Binary and Ordered Response Models in Randomized Experiments: Applications of the Resampling-Based Maximum Likelihood Method

Ito, Takahiro

(Citation)
GSICS Working Paper Series,42:1-22

(Issue Date)
2024-01

(Resource Type)
technical report

(Version)
Version of Record

(URL)
https://hdl.handle.net/20.500.14094/0100486169

GSICS Working Paper Series

Binary and Ordered Response Models in Randomized Experiments: Applications of the Resampling-Based Maximum Likelihood Method

Takahiro ITO

No. 42

January 2024

Graduate School of International Cooperation Studies

Kobe University

Binary and Ordered Response Models in Randomized Experiments:
Applications of the Resampling-Based Maximum Likelihood Method[*]

Takahiro Ito[†]

January 6, 2024

Abstract

This paper formulates a novel distribution-free maximum likelihood estimator for binary and ordered response models and demonstrates its finite sample performance in a Monte Carlo simulation. The simulation examines an ordered response model, focusing on estimating the effect of an exogenous regressor (e.g., randomly assigned treatment status) on the choice probability for an ordered outcome. Estimations are implemented based on a binary specification, which converts the outcome to dichotomous values $\{0, 1\}$, or an ordinal specification, which uses the outcome as is. The simulation results show that the proposed estimator outperforms conventional parametric/semiparametric estimators in most cases for both specifications. The results also show that the superiority of the proposed estimator holds even in the presence of conditionally heteroscedastic variance. In addition, the estimates based on the ordinal specification are always superior to those based on the binary specification in all simulation designs, implying that converting ordered responses to dichotomous responses and estimating based on the binary specification may not be the optimal approach.

JEL codes: C14, C25

Keywords: semiparametric estimation, distribution-free maximum likelihood, binary choice model, ordered response model, Likert-type data, heteroscedastic variance

[*] This study was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research) Grant Number 22K01425.

[†] Graduate School of International Cooperation Studies, Kobe University, 2–1, Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan. E-mail: takahiro.ito@lion.kobe-u.ac.jp, Phone: +81-78-803-7148.

1. Introduction

Over the past decade, there has been a sweeping trend in empirical fields in which estimation assumptions that researchers cannot verify tend to be eschewed. This so-called scientific humility is evident both in the prevalence of randomized experiments and in the dominance of "harmless" estimation methods, especially in microempirical studies. Randomized experiments require no ad hoc assumptions regarding the independence between a regressor of interest and the error, and "harmless" methods attempt to estimate with minimal unverifiable assumptions. Both approaches aim to make the estimates robust to possible model misspecification.

Based on this background, the use of conventional (parametric) maximum likelihood (ML) estimation methods has been increasingly avoided by researchers. The asymptotic properties of ML estimators depend on the distributional assumptions of the errors in the estimation model. However, these assumptions cannot be verified, which reduces the application potential of the conventional ML method, in spite of its many advantages when the model is correctly specified.

Meanwhile, theoretical developments have been ahead of the recent trend in the empirical fields. To address the drawbacks of parametric ML methods, many studies have proposed alternative semiparametric methods for limited dependent variable models, which are typical applications of ML methods. For example, listing only the ML-based methods closely related to this study, several semiparametric estimators have been proposed, such as Cosslett's (1983) infinite-dimensional ML, the sieve ML (Duncan, 1986; Fernandez, 1986; Gallant and Nychka, 1987), Nawata's (1990) grouping-based ML, and the kernel ML (Klein and Spady, 1993; Lee, 1995; Ai, 1997; Ichimura and Thompson, 1998) estimators.[1]

In this study, I propose an alternative distribution-free ML estimator for binary and ordered response models by applying the resampling-based ML (RBML) estimator developed by Ito (2023). Binary and ordered response models are extensively used in empirical fields, especially in the behavioral and experimental social sciences, where Likert scale items are often employed as outcome variables.[2] However, the semiparametric methods proposed thus far are seldom employed, probably due to their practical

---

[1] Other semiparametric methods than ML-based methods include maximum score estimation for discrete choice models (Manski, 1975, 1985; Horowitz, 1992), censored least absolute deviations estimation for censored regression models (Powell, 1984, 1986a; Newey and Powell, 1990), semiparametric least squares estimation (Horowitz, 1986; Ichimura and Lee, 1991; Lee, 1992; Ichimura, 1993), trimmed least squares (and trimmed least absolute deviations) estimation for censored and truncated regression models (Powell, 1986b; Honoré, 1992; Honoré and Powell, 1994), average derivative estimation (Stoker, 1986, 1991; Härdle and Stoker, 1989; Powell et al., 1989), maximum rank correlation estimation (Han, 1987; Sherman, 1993; Cavanagh and Sherman, 1998), and differencing estimation for sample selection models (Robinson, 1988; Ahn and Powell, 1993; and Yatchew, 1997). In addition, for sample selection models, several semiparametric estimations based on the control function approach have been proposed (Lee, 1982; Andrews, 1991; Das et al., 2003; and Newey, 2009).

[2] For example, approximately 10% (or 39) of the 401 articles published in *Journal of Economic Behavior and Organization* in 2020 use Likert-type ordinal variables. Among the 39 articles, 30 articles treat them as cardinal variables without considering the ordinal nature.

inconvenience. Ito's (2023) RBML method has the potential to bridge the gap between the needs in empirical fields and the sparsity of well-performing practical semiparametric estimators.

The key to the theoretical and practical advantages of this method is the use of a parametric likelihood function. By leveraging the asymptotic normality of the mean of resamples obtained by repeating Monte Carlo resampling with replacement from the original sample, the proposed method exploits a parametric likelihood function without any distributional assumption on the error (in the original equation). Thus, as shown by Ito (2023), the estimator possesses asymptotic properties comparable to those of the parametric ML estimator: The proposed estimator is consistent and asymptotically normally distributed (at rate $N^{-1/2}$). The Monte Carlo study by Ito (2023) also showed that the estimator is strongly consistent and efficient compared to probit and other ML-based semiparametric estimators.

In addition, employing a parametric likelihood function can alleviate the convergence problem. While semiparametric methods often have difficulty optimizing the likelihood function due to the complex computations of the unknown function (and probably its undulating shape), the proposed method is expected to have less difficulty maximizing the function, similar to parametric ML methods. Ito's (2023) simulation analysis showed that the RBML method converged in all trials, while there were many cases in which other semiparametric estimators did not converge, even though simple models were used in the simulation.[3]

This study also explores the small-sample performance of the RBML estimator by running a series of simulations for binary and ordered response models. In contrast to the simulation performed by Ito (2023), the Monte Carlo analysis in this study focuses on estimating the marginal impact of the regressor on the choice probability in more realistic situations. Specifically, the simulation is designed to be flexible to determine how discrete choice models should be analyzed in an experimental setting where the outcome variable is ordinal and the regressors are an exogenous treatment variable and other observed and unobserved components that are allowed to be correlated with each other. Moreover, the estimations are implemented using two different specifications. One is a binary specification in which the ordered outcome is converted to dichotomous values $\{0, 1\}$, and the other is an ordinal specification in which the ordered outcome are used as it is. Then, the small-sample performance of the proposed estimator is compared between the two specifications and with that of conventional estimators.

The simulation results comparing the root-mean-square (relative) errors of the estimates show that the RBML estimator performs considerably better than other conventional estimators, such as probit-type and sieve ML estimators, in both the binary and ordinal specifications. The results also show that the superiority of the proposed estimator holds even in the presence of conditionally heteroscedastic variance, as suggested by the theoretical discussion. In addition, the estimates based on the ordinal specification are always superior to those based on the binary specification in all simulation designs. When the outcome

[3] Moreover, the new ML method is free from the perfect prediction (or complete separation) problem: it focuses on variations around the mean of (dependent and explanatory) variables by nature and not the one-to-one correspondence between them.

variable is ordinal, researchers often employ OLS estimation (i.e., a linear probability model) by converting the outcome to binary. However, the above findings indicate that this approach may not always be a good option. In summary, the simulation analysis conducted in this study indicates that RBML estimation is the preferred method for estimating binary and ordinal response models.[4]

The remainder of this paper is organized as follows. Section 2 reviews two approaches to estimating binary and ordered response models. In Section 3, I present an application example of an ordered response model. Section 4 describes the design of the Monte Carlo simulations and reports the results. Then, the conclusions follow in Section 5.

2. Semiparametric Estimation of Ordered Response Models

This section discusses estimation approaches for binary and ordered response models. The main focus is on the case of ordinal outcomes, but without loss of generality it is applicable to the binary case as well. In terms of the model's specification, there are two distinct approaches.

2.1. Latent index approach

Let y_i be an ordinal outcome ($y_i \in \mathcal{Y} = \{1, \cdots, \mathcal{L}\}$) and x_i denote a treatment variable in an experiment (i.e., x_i is randomly assigned to individuals). In addition, $x_i \in \mathcal{X} = [\underline{\mathcal{X}}, \overline{\mathcal{X}}] \subset \mathbb{R}$, where $\underline{\mathcal{X}}$ and $\overline{\mathcal{X}}$ are minimum and maximum values of the treatment status. To examine the effect of x_i on y_i based on semiparametric (or parametric) methods, I start by introducing a latent index with the following conditions:

<u>LIA Assumption:</u>

(a) There is an unknown function that associates x_i (and other determinants, $\mathbf{w}_i \in \mathbb{R}^L$) with the outcome y_i, and the function has a positive or negative monotonic relationship with y_i:

$$\exists\ \varphi_i: \mathbb{R}^{L+1} \to \mathbb{R} \text{ s.t.}$$

$$\forall (y_i, x_i, \mathbf{w}_i), (y_i', x_i', \mathbf{w}_i') \in \mathcal{Y} \times \mathcal{X} \times \mathbb{R}^L;$$

$$\varphi_i(x_i', \mathbf{w}_i') \gtreqless \varphi_i(x_i, \mathbf{w}_i) \Rightarrow y_i' \gtreqless y_i \text{ (positive), or}$$

$$\varphi_i(x_i', \mathbf{w}_i') \lesseqgtr \varphi_i(x_i, \mathbf{w}_i) \Rightarrow y_i' \gtreqless y_i \text{ (negative)}$$

(b) φ_i is bounded and continuous in $\mathcal{X}$.

For example, suppose that y_i is a Likert-type variable that represents a psychometric scale expressed by respondents about a subject of interest in the experiment. Then, $\varphi_i(x_i, \mathbf{w}_i)$ represents respondents' psychological attitude or belief regarding the subject, which is unobservable to econometricians.

---

[4] All simulation results presented in this study can be replicated using a software package for Stata to implement the new distribution-free estimation in linear and discrete response models, which is available on my website. For details on the process of obtaining and using the package, see the Online Supplementary Material I of Ito (2023).

Under the (positive) monotonicity assumption (LIA Assumption (a)), a natural specification of the model is:

$$y_i = \ell \ \text{ if } \ c_{i,\ell} \geq \varphi_i(x_i, \mathbf{w}_i) > c_{i,\ell-1},$$

where $c_{i,\ell}$ ($\ell = 1, \cdots, \mathcal{L}$) are thresholds determining the value of y_i, where $c_{i,\ell}$ increases as ℓ increases ($c_{i,0} < c_{i,1} < \cdots < c_{i,\mathcal{L}}$), $c_{i,0} = -\infty$ and $c_{i,\mathcal{L}} = \infty$. Then, by LIA Assumption (b), applying a generalization of the Taylor expansion (Feller, 1971), we obtain

$$\varphi_i(x_i, \mathbf{w}_i) = \sum_{k=0}^{\infty} \frac{d^k \varphi_i(\bar{x}, \mathbf{w}_i) \cdot (x_i - \bar{x})^k}{k!},$$

where $d^k \varphi_i(\bar{x}, \mathbf{w}_i)$ is defined by

$$d^k \varphi_i(\bar{x}, \mathbf{w}_i) = \begin{cases} \lim\limits_{h \to 0^+} \dfrac{\Delta_h^k \varphi_i(\bar{x}, \mathbf{w}_i)}{h^k} & \text{if } x_i \geq \bar{x} \\ \lim\limits_{h \to 0^-} \dfrac{\Delta_h^k \varphi_i(\bar{x}, \mathbf{w}_i)}{h^k} & \text{if } x_i < \bar{x} \end{cases}.$$

Δ_h^k is the k-th finite difference operator (with respect to $\bar{x} = N^{-1} \sum_{i=1}^{N} x_i$) with step size h, that is, $\Delta_h^k \varphi_i(\bar{x}, \mathbf{w}_i) = \Delta_h^{k-1} \varphi_i(\bar{x} + h, \mathbf{w}_i) - \Delta_h^{k-1} \varphi_i(\bar{x}, \mathbf{w}_i)$ and $\Delta_h^1 \varphi_i(\bar{x}, \mathbf{w}_i) = \Delta_h^0 \varphi_i(\bar{x} + h, \mathbf{w}_i) - \Delta_h^0 \varphi_i(\bar{x}, \mathbf{w}_i) = \varphi_i(\bar{x} + h, \mathbf{w}_i) - \varphi_i(\bar{x}, \mathbf{w}_i)$. Then, the following condition is further assumed.

LIA Assumption:

(c) There is a $K \in \mathbb{N}$ such that $d^k \varphi_i(x, \mathbf{w}_i)$ (for any $k > K$) in the above Taylor equation is negligibly small in $\mathcal{X}$:

$$\forall\, \epsilon > 0, \exists\, K \in \mathbb{N} \text{ s.t. } \forall\, i, \forall\, x \in \mathcal{X}, \forall\, k \in \mathbb{N}; k > K \Rightarrow \left| d^k \varphi_i(x, \mathbf{w}_i) \right| < \epsilon.$$

Thus, under LIA Assumptions (a)-(c), the following equation is derived:

$$\begin{aligned} \varphi_i(x_i, \mathbf{w}_i) &= \left[\sum_{k=0}^{K} \frac{d^k \varphi_i(\bar{x}, \mathbf{w}_i)}{k!} (-\bar{x})^k \right] + \left\{ \sum_{k=1}^{K} \frac{d^k \varphi_i(\bar{x}, \mathbf{w}_i)}{k!} \binom{k}{k-1} (-\bar{x})^{k-1} \right\} x_i \\ &\quad + \left\{ \sum_{k=2}^{K} \frac{d^k \varphi_i(\bar{x}, \mathbf{w}_i)}{k!} \binom{k}{k-2} (-\bar{x})^{k-2} \right\} x_i^2 + \cdots + \sum_{k=K+1}^{\infty} \frac{d^k \varphi_i(\bar{x}, \mathbf{w}_i)}{k!} (x_i - \bar{x})^k \\ &= \alpha_i + \sum_{k=1}^{K} \beta_{i,k} x_i^k + R_{i,K}, \end{aligned}$$

where $\binom{m}{n}$ is an n-combination of an m-element set (i.e., $\binom{m}{n} = \frac{m!}{n!(m-n)!}$), $\alpha_i = \sum_{k=0}^{K} (k!)^{-1} \cdot d^k \varphi_i(\bar{x}, \mathbf{w}_i) \cdot (-\bar{x})^k$, $\beta_{i,j} = \sum_{k=j}^{K} (k!)^{-1} \cdot d^k \varphi_i(\bar{x}, \mathbf{w}_i) \cdot \binom{k}{k-r} \cdot (-\bar{x})^{k-j}$ for $j \leq K$, and $R_{i,K} = \sum_{k=K+1}^{\infty} (k!)^{-1} \cdot d^k \varphi_i(\bar{x}, \mathbf{w}_i) \cdot (x_i - \bar{x})^k$. Therefore, decomposing $\varphi_i(x_i, \mathbf{w}_i)$ into the conditional expectation given x_i and the remaining term yields

$$\varphi_i(x_i, \mathbf{w}_i) = \mathrm{E}[\varphi_i(x_i, \mathbf{w}_i) | x_i] + \varepsilon_i = \mathrm{E}\left[\alpha_i + \sum_{k=1}^{K} \beta_{i,k} x_i^k + R_{i,K} \,\middle|\, x_i \right] + \varepsilon_i = \bar{\alpha} + \sum_{k=1}^{K} \bar{\beta}_k x_i^k + \varepsilon_i',$$

where $\bar{\alpha} = \mathrm{E}[\alpha_i|x_i] = \mathrm{E}[\alpha_i]$, $\bar{\beta}_k = \mathrm{E}[\beta_{i,k}|x_i] = \mathrm{E}[\beta_{i,k}]$, $\varepsilon_i' = (\alpha_i - \bar{\alpha}) + \sum_{k=1}^{K} (\beta_{i,k} - \bar{\beta}_k) x_i^k + \mathrm{E}[R_{i,K}|x_i]$.

Finally, we have

$$\begin{aligned}\Pr[y_i = \ell] &= \Pr[c_{i,\ell} \geq \varphi_i(x_i, \mathbf{w}_i) > c_{i,\ell-1}] \\ &= \Pr\left[\bar{c}_\ell - \bar{\alpha} - \sum_{k=1}^{K} \bar{\beta}_k x_i^k \geq \varepsilon_i' - c_{i,\ell}'\right] - \Pr\left[\bar{c}_{\ell-1} - \bar{\alpha} - \sum_{k=1}^{K} \bar{\beta}_k x_i^k \geq \varepsilon_i' - c_{i,\ell-1}'\right],\end{aligned}$$

where $\bar{c}_\ell = \mathrm{E}[c_{i,\ell}|x_i] = \mathrm{E}[c_{i,\ell}]$, and $c_{i,\ell}' = c_{i,\ell} - \bar{c}_\ell$.

Notably, the key identification condition for $\bar{\beta}_k$ is that $R_{i,K}$ is negligible (by LIA Assumption (c)) and that the treatment variable x_i is independent of the functional form of $\varphi_i(\cdot)$, other factors $\mathbf{w}_i$ and the thresholds $c_{i,\ell}$ (from the randomness of x_i by definition). Thus, $\mathrm{E}[\alpha_i|x_i] = \mathrm{E}[\alpha_i]$, $\mathrm{E}[\beta_{i,k}|x_i] = \mathrm{E}[\beta_{i,k}]$, $\mathrm{E}[c_{i,\ell}|x_i] = \mathrm{E}[c_{i,\ell}]$ and $\mathrm{E}[R_{i,K}|x_i] \approx 0$, which ensures $\mathrm{E}[\varepsilon_i' - c_{i,\ell}'|x_i] \approx 0$. Note that it is possible to control for some observed variables (a subset of $\mathbf{w}_i$) in addition to x_i, and the inclusion of control variables does not affect the causal estimation of x_i. Therefore, given knowledge of $\varphi_i(\cdot)$, distribution-free ML methods, including Ito's (2023) RBML estimation method, can be employed to identify $\bar{\beta}_k$ without further assumptions.[5]

In actual applications, however, we have no information on $\varphi_i(\cdot)$. Thus, the validity of LIA Assumptions (b) and (c) may be extremely questionable. However, the existence of higher-order terms can be evaluated empirically by including them on the right-hand side and performing statistical tests. Furthermore, in some special cases, the causal relationship between x_i and y_i can be estimated without LIA Assumptions (b) and (c). For example, when the treatment status is dichotomous (namely, being treated or not), as in many experimental settings, the equation takes a simple form:[6]

$$y_i = \mathrm{E}[\alpha_i + \beta_i x_i | x_i] + \varepsilon_i = \bar{\alpha} + \bar{\beta} x_i + \varepsilon_i, \qquad (1)$$

where $\alpha_i = \varphi_i(0, \mathbf{w}_i)$ and $\bar{\alpha} = \mathrm{E}[\alpha_i|x_i] = \mathrm{E}[x_i]$; $\beta_i = \varphi_i(1, \mathbf{w}_i) - \varphi_i(0, \mathbf{w}_i)$ and $\bar{\beta} = \mathrm{E}[\beta_i|x_i] = \mathrm{E}[\beta_i]$; and $\varepsilon_i = (\alpha_i - \bar{\alpha}) + (\beta_i - \bar{\beta}) x_i$ and $\mathrm{E}[\varepsilon_i|x_i] = 0$. Additionally, as discussed in Section 3.2, $(\beta_i - \bar{\beta}) x_i$ in ε_i is the source of heteroskedasticity, but the RBML method can address this problem.

2.2. Direct approach

The second approach attempts to estimate the effect of x_i on y_i by directly connecting y_i and (a function

---

[5] If we employ a fully parametric model such as an ordered probit (ordered logit) model to estimate $\bar{\beta}_k$, we must also assume that $(\varepsilon_i' - c_{i,\ell}')$ is identically and independently distributed with a normal (logistic) distribution. When the error component is normally distributed with mean zero and variance σ^2, we have

$$\Pr[y_i = \ell] = \Phi\left(c_\ell^* - \alpha^* - \sum_{k=1}^{K} \beta_k^* x_i^k\right) - \Phi\left(c_{\ell-1}^* - \alpha^* - \sum_{k=1}^{K} \beta_k^* x_i^k\right),$$

where Φ is the cumulative distribution function of the standard normal distribution, $c_\ell^* = \bar{c}_\ell/\sigma$, $\alpha^* = \bar{\alpha}/\sigma$, and $\beta_k^* = \bar{\beta}_k/\sigma$.

[6] In addition, if y_i and x_i are jointly normal, the expected value of y_i given x_i has no second- or higher-order terms, and the equation has the same form as Eq. (1).

of) x_i based on the following assumptions:

DA Assumption:

(a) The values of y_i are cardinal numbers (i.e., y_i is a cardinal variable).

(b) There is an unknown function that associates x_i and other determinants ($\mathbf{w}_i \in \mathbb{R}^L$) with the outcome y_i:

$$\exists\, \psi_i : \mathbb{R}^{L+1} \to \mathbb{R} \text{ s.t. } \forall (y_i, x_i, \mathbf{w}_i) \in \mathcal{Y} \times \mathcal{X} \times \mathbb{R}^L;\ y_i = \psi_i(x_i, \mathbf{w}_i),$$

(c) ψ_i is bounded and continuous in $\mathcal{X}$, and therefore, we have $\psi_i(x_i, \mathbf{w}_i) = \sum_{k=0}^{\infty}\{k!\}^{-1} d^k\psi_i(\bar{x}, \mathbf{w}_i) \cdot (x_i - \bar{x})^k$, where $d^k\psi_i(\bar{x}, \mathbf{w}_i)$ is defined by

$$d^k\psi_i(\bar{x}, \mathbf{w}_i) = \begin{cases} \lim\limits_{h \to 0^+} \dfrac{\Delta_h^k \psi_i(\bar{x}, \mathbf{w}_i)}{h^k} & \text{if } x_i \geq \bar{x} \\ \lim\limits_{h \to 0^-} \dfrac{\Delta_h^k \psi_i(\bar{x}, \mathbf{w}_i)}{h^k} & \text{if } x_i < \bar{x} \end{cases}.$$

(d) There is a $K \in \mathbb{N}$ such that $d^k\psi_i(x, \mathbf{w}_i)$ (for any $k > K$) is negligibly small in $\mathcal{X}$:

$$\forall\, \epsilon > 0, \exists\, K \in \mathbb{N} \text{ s.t. } \forall\, i, \forall\, x \in \mathcal{X}, \forall\, k \in \mathbb{N}; k > K \Rightarrow \left|d^k\psi_i(x, \mathbf{w}_i)\right| < \epsilon.$$

Then, based on the above assumptions, we have

$$y_i = \mathrm{E}[\psi_i(\bar{x}_i', \mathbf{w}_i)|x_i] + v_i = \mathrm{E}\left[\gamma_i + \sum_{k=1}^{K} \delta_{i,k} x_i^k + S_{i,K}\middle| x_i\right] + v_i = \bar{\gamma} + \sum_{k=1}^{K} \bar{\delta}_k x_i^k + v_i',$$

where $\gamma_i = \sum_{k=1}^{K}(k!)^{-1} \cdot d^k\psi_i(\bar{x}, \mathbf{w}_i) \cdot (-\bar{x})^k$, $\delta_{i,l} = \sum_{k=l}^{K}(k!)^{-1} \cdot d^k\psi_i(\bar{x}, \mathbf{w}_i) \cdot \binom{k}{k-l} \cdot (-\bar{x})^{k-l}$ for $l \leq K$, $S_{i,K} = \sum_{k=K+1}^{\infty}(k!)^{-1} d^k\psi_i(\bar{x}, \mathbf{w}_i)(x_i - \bar{x})^k$, $\bar{\gamma} = \mathrm{E}[\gamma_i|x_i] = \mathrm{E}[\gamma_i]$, $\bar{\delta}_j = \mathrm{E}\left[\delta_{i,j}|x_i\right] = \mathrm{E}\left[\delta_{i,j}\right]$, and $v_i' = (\gamma_i - \bar{\gamma}) + \sum_{k=1}^{K}\left(\delta_{i,j} - \bar{\delta}_j\right)x_i^k + \mathrm{E}\left[S_{i,K}|x_i\right]$. Note again that $\mathrm{E}[v_i'|x_i] \approx 0$ because of DA Assumption (d) and the random assignment of x_i (by definition). Thus, with knowledge of $\psi_i(\cdot)$ (with DA Assumptions (b) to (d)), we can estimate the effect of x_i on y_i via ordinary least squares (OLS) estimation.

This approach, however, has a significant flaw: There are serious doubts about the validity of DA Assumption (a). Thus, in this approach, y_i is often summarized as a binary variable, D_i, based on a certain criterion (e.g., $D_i = 1[y_i > \ell']$, $\ell' \in (1, \mathcal{L})$). Since cardinality is not required in the binary case, we can estimate $D_i = \bar{\gamma} + \sum_{k=1}^{K} \bar{\delta}_k x_i^k + v_i$ with only DA Assumptions (b) to (d). This is known as a linear probability model (LPM). Although the LPM uses a limited information on the outcome variable in practice, the identification assumptions are the same as those in the latent index approach. Therefore, when the outcome variable is originally binary, the latent index and direct approaches differ only in their way of approaching the model, while relying on the same assumptions. Thus, to determine which method performs better, they must be tested empirically. In Section 4, I compare the performance of several semiparametric estimators for the binary choice and ordered response models by running a series of Monte Carlo simulations.

3. RBML Estimation

The RBML estimation method proposed by Ito (2023) utilizes a parametric likelihood function by leveraging the asymptotic normality of the mean of re-samples obtained by repeated random drawing with replacement from the original sample. Specifically, the method consists of two main steps: 1) construction of a new dataset through Monte Carlo "in-sample" resampling with replacement and 2) construction and estimation of the likelihood. In this section, I first describe these steps briefly using a simple linear regression model as an example, then discuss the issue of conditionally heteroscedastic variance, and finally present an application example for an ordered response model.

3.1. Procedure

Suppose that the sample consists of independent observations $\{(y_i, x_i)|\ i = 1, \cdots, N\}$ and that the model can be expressed as

$$y_i = \alpha_0 + \beta_0 x_i + \varepsilon_i, \tag{2}$$

where $y_i \in \mathbb{R}$ is an outcome of interest, $x_i \in \mathbb{R}$ is an exogenous treatment status, and ε_i is the error with $\mathrm{E}[\varepsilon_i|x_i] = 0$.

In the first step, a new dataset is constructed as follows:

(i) Randomly draw an observation from $\{(y_i, x_i)\}$ M times with replacement (M is sufficiently large).

(ii) Calculate $\tilde{y} = \sqrt{NM/(N-1)}\left(\sum_{j=1}^{M} y_j /M - \mu_{N,y}\right)$ and $\tilde{x} = \sqrt{NM/(N-1)}\left(\sum_{j=1}^{M} x_j /M - \mu_{N,x}\right)$, where y_j and x_j are the j-th drawn observations and μ_N is the sample average, that is, $\mu_{N,y} = \sum_{i=1}^{N} y_i /N$ and $\mu_{N,x} = \sum_{i=1}^{N} x_i /N$.

(iii) Repeat (i) and (ii) T times to obtain an independent and identically distributed (*i.i.d.*) sample $\{(\tilde{y}_t, \tilde{x}_t)|\ t = 1, \cdots, T\}$ ($T = T^* + N$, where T^* is sufficiently large).

The linear relationship between y_i and x_i in Eq. (2) gives

$$\tilde{y}_t = \beta_0 \tilde{x}_t + \tilde{\varepsilon}_t, \tag{3}$$

where $\tilde{\varepsilon}_t \sim \mathrm{N}(0, \sigma_N^2) \xrightarrow{d} \mathrm{N}(0, \sigma_0^2)$, $\sigma_N^2 = \sum_{i=1}^{N} \varepsilon_i^2 /N$, and $\sigma_0^2 = \lim_{N\to\infty} \mathrm{E}[\sigma_N^2]$. Regarding the distribution property of $\tilde{\varepsilon}_t$, see Ito (2023); in particular, see Proposition 1 and Proposition A1 in his paper for the homoscedasticity (i.e., $\mathrm{E}[\varepsilon_i^2|x_i] = \sigma^2$) and heteroscedasticity cases (i.e., $\mathrm{E}[\varepsilon_i^2|x_i] = \sigma_i^2$), respectively. It is noteworthy that even if ε_i in Eq. (2) has a heteroscedastic variance that depends on x_i (i.e., $\sigma_i^2 = h_i(x_i)$), the variance of $\tilde{\varepsilon}_t$ in Eq. (3) does not depend on $\tilde{x}_t$. This could be of great advantage of Ito's (2023) RBML method over conventional ML methods in estimating discrete choice models. The following section discusses such a case of conditionally heteroscedastic variance.

Finally, in the second step, based on the normality of $\tilde{\varepsilon}_t$, we construct and estimate the likelihood function expressed as:

$$L(\beta, \sigma; \tilde{y}_t, \tilde{x}_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\tilde{y}_t - \beta\tilde{x}_t}{\sigma}\right)^2\right\}.$$

3.2. The issue of heteroscedasticity

This section discusses the implications of the heteroscedasticity issue for the RBML method. The presence of conditionally heteroskedastic variance is particularly problematic when estimating discrete choice models. Suppose that Eq. (2) represents a latent index that determines a binary response outcome, e.g., $d_i = 1[y_i > 0] = 1[\alpha_0 + \beta_0 x_i + \varepsilon_i > 0]$, and that the variance of ε_i depends on x_i, $\mathrm{E}[\varepsilon_i^2|x_i] = h(x_i)$. In this case, it is well known that the probit ML estimation yields the inconsistent estimate of β_0 (Yatchew and Griliches, 1985). There is also the concern that the sign of the marginal effect on the choice probability is not the same as the sign of β_0 (Wooldridge, 2005).

For the RBML method, however, this is not the case. As a typical case of heteroscedastic variance, I use the example presented in Section 2, where the error term in the original equation is expressed as $\varepsilon_i = (\alpha_i - \mathrm{E}[\alpha_i]) + (\beta_i - \mathrm{E}[\beta_i])x_i$. Thus, the issue relies on the distributional properties of α_i, β_i, and x_i. Assuming that x_i (a treatment variable in an experiment) is independent of α_i and β_i, the limit distribution of the error in the RBML estimation ($\tilde{\varepsilon}_t$) is expressed as:

$$\tilde{\varepsilon}_t \xrightarrow{d} \mathrm{N}\left(0, \sigma_\alpha^2 + 2\sigma_{\alpha\beta}\mu_x + \sigma_\beta^2\mu_x^2\right). \quad (4)$$

See Appendix A.1 for details on the assumptions and derivation. This indicates that while the conditional variance of the original error ε_i in Eq. (2) depends on x_i, that of $\tilde{\varepsilon}_t$ in Eq. (3) does not depend on $\tilde{x}_t$. In short, the heteroscedasticity of the error in the original equation does not matter in the RBML estimation.

A simulation exercise confirms the above results. Panels A and B in Figure 1 show the distribution plots of the data around the regression lines expressed by Eqs. (2) and (3) (with an additional control variable w_i), respectively. The simulated data used in this exercise were created with the same design as in Section 4, and larger and brighter hexagons indicate higher observation frequencies. While ε_i in Eq. (2) (measured as the distance from the regression line in Panel A) shows more dispersion with increasing x_i, $\tilde{\varepsilon}_t$ in Eq. (3) (the distance from the regression line in Panel B) seems to be unrelated to the value of $\tilde{x}_t$. The regression results reported at the top of each panel also show that ε_i in the original equation is significantly associated with x_i, but this is not the case for $\tilde{\varepsilon}_t$. Thus, conditionally heteroscedastic variance disappears in the data construction process in the RBML estimation, indicating that the RBML estimators do not suffer from this problem, unlike parametric ML estimators.

[Figure 1 around here]

3.3. Application to ordered response model

The RBML method estimates binary and ordered response models in a very similar manner. Here, the application of the RBML method to an ordered response model is presented; for the application to a binary choice model, see Section 3.2 of Ito (2023).

Suppose there exists a sample $\{(y_i, \mathbf{x}_i) | \; i = 1, \cdots, N\}$, where $y_i \in \{1, \cdots, \mathcal{L}\}$ and $\mathbf{x}_i' \in \mathbb{R}^L$ are independent random variables with finite means and variances. Assuming that LIA Assumption (a) holds, the model can be expressed as follows:

$$y_i = \ell\left[c_{i,\ell} \geq \varphi_i(\mathbf{x}_i) > c_{i,\ell-1}\right], \tag{5}$$

where $\ell[\cdot]$ ($\ell \in \{1, \cdots, \mathcal{L}\}$) is an indicator variable that takes the value of ℓ when the condition inside the brackets is true, $\varphi_i: \mathbb{R}^L \to \mathbb{R}$ is an unobserved index function and $c_{i,\ell}$ is the ℓ-th cutoff point with $c_{i,0} = -\infty$ and $c_{i,\mathcal{L}} = \infty$. Then, I assume that $\mathrm{E}[\varphi_i(\mathbf{x}_i)|\mathbf{x}_i] = \alpha_0 + \mathbf{x}_i\boldsymbol{\beta_0} + \mathrm{E}[\varepsilon_i|\mathbf{x}_i]$, where $\alpha_0 \in \mathbb{R}$, $\boldsymbol{\beta}_0 \in \mathbb{R}^L$ are unknown population parameters to be estimated and $\varepsilon_i \in \mathbb{R}$ is an unobserved component. I also assume that $\mathrm{Rank}[\sum_i^N (1, \mathbf{x}_i)'(1, \mathbf{x}_i)] = L + 1$, $\mathrm{E}[\varepsilon_i|\mathbf{x}_i] = 0$ and $\mathrm{E}[\varepsilon_i^2|\mathbf{x}_i] = \sigma_0^2$ $(< \infty)$.

In ordered response models, the data construction through the Monte Carlo "in-sample" resampling explained in Section 3.1 is performed based on groups classified by the value of y_i, with $\mathcal{L} - 1$ groups from $\{(y_i, \mathbf{x}_i) | \; y_i = 1 \text{ or } y_i = 2\}$ to $\{(y_i, \mathbf{x}_i) | \; y_i = \mathcal{L} - 1 \text{ or } y_i = \mathcal{L}\}$. For example, the outcome variable for the t-th observation from $\{(y_i, \mathbf{x}_i) | \; y_i = \ell \text{ or } y_i = \ell + 1\}$ is expressed as

$$\tilde{y}_{\ell,t} = \sqrt{\frac{N_\ell M_\ell}{N_\ell - 1}} \left(\frac{1}{M_\ell} \sum_{y_{j,t} \in \mathcal{Y}_\ell \cup \mathcal{Y}_{\ell+1}} y_{j,t} - \bar{y}_\ell \right),$$

where $\mathcal{Y}_\ell = \{y_i | \; y_i = \ell\}$, $\mathcal{Y}_{\ell+1} = \{y_i | \; y_i = \ell + 1\}$, $N_\ell = \#[\mathcal{Y}_\ell \cup \mathcal{Y}_{\ell+1}]$, $M_\ell = N_\ell \times M / N_\mathcal{L}$, $N_\mathcal{L} = \sum_{\ell=1}^{\mathcal{L}-1} \#[\mathcal{Y}_\ell \cup \mathcal{Y}_{\ell+1}]$ and $\bar{y}_\ell = N_\ell^{-1} \sum_{y_i \in \mathcal{Y}_\ell \cup \mathcal{Y}_{\ell+1}} y_i$. Letting $\varphi_i(\mathbf{x}_i)$ in Eq. (5) be denoted by y_i^* and $\tilde{y}_{\ell,t}^*$ be the latent index corresponding to $\tilde{y}_{\ell,t}$, we obtain the following relationship (see Appendix A.2 for the derivation):

$$\begin{aligned} \tilde{y}_{\ell,t} > 0 \quad &\text{if } \tilde{y}_{\ell,t}^* = \tilde{\mathbf{x}}_{\ell,t}\boldsymbol{\beta}_0 + \tilde{\varepsilon}_{\ell,t} > \gamma_{\ell,t} \\ \tilde{y}_{\ell,t} \leq 0 \quad &\text{if } \tilde{y}_{\ell,t}^* = \tilde{\mathbf{x}}_{\ell,t}\boldsymbol{\beta}_0 + \tilde{\varepsilon}_{\ell,t} \leq \gamma_{\ell,t}. \end{aligned} \tag{6}$$

where $\tilde{\mathbf{x}}_{\ell,t}$ and $\tilde{\varepsilon}_{\ell,t}$ are the (vector of) explanatory variables and the error term for the t-th observation created in the data construction stage from $\{(y_i, \mathbf{x}_i) | \; y_i = \ell \text{ or } y_i = \ell + 1\}$, $\gamma_{\ell,t}$ is a threshold variable defined in Eq. (A-1) in the appendix that is assumed to be independent of $\tilde{\mathbf{x}}_{\ell,t}$ and $\tilde{\varepsilon}_{\ell,t}$,[7] and $(\gamma_{\ell,t} - \tilde{\varepsilon}_{\ell,t})$ follows $\mathrm{N}(0, \sigma_{N,\ell}^2)$. The relationship in Eq. (6) also holds for other pairwise groups, and each group has T_ℓ $(= N_\ell \times T / N_\mathcal{L})$ observations. Thus, the RBML estimator for the ordered response model is defined as values that satisfy

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{\mathrm{RB}} &= \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ln L_N(\boldsymbol{\theta}; \tilde{\mathbf{y}}, \widetilde{\mathbf{X}}) \\ &= \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{\ell=1}^{\mathcal{L}-1} \sum_{t=1}^{T_\ell} \left[1[\tilde{y}_{\ell,t} \leq 0] \ln \Phi(-\tilde{\mathbf{x}}_{\ell,t}\boldsymbol{\theta}_\ell) + 1[\tilde{y}_{\ell,t} > 0] \ln \Phi(\tilde{\mathbf{x}}_{\ell,t}\boldsymbol{\theta}_\ell) \right], \end{aligned} \tag{7}$$

[7] In the absence of this random threshold assumption, it is impossible to estimate $\boldsymbol{\beta}$ unless an additional identification condition is assumed. Ito (2023) shows that the simulation result strongly supports the random threshold assumption.

where $\boldsymbol{\theta}_\ell = \boldsymbol{\beta}/\sigma_{N,\ell}$, $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^{K(\mathcal{L}-1)}$, which contains the true value $\boldsymbol{\theta}_0$, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Note that in the RBML estimation, the parameters in Eq. (7) are identified up to a scale, as in the conventional ordered probit estimation; however, the scale is different (and also varies with ℓ). In the actual estimation, a weighted average of $\boldsymbol{\theta}_\ell$ is estimated. This is because while it is theoretically possible to estimate different $\boldsymbol{\theta}_\ell$, the relative magnitude of any two coefficients remains the same (i.e., $\forall\, k'$ & k; $\theta_{k',\ell}/\theta_{k,\ell} = \beta_{k'}/\beta_k$) for all ℓ.

4. Monte Carlo Simulation

This section examines the performance of the RBML estimator on small samples by running a series of simulations where the RBML method is applied to (1) binary and (2) ordinal response models. The main purpose of the simulations is to determine how the model should be analyzed when ordinal responses are the outcome variable. The details of the simulation design are described below.

4.1. Simulation design

The outcome variable is assumed to be an ordered categorical response on a scale of one to five, which is determined in the following manner:

$$y_i = \sum_{\ell=1}^{5} \ell\left[c_{i,\ell} \geq y_i^* > c_{i,\ell-1}\right]$$

where y_i^* is a latent variable, as explained below, and $c_{i,\ell}$ represents the ℓ-th cutoff point (threshold). When estimating the model as a binary response model, the dependent variable is converted into a dichotomous variable as $d_i = 1[y_i \geq 4]$.

The cutoff points $c_{i,\ell}$ are randomly drawn from the uniform distribution, $c_{i,1} \in \{c|P_5(\mathbf{y}^*) \leq c \leq P_{15}(\mathbf{y}^*)\}$, $c_{i,2} \in \{c|P_{20}(\mathbf{y}^*) \leq c \leq P_{30}(\mathbf{y}^*)\}$, $c_{i,3} \in \{c|P_{45}(\mathbf{y}^*) \leq c \leq P_{55}(\mathbf{y}^*)\}$, and $c_{i,4} \in \{c|P_{75}(\mathbf{y}^*) \leq c \leq P_{85}(\mathbf{y}^*)\}$, where $P_j(\mathbf{y}^*)$ represents the j-th percentile value of $\{y_i^*: i = 1, \cdots, N\}$. These random cutoff points are heterogeneous across observations.

The focus of this simulation is on the marginal effect of a treatment variable, denoted by x_i, on the above categorical outcome y_i. For the treatment variable x_i, I consider two cases: binary and continuous cases. The binary treatment is defined as $x_i^d = 1[a_i < b]$, where $a_i \sim \mathrm{U}(0,1)$, and $b \sim \mathrm{U}(0.3,0.5)$. Thus, in the population, 40% of the observations are treated ($x_i^d = 1$). For the continuous treatment case, $x_i^c = 1[a_i < b] \times c_i$, where $a_i \sim \mathrm{U}(0,1)$, $b \sim \mathrm{U}(0.3,0.5)$, and $c_i \sim \mathrm{U}(0.5,1.5)$; hence, 40% of the observations are assigned uniform random numbers between 0.5 and 1.5, and the other observations are assigned a value of 0.

Then, the latent variable (y_i^*) is determined by

$$y_i^* = \alpha_i + \beta_i x_i + w_i,$$

where x_i is the treatment variable described above, β_i is the individual treatment effect, and α_i and w_i represent the effects of unobservables and observables. α_i is assumed to follow a continuous uniform distribution with a mean of zero and a variance of three in the population ($\alpha_i \sim \mathrm{U}(-3,3)$, with $\mathrm{E}[\alpha_i] = 0$

and $\mathrm{Var}[\alpha_i] = 3$). For β_i, two cases with different distributional assumptions are considered: In the first case, β is constant, $\beta_i = 0.5$ for all i (Design 1), and in the second case, β_i is heterogeneous and assumed to be a random variable that follows an exponential distribution with a rate parameter of one multiplied by 0.5, that is, $\beta_i \sim 0.5 \cdot \mathrm{Exp}(1)$ with $\mathrm{E}[\beta_i] = 0.5$, $\mathrm{Var}[\beta_i] = 0.25$ (Design 2). The reason why the variance of α_i is much larger than that of β_i is that all unobserved components are considered to be included in α_i. Moreover, if the variance of β_i is too large, a significant fraction of β_i values could be of opposite sign (i.e., negative), which could lead to a negative average impact in a small sample.

Then, w_i follows the beta distribution with shape parameters drawn randomly from $\{1,3,5\}$ and is adjusted to have unit variance in the population. The beta distribution was selected because the skewness and kurtosis of variables from the beta distribution can be negative or positive depending on the combination of the shape parameters. Ito (2023) showed that the RBML estimator is more efficient when the regressors are leptokurtic; therefore, the kurtosis of the regressors is randomly determined in this simulation. The correlation between the observed/unobserved components is set as:

$$\mathrm{Corr}(\mathbf{Z}_i) = \begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & a & 1 & \\ 0 & b & c & 1 \end{pmatrix},$$

where $\mathbf{Z}_i = (x_i, w_i, \alpha_i, \beta_i)$, $a, b \sim \mathrm{U}(-0.2, 0.2)$, and $c \sim \mathrm{U}(0, 0.4)$.

[Table 1 around here]

Table 1 summarizes the simulation design. In the simulation, the sample size in a trial is set to 500 ($N = 500$), and each design consists of 500 independent trials. The descriptive statistics of the variables used in a trial are presented in Table A1 in Appendix A.3.

4.2. Results

The simulation results for the binary and continuous treatment cases are reported in Table 2. For discrete choice models, researchers' interest is generally in the marginal effect on the choice probability, not in the coefficient estimate. In addition, the marginal effect to be estimated in this simulation study differs in different trials due to the design. Therefore, for ease of comparison, I present the root mean relative square error (RMRSE), calculated as $\sqrt{T^{-1}\sum_t^T \{(\widehat{ME}_t - ME_t)/ME_t\}^2}$, where ME_t and $\widehat{ME}_t$ are the marginal effect and its estimate in the t-th trial ($t = 1, \cdots, 500$), respectively. In addition, to compare the small-sample performance of the RBML estimator with that of other parametric and semiparametric estimators, the results based on OLS, probit-type ML, Gallant and Nychka's (1987) Hermite polynomial sieve ML, Klein and Spady's (1993) Nadaraya–Watson kernel ML, and Ichimura's (1993) semiparametric least squares (SLS) estimations are also reported in the tables.[8]

[8] The sieve ML, kernel ML, and SLS estimations are implemented using the –snp–, –sml–, and –sls–

[Table 2 around here]

Table 2 shows that the RBML estimator performs quite well in the small sample case. According to the results based on the binary response specification in Panel A, the RBML estimator outperforms the other estimators, including the OLS, probit, and sieve ML estimators, in terms of the RMRSE. Notably, kernel-based semiparametric estimators such as the kernel ML and SLS estimators tend to be less efficient probably because the unknown functions are computed nonparametrically, but the RBML method, which does not require nonparametric calculations, does not suffer from such efficiency loss.

Moreover, according to Panel B, the results based on the ordered response specification show that the RBML estimator is more efficient than other estimators in most cases. It is also noteworthy that the RMRSE values of the three estimators (RBML, ordered probit, and sieve ML estimators) in Panel B are always smaller than those of the estimators in Panel A. This implies that the ordered response model employing ordinal values of the outcome variable improves the accuracy of the causal estimation. When the outcome variable is ordered response data, researchers often employ a linear probability model (LPM), that is, OLS estimation based on a binary choice specification, by converting the outcome into binary. The simulation results indicate that although the LPM (OLS) estimator performs well compared to the probit estimator in the binary specification, it is less efficient than the three estimators in the ordinal response specification in terms of efficiency.

Based on the simulation analysis, the following two conclusions can be drawn. First, when analyzing an ordered outcome variable, it is recommended to use an ordered response model instead of a LPM based on binary specification. Second, the RBML method can be the best approach to apply in realistic situations where there exist unobserved nonnormal components and heterogeneous treatment effects among individuals.

5. Conclusion

This study formulated the innovative distribution-free ML estimator proposed by Ito (2023) for binary and ordered response models and demonstrated how ordinal dependent variables should be analyzed in experimental settings. Consistent with the simulation results in Ito (2023), the Monte Carlo simulation in this study, which focused on marginal effect estimates, showed that the RBML estimator performs exceedingly well in scenarios with nonnormal unobserved components and heteroskedasticity. The results also showed that estimating a binary choice model by binarizing the ordinal outcome variable may not always be a good option.

Although the simulation designs employed in this study are relatively flexible, they represent only a few possible examples. With this caveat, the RBML estimation method may be the best choice for causal inference in binary and ordered response models. Even when estimating these models by conventional

commands in Stata. See De Luca (2008) for the –snp– and –sml– commands and Barker (2014) for the –sls– command.

methods, the RBML method should be implemented to verify the robustness of their estimation results.

Appendix A: Derivations of Eqs. (4) and (6) and summary statistics of the simulated data

A.1. Derivation of the result in Eq. (4)

Taking M observations from $\{\varepsilon_i | \, i = 1, \cdots, N\}$ with resampling with replacement in the first step (see Section 3.1) is equivalent to taking one observation randomly from $\{\varepsilon_i | \, i = 1, \cdots, N\}$ and repeating M times. Therefore, letting $\mathrm{E}[\alpha_i] = \alpha_0$ and $\mathrm{E}[\beta_i] = \beta_0$, $\tilde{\varepsilon}_t$ can be expressed as

$$\tilde{\varepsilon}_t = \sqrt{\frac{NM}{N-1}}\left(\frac{\sum_{j=1}^{M} \varepsilon_{jt}}{M}\right) = \sqrt{\frac{NM}{N-1}}\frac{\sum_i^N \sum_j^M w_{ijt}\varepsilon_i}{M} = \frac{\sqrt{M'}}{M}\sum_i^N \sum_j^M w_{ijt}\{(\alpha_i - \alpha_0) + (\beta_i - \beta_0)x_i\}$$

$$= \frac{\sqrt{M'}}{M}\sum_i^N \sum_j^M w_{ijt}\{(\alpha_i - \bar{\alpha}_N) + (\bar{\alpha}_N - \alpha_0)$$

$$+(\beta_i - \bar{\beta}_N)(x_i - \bar{x}_N) + (\bar{\beta}_N - \beta_0)x_i + (\beta_i - \bar{\beta}_N)\bar{x}_N\}$$

$$= \tilde{\alpha}_t + \sqrt{M'}(\bar{\alpha}_N - \alpha_0) + \tilde{\rho}_t + \frac{\sqrt{M'}}{M}\sum_i^N \sum_j^M w_{ijt}\{(\bar{\beta}_N - \beta_0)(x_i - \bar{x}_N)$$

$$+(\bar{\beta}_N - \beta_0)\bar{x}_N + (\beta_i - \bar{\beta}_N)\bar{x}_N\}$$

$$= \tilde{\alpha}_t + \sqrt{M'}(\bar{\alpha}_N - \alpha_0) + \tilde{\rho}_t + (\bar{\beta}_N - \beta_0)\tilde{x}_t + \sqrt{M'}(\bar{\beta}_N - \beta_0)\bar{x}_N + \tilde{\beta}_t\bar{x}_N,$$

where w_{ijt} is a random variable that has a value of one if the i-th observation is drawn at the j-th iteration in the t-th resampling stage and zero otherwise (hence, $\sum_i^N \sum_j^M w_{ijt} = M$), $M' = NM/(N-1)$, $\tilde{\alpha}_t = \sqrt{M'}/M \cdot \sum_j^M \sum_i^N w_{ijt}(\alpha_i - \bar{\alpha}_N)$, $\tilde{\rho}_t = \sqrt{M'}/M \cdot \sum_j^M \sum_i^N w_{ijt} \ (\beta_i - \bar{\beta}_N)(x_i - \bar{x}_N)$, $\tilde{\beta}_t = \sqrt{M'}/M \cdot \sum_j^M \sum_i^N w_{ijt}(\beta_i - \bar{\beta}_N)$, and $\bar{m}_N = N^{-1}\sum_i^N m_i$ for $m = \alpha, \beta, x$. Then, if we assume that $\{(\alpha_i, \beta_i, x_i) | \, i = 1, \cdots, N\}$ are *i.i.d.* with an unknown joint distribution with finite mean $\boldsymbol{\mu}$ and finite variance $\boldsymbol{\Sigma}$ such that

$$\boldsymbol{\mu} = (\alpha_0, \beta_0, \mu_x), \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\alpha^2 & & \\ \sigma_{\alpha\beta} & \sigma_\beta^2 & \\ 0 & 0 & \sigma_x^2 \end{pmatrix},$$

Proposition 1 of Ito (2023) indicates that $\tilde{\alpha}_t + \tilde{\beta}_t \xrightarrow{d} \mathrm{N}(0, \sigma_\alpha^2 + 2\sigma_{\alpha\beta} + \sigma_\beta^2)$, $\tilde{\rho}_t \xrightarrow{p} 0$, and $\tilde{x}_t \xrightarrow{d} \mathrm{N}(0, \sigma_x^2)$.

Note that if $\{(\alpha_i, \beta_i, x_i) | \, i = 1, \cdots, N\}$ have different means and variances for each i, namely, they are independent but not identically distributed (*i.n.i.d.*), such as

$$\boldsymbol{\mu}_i = (\alpha_{i,0}, \beta_{i,0}, \mu_{i,x}), \boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_{i,\alpha}^2 & & \\ \sigma_{i,\alpha\beta} & \sigma_{i,\beta}^2 & \\ 0 & 0 & \sigma_{i,x}^2 \end{pmatrix},$$

Proposition A.1 of Ito (2023) is applied. Then, from the law of large number (LLN), $\bar{\alpha}_N \xrightarrow{p} \alpha_0$, $\bar{\beta}_N \xrightarrow{p} \beta_0$, and $\bar{x}_N \xrightarrow{p} \mu_x$. Therefore, the conclusion follows from the Slutzky theorem.

A.2. Derivation of Eq. (6)

Let $\tilde{y}^*_{\ell,t}$ be a latent variable obtained in the t-th resampling stage from $\mathcal{Y}^*_\ell \cup \mathcal{Y}^*_{\ell+1}$ in the RBML data construction. Then, this variable can be expressed as

$$\tilde{y}^*_{\ell,t} = \sqrt{\frac{N_\ell M_\ell}{(N_\ell - 1)}}\left(\frac{1}{M_\ell}\sum_{j=1}^{M_\ell} y_{j,t} - \bar{y}^*_\ell\right)$$

$$= \sqrt{\frac{N_\ell M_\ell}{(N_\ell - 1)}}\left(\frac{1}{M_\ell}\sum_{\substack{y^*_i \in \mathcal{Y}^*_\ell \\ \cup \mathcal{Y}^*_{\ell+1}}}\sum_{j=1}^{M_\ell} w_{ijt} y^*_i - \frac{1}{N_\ell}\left(\sum_{y^*_i \in \mathcal{Y}^*_\ell} y^*_i + \sum_{y^*_i \in \mathcal{Y}^*_{\ell+1}} y^*_i\right)\right)$$

$$= \sqrt{\frac{N_\ell M_\ell}{(N_\ell - 1)}}\left\{\frac{1}{M_\ell}\left(\sum_{y^*_i \in \mathcal{Y}^*_\ell}\sum_{j=1}^{M_\ell} w_{ijt} y^*_i + \sum_{y^*_i \in \mathcal{Y}^*_{\ell+1}}\sum_{j=1}^{M_\ell} w_{ijt} y^*_i\right) - \left(n_{\ell,0}\bar{y}^*_{\ell,0} + n_{\ell,1}\bar{y}^*_{\ell,1}\right)\right\},$$

where $\mathcal{Y}^*_\ell = \{y^*_i \mid y_i = \ell\}$, $\mathcal{Y}^*_{\ell+1} = \{y^*_i \mid y_i = \ell + 1\}$, $N_\ell = \#[\mathcal{Y}^*_\ell \cup \mathcal{Y}^*_{\ell+1}]$, $M_\ell = N_\ell \times M/N_\mathcal{L}$, $N_\mathcal{L} = \sum_{\ell=1}^{\mathcal{L}-1} \#[\mathcal{Y}^*_\ell \cup \mathcal{Y}^*_{\ell+1}]$, $\bar{y}^*_\ell = N_\ell^{-1}\sum_{y^*_i \in \mathcal{Y}^*_\ell \cup \mathcal{Y}^*_{\ell+1}} y^*_i$, $n_{\ell 0} = N_{\ell 0}/N_\ell$, $N_{\ell 0} = \#[\mathcal{Y}^*_\ell]$, $\bar{y}^*_{\ell 0} = N_{\ell 0}^{-1}\sum_{y^*_i \in \mathcal{Y}^*_\ell} y^*_i$, $n_{\ell 1} = N_{\ell 1}/N_\ell$, $N_{\ell 1} = \#[\mathcal{Y}^*_{\ell+1}]$, and $\bar{y}^*_{\ell 1} = N_{\ell 1}^{-1}\sum_{y^*_i \in \mathcal{Y}^*_{\ell+1}} y^*_i$. Then, defining $\gamma_{\ell,t}$ as

$$\gamma_{\ell,t} = N_\ell^{-\frac{1}{2}} \sum_{y^*_i \in \mathcal{Y}^*_\ell}\left\{\left(\frac{N_\ell - 1}{N_\ell^2 M_\ell}\right)^{-\frac{1}{2}}\left(\frac{\sum_{j=1}^{M_\ell} w_{ijt}}{M_\ell} - \frac{1}{N_\ell}\right)(y^*_i - \bar{y}^*_{\ell 0})\right\}$$

$$+ N_\ell^{-\frac{1}{2}} \sum_{y^*_i \in \mathcal{Y}^*_{\ell+1}}\left\{\left(\frac{N_\ell - 1}{N_\ell^2 M_\ell}\right)^{-\frac{1}{2}}\left(\frac{\sum_{j=1}^{M_\ell} w_{ijt}}{M_\ell} - \frac{1}{N_\ell}\right)(y^*_i - \bar{y}^*_{\ell 1})\right\}, \tag{A-1}$$

the above equation for $\tilde{y}_t$ can be rewritten as

$$\tilde{y}^*_{\ell,t} = \gamma_{\ell,t} + \sqrt{\frac{N_\ell M_\ell}{(N_\ell - 1)}}\{m_{\ell 0,t}\bar{y}^*_{\ell 0} + m_{\ell 1,t}\bar{y}^*_{\ell 1} - (n_{\ell 0}\bar{y}^*_{\ell 0} + n_{\ell 1}\bar{y}^*_{\ell 1})\}$$

$$= \gamma_{\ell,t} + \sqrt{\frac{N_\ell M_\ell}{(N_\ell - 1)}}\{(1 - m_{\ell 1,t})\bar{y}^*_{\ell 0} + m_{\ell 1,t}\bar{y}^*_{\ell 1} - ((1 - n_{\ell 1})\bar{y}^*_{\ell 0} + n_{\ell 1}\bar{y}^*_{\ell 1})\}$$

$$= \gamma_{\ell,t} + \sqrt{\frac{N_\ell M_\ell}{(N_\ell - 1)}}(m_{\ell 1,t} - n_{\ell 1})(\bar{y}^*_{\ell 1} - \bar{y}^*_{\ell 0}), \tag{A-2}$$

where $m_{\ell 0,t} = M_{\ell 0,t}/M$, $M_{\ell 0,t}$ is the number of draws from $\mathcal{Y}^*_\ell$ in the t-th resampling stage). Note that $\gamma_{\ell,t} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2_{N,\ell})$, where $\sigma^2_{N,\ell} = n_{\ell 0}\sigma^2_{\ell 0} + n_{\ell 1}\sigma^2_{\ell 1}$, $\sigma^2_{\ell 0} = N_{\ell 0}^{-1}\sum_{y^*_i \in \mathcal{Y}^*_\ell}(y^*_i - \bar{y}^*_{\ell 0})^2$, and $\sigma^2_{\ell 1} =$

$N_{\ell 1}^{-1}\sum_{y_i^* \in \mathcal{Y}_{\ell+1}^*}(y_i^* - \bar{y}_{\ell 1}^*)^2$. If $\{y_i | \, i = 1, \cdots, N\}$ are *i.i.d.*, applying Proposition 1 of Ito (2023) with the finite variance assumption ($\mathrm{Var}[y_i] < \infty$), we obtain the result that $\gamma_{\ell,t} \xrightarrow{d} \mathrm{N}\big(0, n_{\ell 0}\sigma_{\ell 0}^2 + n_{\ell 1}\sigma_{\ell 1}^2\big)$ as N and therefore N_ℓ approach infinity, where $\sigma_{\ell 0}^2 = \mathrm{Var}(y_i^* | \, y_i = \ell)$ and $\sigma_{\ell 1}^2 = \mathrm{Var}(y_i^* | \, y_i = \ell + 1)$. In addition, when $\{y_i : i = 1, \cdots, N\}$ are independent and not identically distributed (*i.n.i.d*), by applying Proposition A1 of Ito (2023) for y_i^* with the additional assumption that $\lim_{N\to\infty} \sum_{i=1}^N \mathrm{E}|y_i^*|^{2+\delta} / \big(\sum_{i=1}^N \mathrm{Var}(y_i^*)\big)^{1+\delta/2} = 0$ for some $\delta > 0$, we have $\gamma_{\ell,t} \xrightarrow{d} \mathrm{N}\big(0, n_{\ell 0}\sigma_{\ell 0}^2 + n_{\ell 1}\sigma_{\ell 1}^2\big)$ as $N \to \infty$ (hence, $N_\ell \to \infty$).

The last expression in Eq. (A-2) implies that when $\tilde{y}_{\ell,t}^* > \gamma_{\ell,t}$, since $\bar{y}_{\ell 1}^* > 0 > \bar{y}_{\ell 0}^*$, we have $\big(m_{\ell 1,t} - n_{\ell 1}\big) > 0$, which means that relatively more observations are taken from $\{(y_i^*, \mathbf{x}_i) | \, y_i = \ell + 1\}$ in the t-th resampling stage than those in the sample, and therefore $\tilde{y}_{\ell,t} > 0$. On the other hand, when $\tilde{y}_{\ell,t}^* \leq \gamma_{\ell,t}$, we have $\big(m_{\ell 1,t} - n_{\ell 1}\big) \leq 0$ and $\tilde{y}_{\ell,t} \leq 0$. Therefore, the introduction of a threshold variable $\gamma_{\ell,t}$ yields Eq. (6).

A.3. Summary statistics of the simulated data in a trial

Table A1 presents summary statistics of the simulated data used in a trial in the simulation analysis conducted in Section 4.

[Table A3 around here]

References

Ahn, H. and Powell, J.L. 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1–2): 3–29.

Andrews, D.W.K. 1991. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica*, 59(2): 307–345.

Barker, M. 2014. 2014. SLS: Stata module to perform semiparametric least squares. *Statistical Software Components*, S457927, Boston: Boston College Department of Economics.

Cavanagh, C. and Sherman, R.P. 1998. Rank estimators for monotonic index models. *Journal of Econometrics*, 84(2): 351–381.

Cosslett, S.R. 1983. Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica*, 51(3): 765–782.

Das, M., Newey, W.K., and Vella, F. 2003. Nonparametric estimation of sample selection models. *Review of Economic Studies*, 70(1): 33–58.

De Luca, G. 2008. SNP and SML estimation of univariate and bivariate binary-choice models. *Stata Journal*, 8(2): 190–220.

Duncan, G.M. 1986. A semi-parametric censored regression estimator. *Journal of Econometrics*, 32(1): 5–34.

Feller, W. 1971. *An introduction to probability theory and its applications. vol. 2.* New York: John Wiley.

Fernandez, L. 1986. Non-parametric maximum likelihood estimation of censored regression models. *Journal of Econometrics*, 32(1): 35–57.

Gallant, A.R. and Nychka, D.W. 1987. Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2): 363–390.

Han, A.K. 1987. Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics*, 35(2–3): 303–316.

Härdle, W. and Stoker, T.M. 1989. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408): 986–995.

Honoré, B.E. 1992. Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, 60(3): 533–565.

Honoré, B.E. and Powell, J.L. 1994. Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, 64(2): 241–278.

Horowitz, J.L. 1986. A distribution-free least squares estimator for censored linear regression models. *Journal of Econometrics*, 32(1): 59–84.

Horowitz, J.L. 1992. A smoothed maximum score estimator for the binary response model. *Econometrica*, 60(3): 505–531.

Ichimura, H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58(1–2): 71–120.

Ichimura, H. and Lee, L.-F. 1991. Semiparametric least squares estimation of multiple index models: Single equation estimation. In: Barnett, W.A., Powell, J.L., and Tauchen, G.E. (Eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, 3–49. Cambridge, MA: Cambridge University Press.

Ito, T. 2023. Resampling-Based Maximum Likelihood Estimation. Unpublished Manuscript, Kobe University. (available at: http://dx.doi.org/10.2139/ssrn.3967838)

Klein, R.W. and Spady, R. H. 1993. An efficient semiparametric estimator for binary response models. *Econometrica*, 61(2): 387–421.

Lee, L.-F. 1982. Some approaches to the correction of selectivity bias. *Review of Economic Studies*, 49: 355–372.

Lee, L.-F. 1992. Semiparametric nonlinear least squares estimation of truncation regression models. *Econometric Theory*, 8(1): 52–94.

Lee, L.-F. 1995. Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. *Journal of Econometrics*, 65(2): 381–428.

Manski, C.F. 1975. The maximum score estimator of the stochastic utility model of choice. *Journal of Econometrics*, 3(3): 205–228.

Manski, C.F. 1985. Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27(3): 313–333.

Nawata, K. 1990. Robust estimation based on grouped-adjusted data in censored regression models. *Journal of Econometrics*, 43(3): 337–362.

Newey, W.K. 2009. Two-step series estimation of sample selection models. *Econometrics Journal*, 12(s1): S217–S229.

Newey, W.K. and Powell, J.L. 1990. Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory*, 6(3): 295–317.

Powell, J.L. 1984. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3): 303–325.

Powell, J.L. 1986a. Censored regression quantiles. *Journal of Econometrics* 32(1): 143–155.

Powell, J.L. 1986b. Symmetrically trimmed least squares estimation for Tobit models. *Econometrica*, 54(6): 1435–1460.

Powell, J., Stock, J., and Stoker, T.M. 1989. Semiparametric estimation of index coefficients. *Econometrica*, 57(6): 1403–1430.

Robinson, P.M. 1988. Root-n-consistent semiparametric regression. *Econometrica* 56(4): 931–954.

Sherman, R.P. 1993. The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61(1): 123–137.

Stoker, T.M. 1986. Consistent estimation of scaled coefficients. *Econometrica*, 54(6): 1461–1481.

Stoker, T.M. 1991. Equivalence of direct, indirect, and slope estimators of average derivatives. In: Barnett, W.A., Powell, J.L., and Tauchen, G.E. (Eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, 99–118. Cambridge, MA: Cambridge University Press.

Yatchew, A. 1997. An elementary estimator of the partial linear model. *Economics Letters*, 57: 135–143.

Yatchew, A. and Griliches, Z. 1985. Specification error in probit models. *Review of Economics and Statistics*, 67(1): 134–139.

Wooldridge, J.M. 2005. Unobserved heterogeneity and estimation of average partial effects. In: Andrews, D.W.K. and Stock, J.H. (Eds.) *Identification and inference for econometric models: Essays in Honor of Thomas Rothenberg*, 27–55. Cambridge, MA: Cambridge University Press.

Tables and Figures

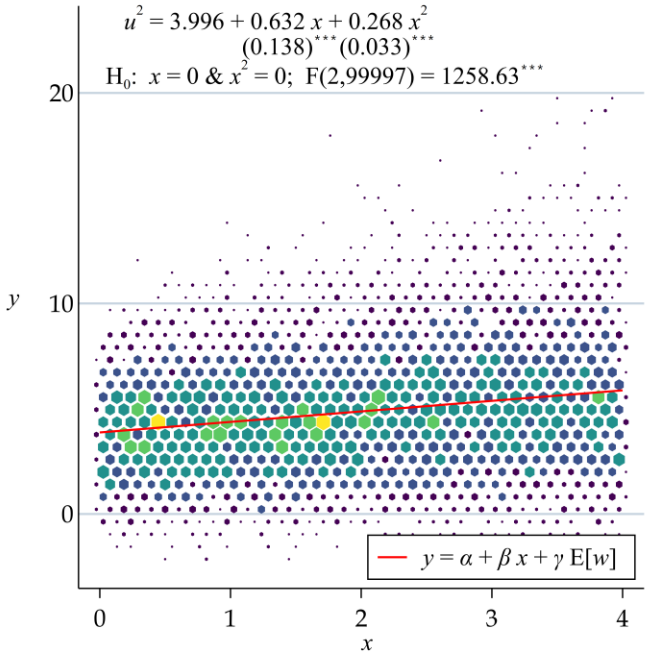
Table 1: Model description

A) Dependent variable	$y_i = 1[y_i^* < c_{i,1}] + 2[c_{i,1} \leq y_i^* < c_{i,2}]$ $+3[c_{i,2} \leq y_i^* < c_{i,3}] + 4[c_{i,3} \leq y_i^* < c_{i,4}] + 5[c_{i,4} \leq y_i^*]$
B) Cutoff points	$c_{i,1} \sim \mathrm{U}(P_5(\mathbf{y}^*), P_{15}(\mathbf{y}^*))$, $c_{i,2} \sim \mathrm{U}(P_{20}(\mathbf{y}^*), P_{30}(\mathbf{y}^*))$, $c_{i,3} \sim \mathrm{U}(P_{45}(\mathbf{y}^*), P_{55}(\mathbf{y}^*))$, and $c_{i,4} \sim \mathrm{U}(P_{75}(\mathbf{y}^*), P_{85}(\mathbf{y}^*))$
C) Latent variable	$y_i^* = \alpha_i + x_i\beta_i + w_i$
D) Explanatory variables	
(1) Binary treatment (x_i^d)	$x_i^d = 1[a_i < b]$, where $a_i \sim \mathrm{U}(0,1)$ and $b \sim \mathrm{U}(0.3,0.5)$, with $\mathrm{E}[x_i^d] = b$ and$\mathrm{Var}[x_i^d] = b(1-b)$
(2) Continuous treatment (x_i^c)	$x_i^c = 1[a_i < b] \times c_i$, where $a_i \sim \mathrm{U}(0,1)$, $b \sim \mathrm{U}(0.3,0.5)$, and $c_i \sim \mathrm{U}(0.5,1.5)$, with $\mathrm{E}[x_i^c] = b$ and$\mathrm{Var}[x_i^c] = b(13/12 - b)$
(3) Control variable (w_i)	$w_i = a_i/\sqrt{(b \cdot c)/\{(b+c)^2(b+c+1)\}}$, where $a_i \sim \mathrm{Beta}(b,c)$, $b, c \in \{1,3,5\}$, with $\mathrm{E}[w_i] = \sqrt{b(b+c+1)/\theta}$ and$\mathrm{Var}[w_i] = 1$
E) Coefficients (α_i and β_i)	
(1) Heterogenous α_i	$\alpha_i \sim \mathrm{U}(-3,3)$ with $\mathrm{E}[\alpha_i] = 0$ and$\mathrm{Var}[\alpha_i] = 3$
(2) Heterogenous β_i	$\beta_i = 0.5a_i$, where $a_i \sim \mathrm{Exp}(1)$, with $\mathrm{E}[\beta_i] = 0.5$ and$\mathrm{Var}[\beta_i] = 0.25$
F) Correlation among variables and coefficients	$\mathrm{Corr}(\mathbf{X}) = \begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & a & 1 & \\ 0 & b & c & 1 \end{pmatrix}$, where $\mathbf{X}_i = (x_i, w_i, \alpha_i, \beta_i)$, $a, b \sim \mathrm{U}(-0.2,0.2)$, and$c \sim \mathrm{U}(0,0.4)$

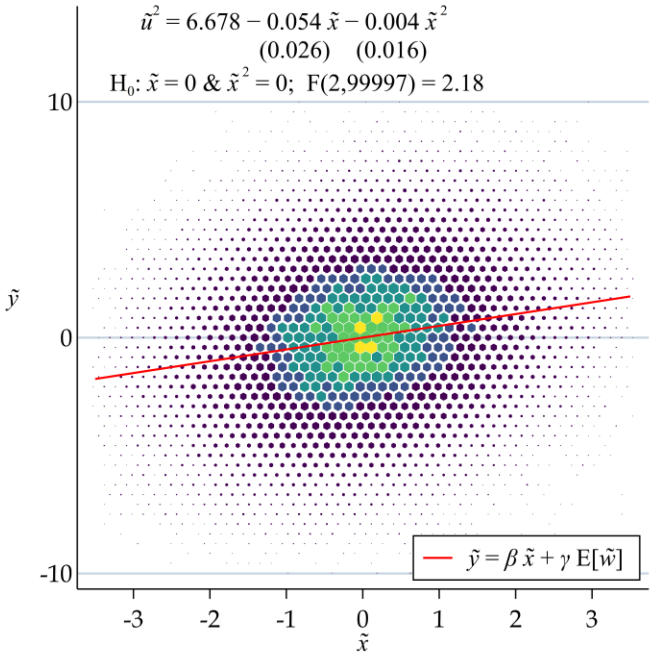
Notes: $P_k(\mathbf{y}^*)$ is the k-th percentile value of $\mathbf{y}^* = \{y_i^* \mid i = 1, \cdots, N\}$. $U(a,b)$ is the continuous uniform distribution in the interval $[a,b]$. $\mathrm{Beta}(a,b)$ is the beta distribution with parameters a and b. $\chi^2(a)$ is the chi-square distribution with a degrees of freedom. $\mathrm{Exp}(a)$ is the exponential distribution with parameter $a > 0$. $\mathrm{U}(a, a+1, \cdots, b-1, b)$, where a and b are integers and $a < b$, is the discrete uniform distribution from a to b.

Table 2: Simulation results for the root mean square relative errors of the estimates

	(1)	(2)	(3)	(4)
The treatment variable z:	Binary		Continuous	
Heterogenous α_i (nonnormal error)	Yes	Yes	Yes	Yes
Heterogenous β_i (heteroscedasticity)	No	Yes	No	Yes
A) Based on the binary response model (y_i is converted into a binary outcome, $d_i = 1[y_i \geq 4]$)				
ANML (M=T=100,000)	0.497	0.526	0.451	0.487
LPM (OLS)	0.507	0.546	0.453	0.494
Probit	0.507	0.546	0.455	0.496
Sieve ML (Gallant and Nychka, 1987)	0.511	0.565	0.462	0.507
Kernel ML (Klein and Spady, 1993)	0.570	0.622	0.511	0.580
SLS (Ichimura, 1994)	0.586	0.646	0.552	0.601
B) Based on the ordered response model ($y_i \in \{1,2,3,4,5\}$)				
ANML (M=T=100,000)	0.379	0.397	0.371	0.406
Ordered probit	0.467	0.489	0.429	0.467
Sieve ML (Gallant and Nychka, 1987)	0.391	0.423	0.358	0.415



A) Original data



B) Data generated through the RBML method

Figure 1: Data distribution and heteroscedasticity

Notes: Panel A uses the simulated data based on the same design as in Section 4, and Panel B uses the data generated in the RBML estimation process (i.e., the first step described in the text) from the data used in Panel A. Larger and brighter hexagons indicate higher observation frequencies. At the top of each panel, the result of regressing the errors (measured as the distance from the regression line) on the variable of interest and its square term is reported.

Table A1: Summary statistics of the simulated data in a trial

Variable	Obs.	Mean	Std. Dev.	Min	Max
Outcome variables: $y_i \sim \mathrm{Uniform}\{1,5\}$ and $d_i = 1[y_i \geq 4]$					
[Design 1] y_i	500	3.372	1.210	1.000	5.000
d_i	500	0.500	0.501	0.000	1.000
[Design 2] y_i	500	3.334	1.214	1.000	5.000
d_i	500	0.480	0.500	0.000	1.000
Latent variables					
[Design 1] y_i^*	500	3.557	1.959	−1.044	9.074
[Design 2] y_i^*	500	3.569	2.043	−1.044	10.335
Explanatory variables					
x_i^d (binary treatment)	500	0.444	0.471	0.000	1.000
x_i^c (continuous treatment)	500	0.169	0.295	0.000	1.499
w_i	500	3.319	0.959	0.463	5.855
Error/Unobservables					
α_i (uniform distribution)	500	0.153	1.743	−2.995	2.995
β_i (exponential distribution)	500	0.499	0.501	0.001	3.956
Cutoff variables					
$c_{i,1}$	500	0.794	0.284	0.300	1.276
$c_{i,2}$	500	1.944	0.209	1.588	2.312
$c_{i,3}$	500	3.633	0.215	3.253	4.002
$c_{i,4}$	500	5.348	0.180	5.032	5.657