 5000	Homoskedastic	-2.1585	-2.0998	0.1131	(-2.3201, -1.8770)	-7811.57	(2529, 2471)
	Heteroskedastic	-2.1585	-2.1142	0.1097	(-2.3281, -1.8978)	-7621.34	(2529, 2471)
	RDRobust	-2.1585	-1.9219	0.1603	(-2.2203, -1.5017)		(2424, 2576)
n = 50000	Homoskedastic	-2.1585	-2.148	0.0393	(-2.2242, -2.0698)	-54592.15	(24815, 25185)
	Heteroskedastic	-2.1585	-2.1579	0.0292	(-2.2149, -2.1002)	-48880.91	(24815, 25185)
	RDRobust	-2.1585	-2.1474	0.0323	(-2.2132, -2.0644)		(7429, 7479)

To demonstrate the pitfalls of only employing data within a small band around w^*, we present a study with a sample size $n = 5000$ from the following DGP

$$g_0(w) = \sin(w) + \exp\left(-20(w+0.5)^2\right), \quad g_1(w) = 1.2 - \sin(w) - \exp\left(-20(w-0.5)^2\right),$$
$$\gamma_0 = (-2, 2, 1)', \quad \gamma_1 = (-2, 2, -1)', \quad \beta_j \sim N(0, I), \quad X \sim N(0, I), \quad Z = (1, ||w| - 1.5|, X1).$$

in a case where the cutoff point is in a low density region of w presented in Figure 1. In this scenario, the paucity of observations around the cutoff is compounded by more pronounced heteroskedasticity in that region. The generated data passed the density test (McCrary, 2008) with p-value 0.4655. The estimated RD ATE is provided in Table 3. The estimated parameter $\hat{g}_0$ and $\hat{g}_1$ can be found in Figure 2. In this context, the estimates from the homoskedastic model can be adversely affected by the outliers near the cutoff point, which ultimately resulted in significantly biased RD ATE estimates. On the other hand, the figure shows that the heteroskedastic model can estimate the true function well, owing to the fact that the data points are weighted correctly in the heteroskedastic context. In this scenario, RDRobust yielded a notably wide 95% confidence interval, primarily due to the dramatically smaller number of data points near the cutoff. We take this as a warning to account for the distribution of w as well as the behavior of the heteroskedasticity around the cutoff in determining the merits of alternative estimators.

Figure 1: Data and Running Variable Density



(a) Data

(b) Running Variable Density

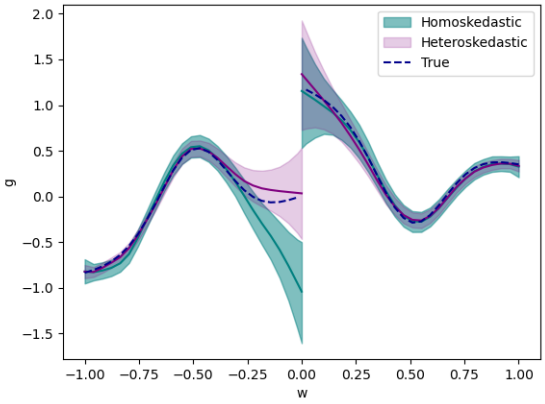
Finally, we simulate data for settings with binary outcomes using (6) with

$$g_0(w) = 1 - \sin(w+1) + (w+1)^2, \quad g_1(w) = -1 - \sin(w) + w^2, \quad w \sim U[-1, 1],$$
$$\gamma_0 = \gamma_1 = 2, \quad \beta_j \sim N(0, I), \quad X \sim N(0, I), \quad Z = (||w| - 1.5|).$$

Table 3: RD ATE Results

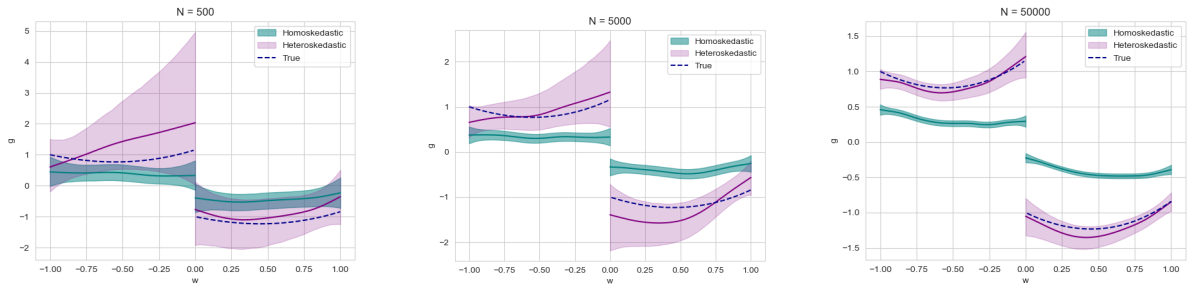
Model	True ATE	RD ATE	SD	95% CI	Marg. Like.	Obs.
Homoskedastic	1.2005	2.2085	0.4174	(1.3831, 3.0230)	-7642.76	(2489, 2511)
Heteroskedastic	1.2005	1.3185	0.3986	(0.5307, 2.0935)	-6066.56	(2489, 2511)
RDRobust	1.2005	2.0958	0.8682	(-0.1510, 4.1174)		(202, 190)

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



Estimates of the nonparametric functions $\hat{g}_0$ and $\hat{g}_1$ are depicted in Figure 3. Table 4 presents the RD ATE and marginal likelihood statistics. As the sample size increases, the heteroskedastic model provides a closer approximation, whereas both the homoskedastic model and RDRobust exhibit inconsistencies while also significantly understating the estimation variability. The impact of ignoring heteroskedasticity is amplified when the non-linear features of the model become more prominent. With larger samples, the evidence in support of the heteroskedastic specification grows stronger as demonstrated by the marginal likelihood results in Table 4.

Figure 3: Estimated g with Binary Outcome Variable



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

Table 4: RD ATE with Binary Outcome Variable

	Model	True ATE	RD ATE	SD	95% CI	Marg. Like.	Obs.
n = 500	Homoskedastic	-0.2303	-0.1931	0.1003	(-0.3877,-0.0056)	-306.18	(262, 238)
	Heteroskedastic	-0.2303	-0.2377	0.1005	(-0.4327,-0.0382)	-301.29	(262, 238)
	RDRobust	-0.2303	-0.2262	0.1318	(-0.5405,0.0801)		(109, 91)
n = 5000	Homoskedastic	-0.2579	-0.2212	0.0433	(-0.3061, -0.1363)	-2811.87	(2529, 2471)
	Heteroskedastic	-0.2579	-0.2997	0.0488	(-0.3945,-0.2037)	-2770.39	(2529, 2471)
	RDRobust	-0.2579	-0.2218	0.0480	(-0.3411,-0.1167)		(941, 932)
n = 50000	Homoskedastic	-0.2566	-0.1684	0.0171	(-0.2019,-0.1348)	-27844.76	(24815, 25185)
	Heteroskedastic	-0.2566	-0.2631	0.0195	(-0.3011,-0.2247)	-27470.87	(24815, 25185)
	RDRobust	-0.2566	-0.1756	0.0174	(-0.2194,-0.1386)		(6949, 7155)

2.2 Application: The Effect of Academic Probation on Student Performance

Academic probation is commonly used as a catalyst to motivate students and improve effort levels. Fletcher and Tokmouline (2018) and Wright (2020) employed RDD to evaluate the impact of academic probation on academic performance. We use public data from the Texas Higher Education Opportunity Project (THEOP) to study the impact of academic probation on students' academic performance. We performed analysis employing both homoskedastic and heteroskedastic models, using model comparison techniques to assess their practical relevance in this context.

The treated group consists of students who received academic probation at the end of their first semester. The outcome variables are the students' GPA in two subsequent semesters as well as their graduation status. We use the longitudinal administrative data from University of Texas, Austin, for students admitted from 1991 through 2000. Students are placed on probation (the treatment) if their cumulative GPA falls below 2.0; such students must raise their GPA above the threshold or face dismissal from the university. Covariates in this setting include the student's gender, citizenship, race, standardized SAT score, high school decile, an indicator of private high school attendance, and an indicator if the student has a major in the first semester. Summary of statistics for the data are presented in Tables 5, 6 and 7.

One key underlying assumption for the sharp RD design is that the students near the threshold can not manipulate their GPA. Following McCrary (2008), a density test of the running variable was performed, resulting in a t-statistic of -1.0027 and a corresponding p-value of 0.3160. Therefore, there is no evidence to suggest that students manipulate their GPA to avoid the treatment. Additionally, since the GPA data is rounded to the nearest tenth, it is possible that some students

with GPA of $[1.95, 2.05)$ are misclassified. To address this issue, students who have a first semester cumulative GPA of exactly 2.0 are eliminated from the sample (Fletcher and Tokmouline, 2018), while their covariates X and Z are retained for the purpose of treatment effect estimation.

Table 5: Summary of Statistics: Second Semester GPA

	Control (37980)		**Treatment (5789)**	
Variable	**Mean**	**SD**	**Mean**	**SD**
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 6: Summary of Statistics: Third Semester GPA

	Control (36738)		**Treatment (4366)**	
Variable	**Mean**	**SD**	**Mean**	**SD**
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

Our analysis centers on the effect of heteroskedasticity and sidesteps potential complications that may arise due to sample selection or endogeneity related to the decision to stay in school (see, e.g., Dong, 2019). Such complications are unlikely to be important at short time horizons, such as the second and third semester, where we see a strong impact on GPA, but could be relevant in the longer run where our results are inconclusive. In particular, the estimated functions $\hat{g}_0$ and $\hat{g}_1$ are represented in Figure 4, whereas estimates of the RD ATE and the marginal likelihoods are provided in Table 8. Notably, academic probation is practically relevant with a considerable positive effect on subsequent semester GPAs in all the Bayesian models, while the estimated impact from RDRobust is not always of the expected sign, and is generally lacks statistical significance.

Table 7: Summary of Statistics: Graduation

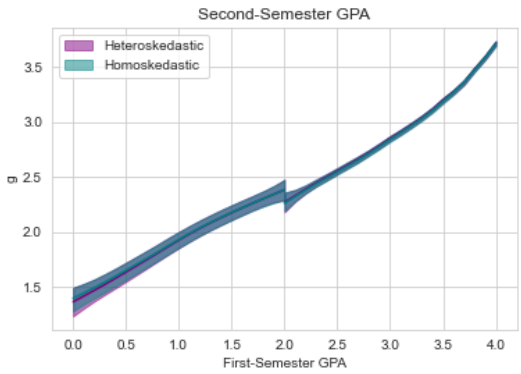
Variable	Control (38525) Mean	Control (38525) SD	Treatment (6494) Mean	Treatment (6494) SD
Female	0.517	0.500	0.387	0.487
Non US Citizen	0.003	0.052	0.001	0.037
Minority	0.032	0.175	0.066	0.248
SAT (Standardized)	0.084	0.997	-0.378	0.917
Second Decile (High School)	0.264	0.441	0.305	0.461
Third Decile (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

The impact of academic probation on graduation rates is indeterminate according to the results of all models. Additionally, the marginal likelihoods suggest that the heteroskedastic model, with the constraint that $\beta_0 = \beta_1$, is preferred in all scenarios.

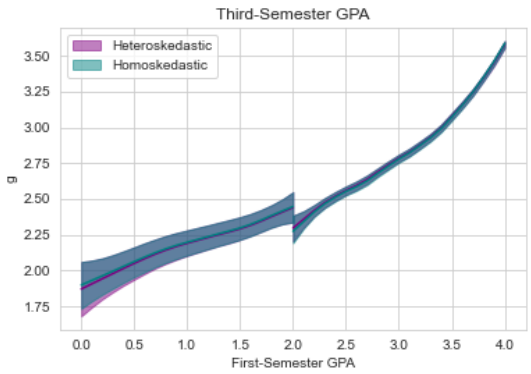
Table 8: RD ATE Results

Outcomes	Models	RD ATE	Std.	95% CI	Marg. Like.	Obs.
GPA (2nd Semester)	Homoskedastic	0.1709	0.0489	(0.0745, 0.2664)	-42714.08	(37980, 5789)
	Homoskedastic ($\beta_0 = \beta_1$)	0.1705	0.0501	(0.0715, 0.2684)	-42677.36	(37980, 5789)
	Heteroskedastic	0.1731	0.0556	(0.0641, 0.2825)	-41248.43	(37980, 5789)
	Heteroskedastic($\beta_0 = \beta_1$)	0.1709	0.0564	(0.0603, 0.2820)	-41211.59	(37980, 5789)
	RDRobust	-0.1434	0.1022	(-0.4538, 0.2644)		(1937, 3030)
GPA (3rd Semester)	Homoskedastic	0.1567	0.0570	(0.0461, 0.2700)	-44410.90	(36738, 4366)
	Homoskedastic ($\beta_0 = \beta_1$)	0.1410	0.0580	(0.0282, 0.2561)	-44383.80	(36738, 4366)
	Heteroskedastic	0.1391	0.0597	(0.0239, 0.2581)	-43617.56	(36738, 4366)
	Heteroskedastic($\beta_0 = \beta_1$)	0.1232	0.0399	(0.0063, 0.2436)	-43592.42	(36738, 4366)
	RDRobust	-0.1357	0.1070	(-0.4276, 0.3372)		(1671, 2824)
4-Year Graduation	Homoskedastic	-0.0198	0.0224	(-0.0633, 0.0246)	-27659.71	(38525, 6494)
	Homoskedastic ($\beta_0 = \beta_1$)	-0.0249	0.0223	(-0.0678, 0.0195)	-27631.28	(38525, 6494)
	Heteroskedastic	-0.0286	0.0218	(-0.0706, 0.0147)	-27642.73	(38525, 6494)
	Heteroskedastic($\beta_0 = \beta_1$)	-0.0301	0.0214	(-0.0717, 0.0125)	-27619.22	(38525, 6494)
	RDRobust	-0.0343	0.0535	(-0.3026, 0.0796)		(2022, 3138)
Graduation	Homoskedastic	0.0213	0.0238	(-0.0254, 0.0679)	-21773.71	(38525, 6494)
	Homoskedastic ($\beta_0 = \beta_1$)	0.0129	0.0239	(-0.0338, 0.0601)	-21752.81	(38525, 6494)
	Heteroskedastic	0.0070	0.0233	(-0.0386, 0.0525)	-21754.73	(38525, 6494)
	Heteroskedastic($\beta_0 = \beta_1$)	0.0114	0.0237	(-0.0352, 0.0578)	-21749.15	(38525, 6494)
	RDRobust	-0.0074	0.0627	(-0.3038, 0.1396)		(2022, 3138)

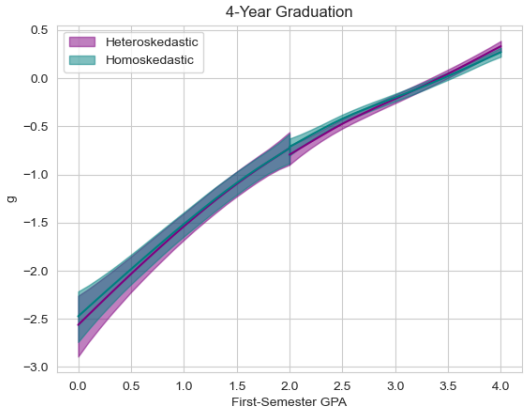
Figure 4: Estimated Nonparametric Parameters



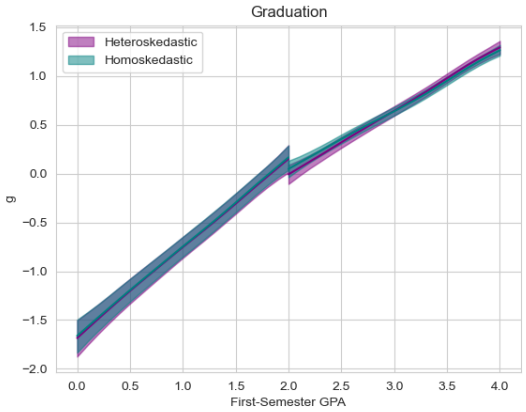
(a) Second-Semester GPA



(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 RD model and outline the corresponding estimation algorithm. To help illustrate this model, we employ simulations. 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) who never take the treatment, and always-takers (denoted as a) who always take the treatment. The treatment status for compliers is $T = \mathbb{1}\{w \geq w^* | s = c\}$. The sample data in the fuzzy RD case is summarized by Table 9.

Table 9: Fuzzy RD Data

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

The model can be specified as

$$s = c : T = \mathbb{1}\{w \geq w^*\}, \quad y_{ij} = g_j(w_i) + x_i'\beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2\right), \quad \ln\left(\sigma_{ij}^2\right) = Z_{ij}'\gamma_j$$

$$s = a : T = 1, \quad y_i = g_a(w_i) + x_i'\beta_a + \varepsilon_{ia}, \quad \varepsilon_{ia} \sim N\left(0, \sigma_{ia}^2\right), \quad \ln\left(\sigma_{ia}^2\right) = Z_{ia}'\gamma_a$$

$$s = n : T = 0, \quad y_i = g_n(w_i) + x_i'\beta_n + \varepsilon_{in}, \quad \varepsilon_{in} \sim N\left(0, \sigma_{in}^2\right), \quad \ln\left(\sigma_{in}^2\right) = Z_{in}'\gamma_n$$

$$P(s = k) = q_k > 0, \quad \forall k \in \{c, a, n\}, \quad q_c + q_n + q_a = 1.$$

The RD ATE is defined as

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

The likelihood function is expressed as

$$\begin{aligned}L = &\prod_{i \in I_{00}} \left(q_c f_N\left(y_i | g_0(w_i) + x_i'\beta_0, \sigma_{i0}^2\right) + q_n f_N\left(y_i | g_n(w_i) + x_i'\beta_n, \sigma_{in}^2\right)\right) \\ &\prod_{i \in I_{10}} q_n f_N\left(y_i | g_n(w_i) + x_i'\beta_n, \sigma_{in}^2\right) \prod_{i \in I_{01}} q_a f_N\left(y_i | g_a(w_i) + x_i'\beta_a, \sigma_{ia}^2\right) \\ &\prod_{i \in I_{11}} \left(q_c \phi\left(y_i | g_1(w_i) + x_i'\beta_1, \sigma_{i1}^2\right) + q_a f_N\left(y_i | g_a(w_i) + x_i'\beta_a, \sigma_{ia}^2\right)\right)\end{aligned}$$

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

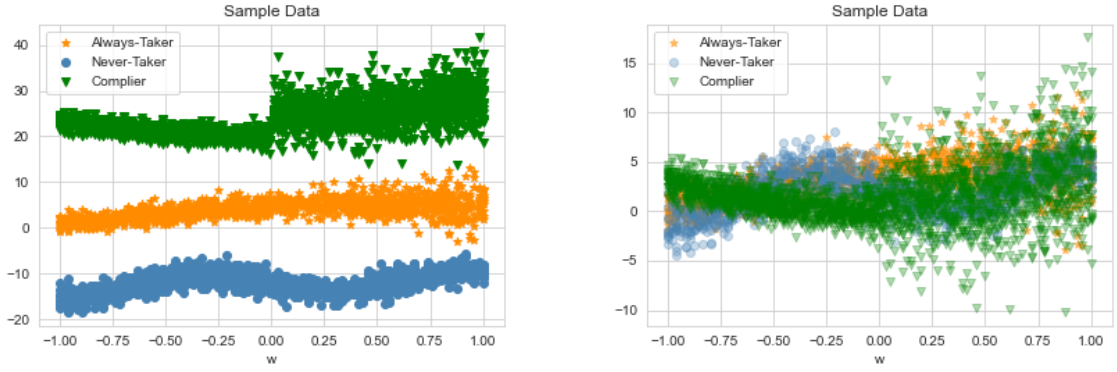
We specify the prior distribution of $q = (q_a, q_n, q_c)$ as $q \sim Dir(n_{a0}, n_{n0}, n_{c0})$. All the other parameters follows the same prior distribution as discussed in section 2.1. The vectors g_a and g_n are of dimension $m = m_0 + m_1$, where m_0 is the dimension of vector v_0, and m_1 is the dimension of vector v_1. The unique ordered values of the running variable w_0 and w_1, denoted by v_0 and v_1. The function evaluations g_0 and g_1 are vectors of dimension m_0 and m_1, respectively. In the estimation process, the nonparametric functions in each group are updated using both the prior information and observations that were categorized into this specific group in each iteration. The

posterior distribution for the type variables s_i, $i = 1, \ldots, n$, is specified as

$$\begin{aligned} Pr\left(s_i = c|y_i, g_j, \beta_j, \tau_j^2, \gamma_j\right) &\propto q_c f_N\left(y_i|g_i(w_i) + x_i'\beta_j, \sigma_{ij}^2\right), \\ Pr\left(s_i = n|y_i, g_n, \beta_n, \tau_n^2, \gamma_n\right) &\propto q_n f_N\left(y_i|g_n(w_i) + x_i'\beta_n, \sigma_{in}^2\right), \\ Pr\left(s_i = a|y_i, g_a, \beta_a, \tau_a^2, \gamma_a\right) &\propto q_a f_N\left(y_i|g_a(w_i) + x_i'\beta_a, \sigma_{ia}^2\right). \end{aligned} \tag{11}$$

The joint posterior distribution can be sampled as in Algorithm 3. We conducted a simulation study to assess the algorithm's performance and evaluate the influence of heteroskedasticity within the framework of the fuzzy RD model. 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 nonparametric functions estimated for each group are presented in Figures 6 and 7. In this context, it is important to highlight the possibility of misclassification and label switching, which can occur when the clusters are not well-separated (Celeux, 1998). Hence, we advise practitioners to carefully examine their results prior to drawing definitive conclusions in applications where clusters are not well-separated, as there is no current consensus solution to the problems present in this context.

Figure 5: Simulated Data



(a) Well-separated Sample

(b) Poorly-separated Sample

3 Rubin causal model (Roy-type model)

In this section, we introduce a potential outcome framework (Roy, 1951; Rubin, 1974, 1977, 1978, 2004, 2005) with self-selection for estimating the treatment effect, following the approach outlined in Chib (2007). We assume that there are two potential outcome variables y_0 and y_1 for the treated

Algorithm 3 (Semi-parametric Fuzzy RDD)

---

(1) Sample the type variables $\{s_i\}$ from the posterior distribution in Equation (11).

(2) Sample $q = (q_a, q_n, q_c) \sim Dir\left(n_{a0} + n_a, n_{n0} + n_n, n_{c0} + n_c\right)$, 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 $\left[g_j|y_j, \beta_j, \tau_j^2, \gamma_j\right] \sim N\left(\hat{g}_j, \hat{G}_j\right)$, where $\hat{G}_j = \left(\frac{K_j}{\tau_j^2} + Q_j'\Omega_j^{-1}Q_j\right)^{-1}$ and $\hat{g}_j = \hat{G}_j\left(\frac{1}{\tau_j^2}K_j g_{j0} + Q_j'\Omega_j^{-1}\left(y_j - X_j\beta_j\right)\right)$. We repeat this step for all the compliers and never-takers.

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

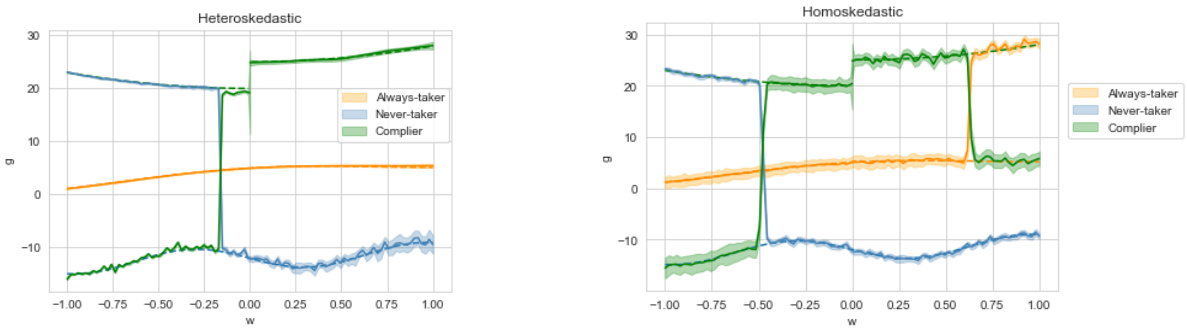
(5) Sample $\left[\tau_j^2|g_j\right] \sim IG\left(\frac{t_{\nu_{j0}} + m_j}{2}, \frac{t_{d_{j0}} + (g_j - g_{j0})'K_j(g_j - g_{j0})}{2}\right)$. Repeat this step to sample $\left[\tau_a^2|g_a\right]$ and $\left[\tau_n^2|g_n\right]$.

(6) Sample $[\gamma_j|y_j, g_j, \beta_j]$ using an MH step by drawing a proposal value $\gamma_j^\dagger \sim q\left(\gamma_j|\hat{\gamma}_j, V_j\right)$, where $\hat{\gamma}_j$ and V_j are computed as in (5) using the current value of γ_j. Also use $\gamma_j^\dagger$ in equation (5) to produce $\hat{\gamma}_j^\dagger$ and accept the proposed $\gamma_j^\dagger$ with probability

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

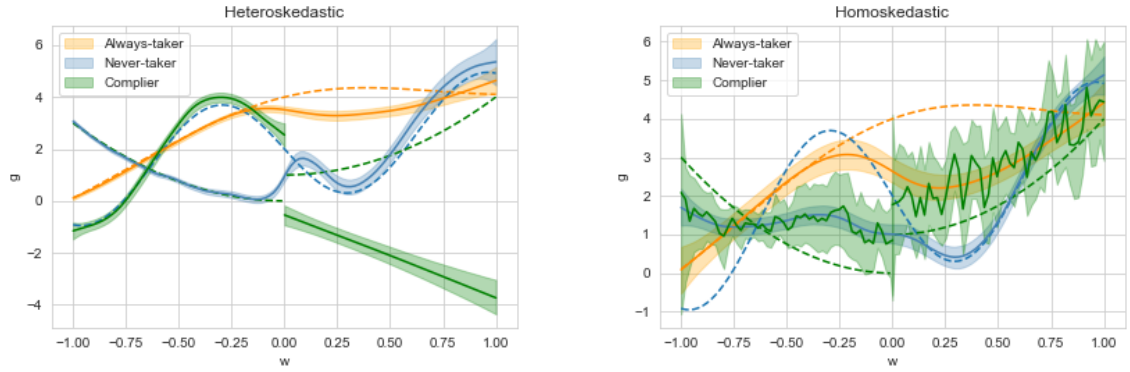
otherwise the current value γ_j is repeated in the next MCMC iteration. We repeat this step for always-takers and never-takers to sample $[\gamma_a|y_a, g_a, \beta_a]$ and $[\gamma_n|y_n, g_n, \beta_n]$.

Figure 6: Well-separated Data



(a) Heteroskedastic (b) Homoskedastic

Figure 7: Poorly-separated Data



(a) Heteroskedastic (b) Homoskedastic

and untreated states. The binary treatment status T is determined by a latent variable T^* and $T = \mathbb{1}\{T^* \geq 0\}$. We also assume that there is heteroskedasticity in both the treatment assignment and the outcome.

The model can be represented as

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

where

$$y_i = \begin{pmatrix} T_i^* \\ y_{0i} \\ y_{1i} \end{pmatrix}, \quad X_i = \begin{pmatrix} x'_{Ti} & 0 & 0 \\ 0 & x'_{0i} & 0 \\ 0 & 0 & x'_{1i} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta'_T \\ \beta'_0 \\ \beta'_1 \end{pmatrix} \quad \text{, and } \varepsilon = \begin{pmatrix} \varepsilon_{Ti} \\ \varepsilon_{0i} \\ \varepsilon_{1i} \end{pmatrix}.$$

Let N_j 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

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

Due to the missing counterfactuals, ω_{01i} is not identified. To simplify the notation in the subsequent discussion, we introduce the following matrices

$$\Omega_{0i} = \begin{pmatrix} \omega_{TTi} & \omega_{T0i} \\ \omega_{0Ti} & \omega_{00i} \end{pmatrix}, \quad \Omega_1 = \begin{pmatrix} \omega_{TTi} & \omega_{T1i} \\ \omega_{1Ti} & \omega_{11i} \end{pmatrix}, \quad J_0 = \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & I \end{pmatrix}, \quad J_1 = \begin{pmatrix} 0 & I & 0 \\ 0 & 0 & I \end{pmatrix},$$
$$\tilde{X}_{i0} = \begin{pmatrix} x'_{Ti} & 0 \\ 0 & x'_{0i} \end{pmatrix}, \quad \tilde{X}_{i1} = \begin{pmatrix} x'_{Ti} & 0 \\ 0 & x'_{1i} \end{pmatrix}, \quad \tilde{y}_{i0} = \begin{pmatrix} T_i^* \\ y_{0i} \end{pmatrix}, \quad \tilde{y}_{i1} = \begin{pmatrix} T_i^* \\ y_{1i} \end{pmatrix}.$$

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

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

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

$$\Omega_0 = \begin{pmatrix} 1 & \omega_{T0} \\ \omega_{0T} & \omega_{00} \end{pmatrix}, \qquad \Omega_1 = \begin{pmatrix} 1 & \omega_{T1} \\ \omega_{1T} & \omega_{11} \end{pmatrix},$$
$$\Omega_{22\cdot 1} = \omega_{11} - \omega_{1T}\omega_{T1}, \quad \Omega_{22\cdot 0} = \omega_{00} - \omega_{0T}\omega_{T0}.$$

We assume that $\Omega_{22\cdot j} \sim IG\left(\frac{r_j}{2}, \frac{R_j}{2}\right)$ for $j \in \{0, 1\}$.[1] Under this assumption, the conditional distribution of $\omega_{jT} | \Omega_{jj\cdot 2} \sim N\left(q_j, \Omega_{22\cdot j}\right)$. Furthermore, we specify the prior distribution of β as β is $\beta \sim N\left(b_0, B_0\right)$. The estimation algorithm is summarized in Algorithm 4.

To expand the model to the case of multivariate heteroskedasticity, we decompose the covariance matrices (Chan and Jeliazkov, 2009b) as

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

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

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

[1] If $\Omega_{22\cdot j}$ were a matrix, the generalization would be $\Omega_{22\cdot j} \sim IW\left(r_j, R_j\right)$.

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

---

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

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

(3) For $i \in N_j$,

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

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

(4) For $i \in N_j$,

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

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

The model can be rewritten as

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

or

$$T_i^* = x_{Ti}'\beta_T + \psi_{Ti}, \quad y_{ji} = x_{ji}'\beta_j + a_{jT}\psi_{Ti} + \psi_{ji}, \quad \psi_{ji} \sim N(0, \lambda_{ji}), \quad \psi_{Ti} \sim N\left(0, \lambda_{Ti}\right),$$
$$\lambda_{ji} = \exp\left(Z_{ji}'\gamma_j\right), \quad \lambda_{Tj} = \exp\left(Z_{Ti}'\gamma_T\right).$$

The prior distributions are specified as

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

and the estimation algorithm is detailed in Algorithm 5. In our model, the average treatment effect (ATE) and the average treatment effect on the treated (ATT) are defined as

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

and they can be estimated using the MCMC output.

Algorithm 5 (Rubin Causal (Roy-Type) Model with Heteroskedasticity)

---

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

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

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

(4) Sample $[\gamma_T|y_j, g_j, \beta_j]$ using an MH step by drawing a proposal value $\gamma_T^\dagger \sim q\left(\gamma_T|\hat{\gamma}_T, V_T\right)$, where $e_i = T_i^* - x_{Ti}'\beta_T$ and $\hat{\gamma}_T$ and V_T are defined similarly to (5) using the current value of γ_j and T_i^*. Also use $\gamma_T^\dagger$ in equation (5) to produce $\hat{\gamma}_j^\dagger$ and accept the proposed $\gamma_T^\dagger$ with probability

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

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

(5) Let $e_{ji} = \left(y_{ji} - x_{ji}'\beta_j - a_{jT}\psi_{Ti}\right)$, $\eta_{ji} = z_{ji}'\gamma_j + \frac{e_{ji}^2 - \omega_{jji}}{\omega_{jji}}$, and $\eta_j = \left(\eta_{j1}, \ldots, \eta_{jn_j}\right)'$ and sample $[\gamma_j|a_{jT}, \beta_j]$ similarly to Step (4).

In the homoskedastic case, the posterior ordinate can be estimated similarly to Section 2.1.1 using the CRT method. To estimate the marginal likelihood in the heteroskedastic model, let θ denote the parameters (β, a_{1T}, a_{0T}). The posterior ordinate of $\hat{\pi}(\theta^*|y, T^*)$ can also be estimated using the CRT method as discussed in Section 2.1.1. The posterior ordinate of $\hat{\pi}\left(\gamma_T^*|y, \theta^*, T^*\right), \hat{\pi}\left(\gamma_1^*|y, \theta^*, T^*\right)$, and $\hat{\pi}\left(\gamma_0^*|y, \theta^*, T^*\right)$ can be sampled similarly to Section 2.1.1.

3.1 Simulation Study

In this section, we performed a simulation study to gauge MCMC efficiency, evaluate the consequences of neglected heteroskedasticity, and validate the model comparison technique. We report mean, standard deviations, and 95% credible intervals of the posterior distribution for the treatment effects in each model.

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

$$X_0 \sim (1, N(0,1)), \quad X_1 = X_0, \quad X_d = (X_0, N(0,1)), \quad \beta \sim N(0, I), \quad \gamma_0 \sim N(0, I),$$
$$\gamma_1 \sim N(0, I), \quad \gamma_d \sim N(0, I), \quad a_{0T} = -0.2, \quad a_{1T} = 0.2.$$

The estimated ATE and ATT are summarized in Table 10. In all cases, the heteroskedastic model emerges as the recommended choice based on the marginal likelihood results. These findings highlight that ignoring heteroskedasticity when it is present, leads to inconsistent and biased estimates for both ATE and ATT.

Table 10: Treatment Effects Estimation (Simulation)

	Model		TRUE	Estimated	SD	95% CI	Marg. Like.
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)	

3.2 Application: The Effect of Medigap on Healthcare Expenditure

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

compared to the basic Medicare policy, and individuals often choose them in anticipation of reducing out-of-pocket healthcare costs.

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

We incorporate self-perceived health status variables, the number of chronic conditions, location, and various demographic variables as covariates that influence healthcare expenditures. We assume that family income only affects the purchase of the private insurance, and does not alter health care utilization directly. Variable definitions and summary statistics are presented in Table 11. In the regression, age is standardized, and due to excessive right skew, expenditure (in thousands of dollars) is stabilized using the square root transformation (Amaratunga and Cabrera, 2001). Additionally, we consider models where the variance of treatment assignment depends on family income, and the variance of healthcare expenditure depends on age and the number of chronic conditions. In this application, heteroskedasticity remains our primary focus, although modeling could be generalized to explicitly model the choice of specific Medigap plans based on their anticipated healthcare expenditures and accommodate potential endogeneity of the Medicaid variable.

The estimated ATE and ATT are presented in Table 12. 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. Parameter estimates are reported in Appendix B.

4 Propensity Score

Let $p(x) \equiv \Pr(T = 1|x)$ denote the propensity score, which represents the conditional probability of assignment to treatment given the covariates x (Rosenbaum and Rubin, 1983). There are two popular methods that utilize propensity score to mitigate the selection bias: propensity score matching and inverse probability of treatment weighting (IPTW). The key ideas behind these models are captured in Figure 8. Specifically, Figure 8a depicts the key assumption of selection on observables,

Table 11: Variable Definition and Summary Statistics

Variables	Description	2020 5019		2018, 2019 10226	
		Mean	SD	Mean	SD
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

Table 12: Treatment Effects Estimation

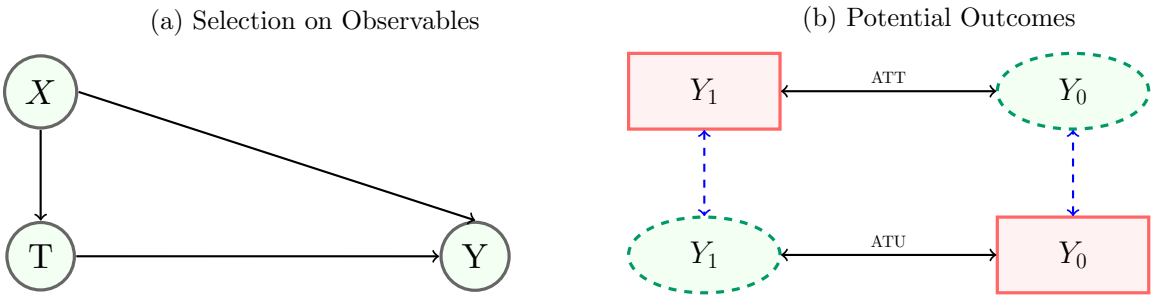
Year	**Model**		**Estimate**	**SD**	**95% CI**	**Marg. Like.**
2018, 2019	Homoskedastic	ATE	0.0162	0.0014	(0.0133, 0.0188)	22243.61
		ATT	0.0170	0.0013	(0.0144, 0.0195)	
	Heteroskedastic	ATE	-0.0058	0.0009	(-0.0076, -0.0040)	25365.03
		ATT	-0.0016	0.0009	(-0.0034, 0.0001)	
2020	Homoskedastic	ATE	0.0075	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)	

whereas Figure 8b demonstrates that the fundamental problem of estimating treatment effects is caused by the missing counterfactuals. In practice, we have the observed treated and untreated outcomes denoted by the rectangles in Figure 8b, whereas the dashed ovals are the unobserved counterfactuals. The idea behind matching observations on the basis of the propensity score is to generate sub-samples from the observed treated and untreated groups that are comparable to one another as a means of uncovering the unobserved counterfactuals and estimating the desired treatment effect.

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

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

Figure 8: Illustration



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

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

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

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

To study this issue, we employ a heteroskedastic model for the propensity score. Owing to the

nonlinearity of the setting, Jensen's inequality implies that erroneously omitting heteroskedasticity will impact the bias and consistency properties of estimators and can not be dealt with by simply adjusting the standard errors. For $i = 1, \ldots, n$, the heteroskedastic probit model is specified as

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

In the homoskedastic case, for identification purposes we impose the constraint that the variance of ν_i equals 1. In the heteroskedastic model, we assume that $var(\nu_i) = \exp\left(z_i'\gamma\right)$ and for identification z_i does not include a constant term. We specify the prior distributions $\beta \sim N\left(b_0, B_0\right)$ and $\gamma \sim N\left(\gamma_0, \Gamma_0\right)$. Algorithms 6 and 7 provide details on the propensity score estimation following the data augmentation approach provided by Albert and Chib (1993).

Algorithm 6 (Bayesian Propensity Score Estimation with Homoskedasticity)

---

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

(2) Sample $T_i^* \sim TN\left(\mu_i, 1\right)$, 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.

An estimate of the marginal likelihood for the homoskedastic model and heteroskedastic models is easily obtained as a simplification as the approach in Section 2.1.1.

4.1 Simulation

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

4.1.1 Simulation: Propensity Score Matching

Empirical researchers typically proceed by incorporating higher-order and interaction terms to improve the balance of the matched samples if it failed in the beginning (Dehejia and Wahba, 1999; Caliendo and Kopeinig, 2008). In this section, we illustrate that neglected heteroskedasticity can lead to the emergence of imbalanced samples. The specification of a model with heteroskedasticity

Algorithm 7 (Bayesian Propensity Score Estimation with Heteroskedasticity)

---

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

(2) Sample $T_i^* \sim TN\left(\mu_i, \exp(z_i'\gamma)\right)$, 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 an MH step by drawing a proposal value $\gamma^\dagger \sim q\left(\gamma|\hat{\gamma}_T, V_T\right)$, where $e_i = T_i^* - x_i'\beta$ and $\hat{\gamma}$ and V are defined similarly as in (5) using the current value of γ_j and T_i^*. Also use $\gamma_T^\dagger$ in equation (5) to produce $\hat{\gamma}_j^\dagger$ and accept the proposed $\gamma_T^\dagger$ with probability

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

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

---

is one step in addressing misspecification in addition to other possible approaches that can be taken, such as considering the problem of variable selection or misspecification of the mean function.

The simulation study is based on the data in Dehejia and Wahba (1999), which comes from the National Supported Work Demonstration (NSW) and the panel study of income dynamics (PSID). The treatment T is the NSW participation. We believe that the variables age, 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 observations in the control group. We assume that DGP for the treatment assignment is

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

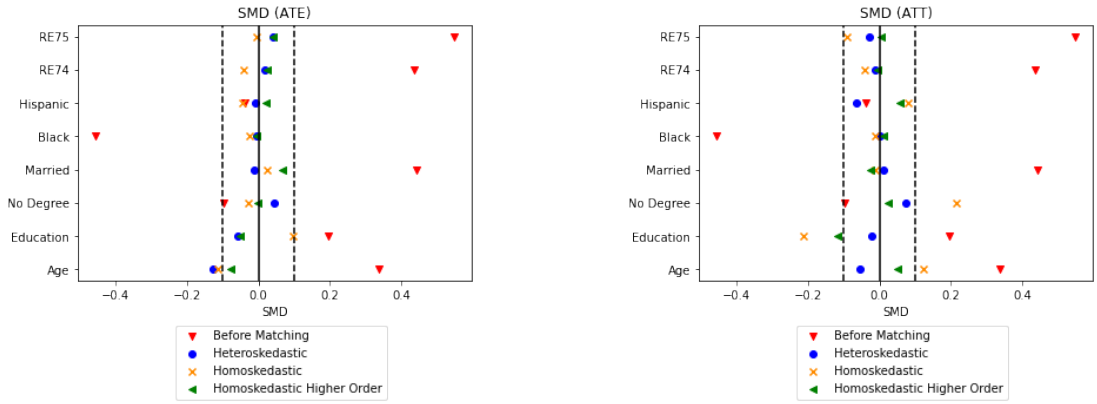
We use the standardized mean difference (SMD) as a balance measure. (Rosenbaum and Rubin (1985) and Thoemmes (2012)). A SMD exceeding 0.1 can be considered as a sign of imbalance. (Zhang et al. (2019)). SMD is calculated as

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

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, respectively. We employ nearest neighbor matching with replacement with a radius of 0.2 times the standard deviation of the estimated propensity score (Austin, 2011; Chaudhuri and Howley, 2022).

Three models are estimated: the correctly specified heteroskedastic model, the homoskedastic model with all covariates, and the extended homoskedastic model incorporating all covariates, age squared (age^2), and interactions between age and other covariates. Figure 9 shows the SMD before and after matching. Before matching, the covariates are imbalanced. The heteroskedastic model effectively enhances balance within both the ATE and ATT samples. In this example, the homoskedastic model falls short of achieving balance in the ATT estimation sample. However, by including higher order and interaction terms, the balance is improved.

Figure 9: SMD



(a) SMD (ATE)

(b) SMD (ATT)

4.1.2 Simulation: Inverse Probability of Treatment Weighting

In this section, we illustrate that neglected heteroskedasticity can lead to biased and inconsistent treatment effect estimator. We assume that the DGP for the treatment assignment is

$$x_i \sim \left(1, TN_{(1,+\infty)}\left(1,1\right), TN_{(1,+\infty)}\left(1,1\right)\right)', \quad \beta = (0.9, 1.2, -1.2)', \quad Z = (X_2), \quad \gamma = 0.6$$

$$T_i = \mathbb{1}\left\{x_i'\beta + \nu_i \geq 0\right\}, \quad \nu_i \sim N\left(0, \exp\left(z_i'\gamma\right)\right)$$

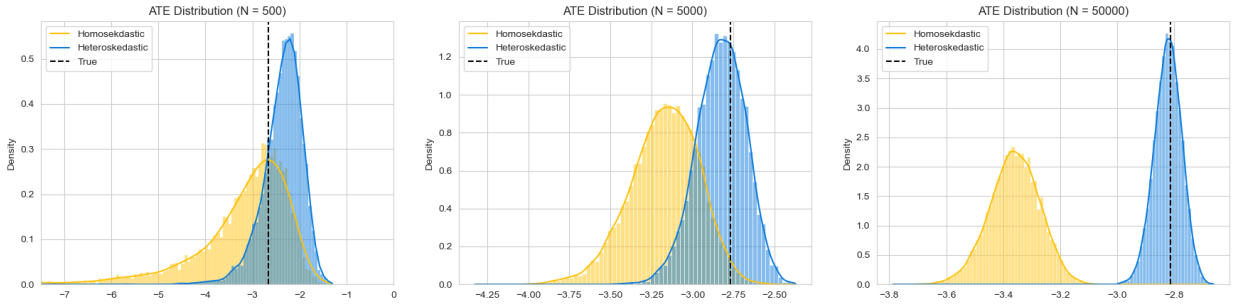
Table 13 summarizes the estimated ATE, and Figure 10 depicts the histogram of the estimated ATE. Marginal likelihood results consistently favor the heteroskedastic model in all scenarios. These findings demonstrated that the ignored heteroskedasticity can lead to biased and inconsistent ATE

estimates. Notably, the heteroskedastic model exhibits accurate ATE representation, in contrast to the homoskedastic model's failure in capturing the true ATE, especially in larger samples.

Table 13: Treatment Effects Estimation (Simulation)

	Model	True ATE	Estimate	SD	95% CI	Marg. Like.
n = 500	Heteroskedastic	-2.6609	-2.3635	0.4591	(-3.3470, -1.7043)	-247.68
	Homoskedastic	-2.6609	-3.2416	1.2067	(-6.2450, -1.9550)	-252.71
n = 5000	Heteroskedastic	-2.7663	-2.8267	0.1489	(-3.1367, -2.5564)	-2421.02
	Homoskedastic	-2.7663	-3.1711	0.2090	(-3.6145, -2.8002)	-2434.46
n = 50000	Heteroskedastic	-2.8143	-2.8222	0.0476	(-2.9183, -2.7315)	-23005.06
	Homoskedastic	-2.8143	-3.3607	0.0878	(-3.5398, -3.1963)	-23116.63

Figure 10: Treatment Distribution



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

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

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

the level of mental well-being in the evaluated individuals. Summary statistics of key variables are presented in Table 14, and analysis by propensity score matching and IPTW methods are presented in Sections 4.2.1 and 4.2.2.

Table 14: Summary of Statistics

	Control Group (12423)		Treatment Group (9562)	
Variables	**Mean**	**SD**	**Mean**	**SD**
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

4.2.1 Application: Propensity Score Matching

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

The estimated impact of COVID-19 vaccination on mental well-being is presented in Table 15. The marginal likelihood results suggest that the heteroskedastic model with $var(\nu_i) = \exp(z_i'\gamma)$ fits the data best. These results suggest that COVID-19 vaccination is expected to improve mental health. Figure 11 and 12 show the SMD before and after matching. The figures show 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 better than the alternatives.

Table 15: Impact of COVID-19 Vaccination on Mental Health (PSM)

	ATE			ATT			
Model	**Mean**	**SD**	**95% CI**	**Mean**	**SD**	**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 11: SMD (ATE Sample)

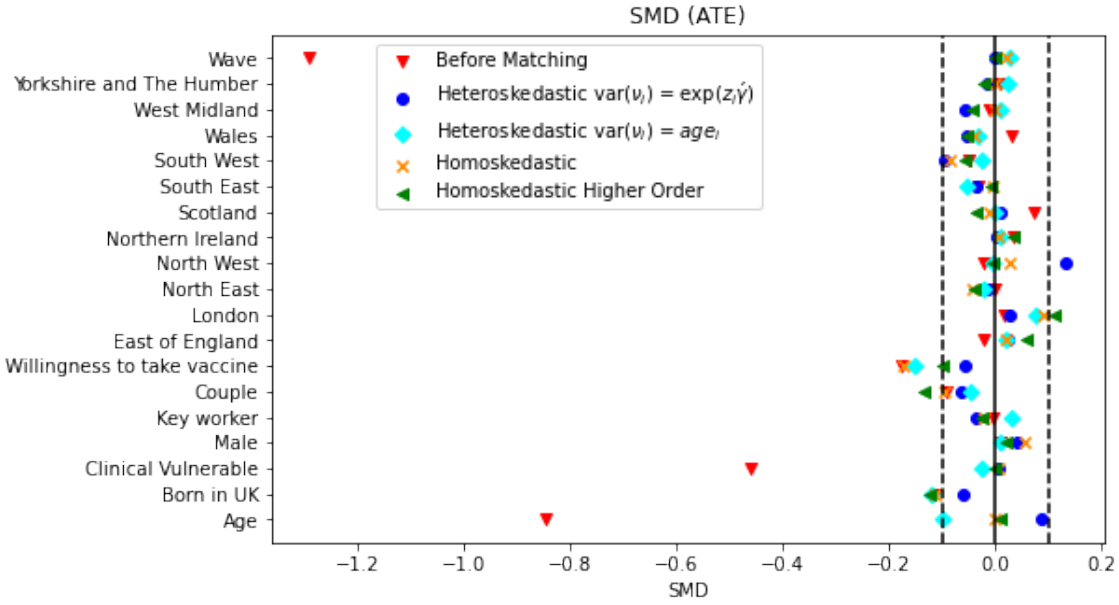
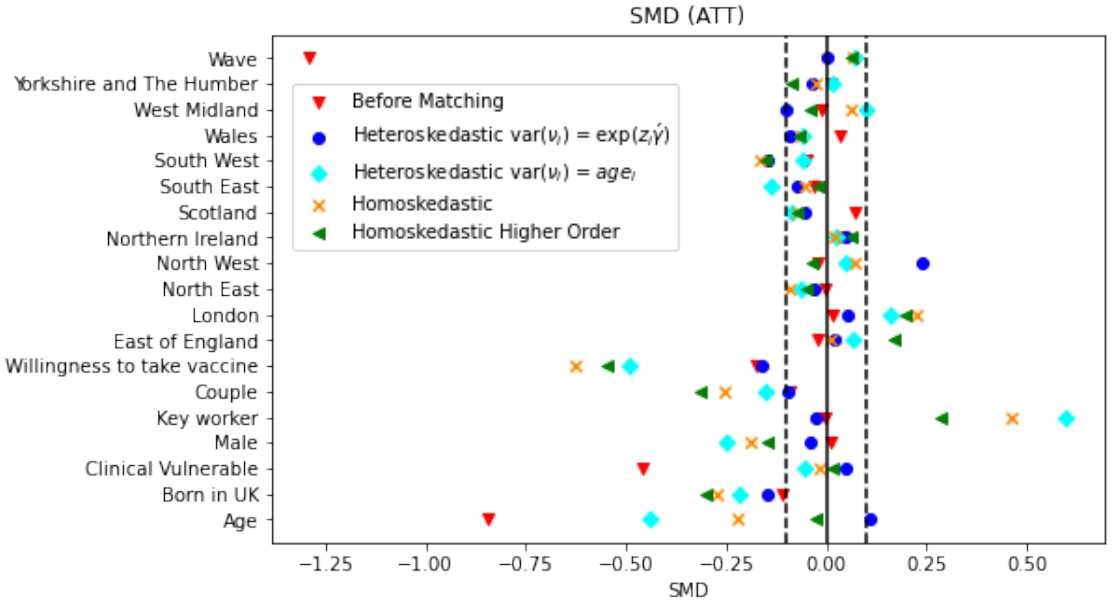


Figure 12: SMD (ATT Sample)



4.2.2 Application: Inverse Probability of Treatment Weighting

We computed the Average Treatment Effect (ATE) utilizing both a heteroskedastic model and a homoskedastic model to assess the impact of COVID-19 vaccination on mental well-being, as outlined

in Table 16. The marginal likelihood strongly supports the heteroskedastic model, and its estimate closely aligns with the findings from propensity score matching. However, the ATE distribution's standard deviation, significantly larger than that in the propensity score matching section, renders the impact of COVID-19 vaccination less precise than the propensity score matching method. The table also demonstrates a well-known problem with inverse probability estimators, namely their large variance because there is no guarantee that the inverse probability must be bounded. Moreover, the estimated treatment effect in the homoskedastic model lacks interpretability, because it lies entirely outside the range of the mental health score $[0, 36]$. This underscores the importance of heteroskedasticity in this context.

Table 16: Impact of COVID-19 Vaccination on Mental Health (IPTW)

Model	Mean	SD	95% CI	Marginal Likelihood
Heteroskedastic	3.4518	1.8518	(-0.3327, 6.9847)	-7633.95
Homoskedastic	54.2305	8.3413	(40.2458, 72.3875)	-8753.28

5 Conclusion

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

Simulation studies have been provided in order to evaluate the empirical consequences of omitted heteroskedasticity, assess the performance of the proposed estimation algorithms, and validate the proposed model comparison techniques. Our investigation has revealed that when non-linearity is pronounced, ignoring heteroskedasticity can result in biased estimates of treatment effects. We

also find that the proposed MCMC methods perform well and can recover the true parameters and models used in generating the data.

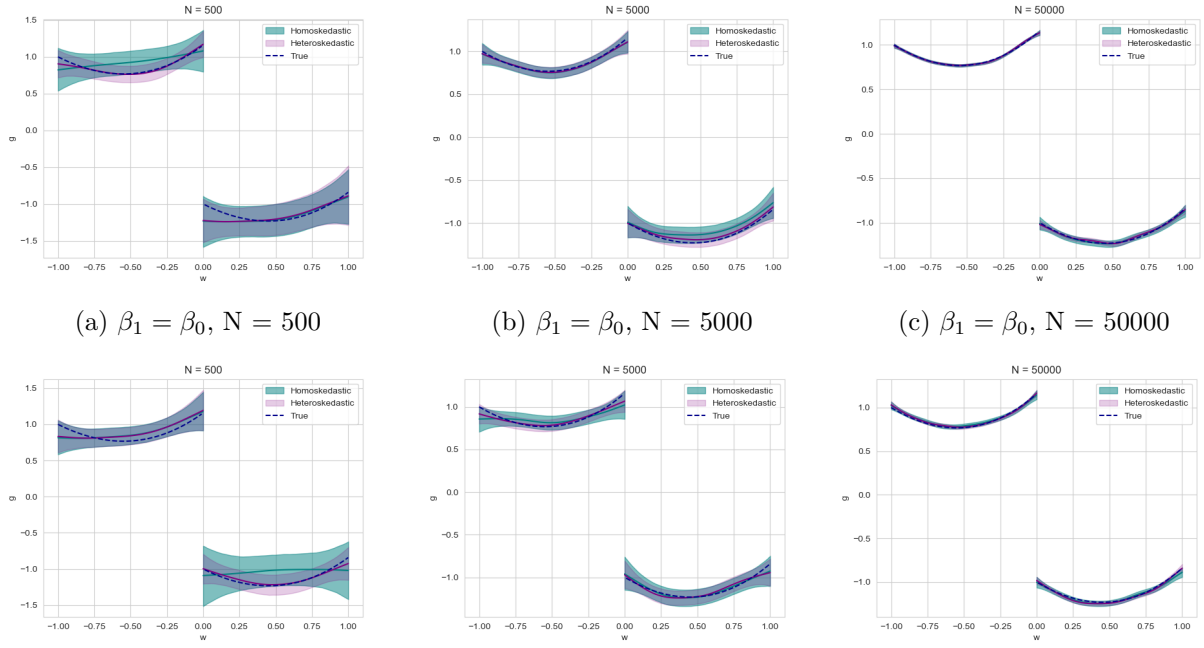
To assess the practical applicability and relevance of our methods, the paper has devoted considerable attention to several applications. In particular, we have explored the impact of academic probation on students' academic performance, the effects of Medigap policies on out-of-pocket healthcare expenditures, and the influence of COVID-19 vaccination on mental well-being. RDD results suggest that academic probation improves subsequent semester GPA, while exhibiting no discernible impact on graduation rates. Using a Rubin causal model (Roy-type model) in our second application, we find that that Medigap policies are expected to reduce out-of-pocket healthcare expenditures. Finally, results from propensity score matching and inverse probability of treatment weighting indicate that COVID-19 vaccination improved the mental well-being of vaccine recipients in the UK. Based on model comparisons in each application, we found that heteroskedastic models were favored in all settings. The results emphasize the importance of allowing for heteroskedasticity in observational observational studies of causal effects and demonstrate that the presence of heteroskedasticity can be uncovered through model comparisons. While our analysis has primarily centered on the effects of heteroskedasticity, we believe that other concerns such as sample selection or endogeneity may also be present in many settings. We intend to study their impact, as well as their interactions with heteroskedasticity, on treatment effect estimation in future work.

Appendix A Sharp RDD: Additional Simulation Results

Table 17: ATE with Continuous Outcome Variable (Sensitivity Check)

	Model	True ATE	RD ATE	SD	95% CI
n = 500	Homoskedastic	-2.2357	-2.3684	0.2494	(-2.8652, -1.8849)
	Heteroskedastic	-2.2357	-2.4144	0.2309	(-2.8728, 1.9686)
	RDRobust	-2.2357	-2.6305	0.6780	(-4.1775, -1.0734)
n = 5000	Homoskedastic	-2.2195	-2.0407	0.1271	(-2.2888, -1.7895)
	Heteroskedastic	-2.2195	-2.0941	0.1061	(-2.2996, -1.8834)
	RDRobust	-2.2195	-2.1892	0.2142	(-2.7119, -1.7242)
n = 50000	Homoskedastic	-2.2092	-2.1986	0.0409	(-2.2793, -2.1191)
	Heteroskedastic	-2.2092	-2.1991	0.0288	(-2.2557, -2.1427)
	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) $\beta_1 = \beta_0$, N = 5000 (c) $\beta_1 = \beta_0$, N = 50000

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

Appendix B MEPS: Coefficients

Table 18: Heteroskedastic Model (2020)

	Parameter	Mean	SD	95% CI			Parameter	Mean	SD	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 19: Heteroskedastic Model (2018, 2019)

	Parameter	Mean	SD	95% CI			Parameter	Mean	SD	95% CI	
	const	-0.7571	0.0636	-0.8834	-0.6312	γ_T	FAMINC	0.1347	0.0055	0.1240	0.1457
	AGE	-0.0052	0.0207	-0.0456	0.0349		const	-8.4420	0.0786	-8.5970	-8.2916
	FAMINC	0.1512	0.0091	0.1334	0.1691	γ_0	AGE	0.3319	0.0311	0.2727	0.3930
	NUM_VISIT	0.0020	0.0011	-0.0001	0.0041		NUM_CHRON	0.0452	0.0138	0.0181	0.0719
	NUM_CHRON	0.0191	0.0093	0.0012	0.0376		const	-8.9472	0.1254	-9.2025	-8.7126
	EXCHLTH	-0.0787	0.0610	-0.1980	0.0405	γ_1	AGE	0.5289	0.0501	0.4290	0.6243
	POORHLTH	-0.0513	0.0915	-0.2332	0.1261		NUM_CHRON	0.0300	0.0226	-0.0136	0.0744
	EXCMHLTH	0.0584	0.0493	-0.0380	0.1543						
β_T	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						
	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						
β_0	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						
	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						
β_1	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						

Table 20: Homoskedastic Model

		2020				2018, 2019			
	Parameter	Mean	SD	95% CI		Mean	SD	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

Ahamada, I. and Flachaire, E. (2010), *Non-Parametric Econometrics*, Oxford: Oxford University Press.

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

Amaratunga, D. and Cabrera, J. (2001), "Outlier Resistance, Standardization, and Modeling Issues for DNA Microarray Data," *Statistics in Genetics and in the Environmental Sciences*, pp. 17–26.

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

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

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems," *The annals of statistics*, pp. 1152–1174.

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.

Celeux, G. (1998), "Bayesian Inference for Mixture: The Label Switching Problem," in *COMPSTAT: Proceedings in Computational Statistics 13th Symposium held in Bristol, Great Britain, 1998*, pp. 227–232, Springer.

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 Wellbeing," *European Economic Review*, 150, 104293.

Chesnaye, N. C., Stel, V. S., Tripepi, G., Dekker, F. W., Fu, E. L., Zoccali, C., and Jager, K. J. (2022), "An introduction to Inverse Probability of Treatment Weighting in Observational Research," *Clinical Kidney Journal*, 15, 14–20.

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 Greenberg, E. (2010), "Additive Cubic Spline Regression with Dirichlet Process Mixture Errors," *Journal of Econometrics*, 156, 322–336.

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.

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

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The annals of statistics*, pp. 209–230.

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.

Gelman, A. and Imbens, G. (2019), "Why High-Order Polynomials Should not be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 37, 447–456.

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.

Jeliazkov, I. and Vossmeyer, A. (2018), "The impact of estimation uncertainty on covariate effects in nonlinear models," *Statistical Papers*, 59.

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 Poirier, D. J. (2004), "Bayesian Variants of Some Classical Semiparametric Regression Techniques," *Journal of Econometrics*, 123, 259–282.

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

Koop, G., Poirier, D. J., and Tobias, J. (2005), "Semiparametric Bayesian Inference in Multiple Equation Models," *Journal of Applied Econometrics*, 20, 723–747.

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.

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

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

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

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

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.

Rue, H. and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, CRC press.

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

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

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

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.

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

Vossmeyer, 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.

Williams, C. K. I. (1998), "Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond," in *Learning in Graphical Models*, ed. M. I. Jordan, 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.