

Article

Estimating the Complier Average Causal Effect with Non-Ignorable Missing Outcomes Using Likelihood Analysis

Jierui Du ¹, Gao Wen ¹ and Xin Liang ^{2,*}

¹ School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China; dujierui@gzhu.edu.cn (J.D.); gaowenjasj@163.com (G.W.)

² School of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, China

* Correspondence: liangxin@gxnu.edu.cn

Abstract: Missing data problems arise in randomized trials, which complicates the inference of causal effects if the missing mechanism is non-ignorable. We tackle the challenge of identifying and estimating the complier average causal effect parameters under non-ignorable missingness by increasing the covariates to mitigate the sensitivity to the violation of specific identification assumptions. The missing data mechanism is assumed to follow a logistic model, wherein the absence of the outcome is explained by the outcome itself, the treatment received, and the covariates. We establish the identifiability of the models under mild conditions by assuming that the outcome follows a normal distribution. We develop a computational method to estimate model parameters through a two-step likelihood estimation approach, employing subgroup analysis. The bootstrap method is employed for variance estimation, and the effectiveness of our approach is confirmed through simulation. We applied the proposed method to analyze the household income dataset from the Chinese Household Income Project Survey 2013.

Keywords: complier average causal effect; non-ignorable missingness; identifiability; likelihood; subgroup analysis



Citation: Du, J.; Wen, G.; Liang, X. Estimating the Complier Average Causal Effect with Non-Ignorable Missing Outcomes Using Likelihood Analysis. *Mathematics* **2024**, *12*, 1300. <https://doi.org/10.3390/math12091300>

Academic Editor: Manuel Alberto M. Ferreira

Received: 1 April 2024

Revised: 22 April 2024

Accepted: 23 April 2024

Published: 25 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Non-compliance and missing data are phenomena which usually occur in studies of economics, medicine, and public health. Non-compliance arises when certain participants do not adhere to their prescribed treatments, while the problem of missing data arises when the study's researchers are unable to gather information for some participants [1]. These unmeasured confounding variables may complicate the inference of causal effects if the missing mechanism is non-ignorable [2,3]. The missingness is named ignorable if it depends on the observed data only; otherwise, it is named non-ignorable [4,5]. Identifying the complier average causal effect becomes challenging in the presence of both non-compliance and non-ignorable missing values, as it is impossible to identify the full data distribution or the causal effect without additional assumptions.

To address non-compliance, one can assume that the response in the dataset follows a specific distribution, such as exponential families, and use maximum likelihood estimation. The authors in [6–8] presented the estimation of the complier average causal effect (CACE) with the maximum likelihood estimation method with the EM algorithm [9]. The maximum likelihood method provides the advantage of relaxing the exclusion restrictions, which can often be unrealistic, particularly in natural experiment scenarios [10]. Another approach to addressing non-compliance is the instrumental variable method [11]. Utilizing instrumental variables is a viable method when stringent assumptions about the response distribution could result in mis-specification.

The shadow variable strategy is a mainstream solution for establishing identifiability in cases of non-ignorable missingness [12]. Analogous approaches also employ instrumental variables, as suggested by [13]. Nonetheless, the selection of appropriate instrumental or shadow variables poses challenges, particularly among a multitude of covariates [14]. The authors of [15] illustrate that stronger assumptions regarding the response mechanism enable the establishment of identifiability derived from the distribution of the observed data. When the response adheres to exponential families, as noted by [16,17], the identifiability under non-ignorable missingness is readily attainable without the use of instrumental variables.

It is critically important to integrate causal inference with research on non-ignorable missing data [18,19]. The authors of [20–22] introduced estimators for the complier average causal effect parameters, assuming that the missing data are ignorable. However, if the missingness is non-ignorable, these methods may yield biased estimators. The recent literature has introduced strategies for addressing missing covariates or outcomes in the context of non-ignorable missingness, which includes but is not limited to the following papers. The study in [1] initially explored the identifiability of parameters in randomized clinical trials characterized by non-compliance and non-ignorable missing binary outcome variables. The study in [2] examined semi-parametric identifiability and developed an estimation method for the complier average causal effect in randomized clinical trials, addressing the challenge of non-ignorable missingness in continuous outcomes. The study in [23] formulated a method using a shadow variable to identify and estimate the complier average causal effect parameters arising from outcomes that represent non-ignorable missingness. The study in [19] tackled the challenge of non-ignorable missingness confounders in causal analysis from observational data, demonstrating that causal effects are ascertainable in scenarios wherein the missingness mechanism is outcome-independent, conditional on the treatment and potential missing confounders. The study in [24] developed semi-parametric estimators to determine the average causal effect, accounting for non-ignorable missing confounders, based on the premise that the missingness of data is independent of the outcome. The study in [25] addressed the problem of identifying the treatment benefit rate and treatment harm rate in scenarios wherein treatment, endpoints, or covariates are missing. The study in [26] investigates a treatment-independent missingness assumption, which facilitates the identification of causal effects in situations wherein confounders represent non-ignorable missingness.

In the existing literature, limited attention has been given to situations wherein a general missing mechanism model is selected to address non-ignorable missingness. The study in [2] utilized an outcome-dependent non-ignorable missingness model without incorporating an auxiliary variable. The study in [3] proposed using an auxiliary variable that serves as both a shadow variable for non-ignorable missingness and an instrumental variable for causal effects. However, the models used in the above approach assume the absence of covariates, despite their common presence in practical applied research. In this paper, we examine the complier average causal effect parameters under non-ignorable missingness by including covariates to mitigate the sensitivity to the violation of specific identification assumptions. The missing data mechanism is assumed to follow a logistic model, wherein the absence of the outcome is explained by the outcome itself, the treatment received, and the covariates. We establish the identifiability of the models under mild conditions by assuming that the outcome follows a normal distribution. We develop a computational method to estimate model parameters through a two-step likelihood estimation approach, employing subgroup analysis. The bootstrap method is employed for variance estimation, and the effectiveness of our approach is confirmed through simulation. The proposed method is applied to analyze the household income dataset from the Chinese Household Income Project Survey 2013.

The rest of this article is organized as follows. Section 2 introduces the general modeling framework, with notation and assumptions. In Section 3, we give the theoretical results on the identifiability of the parameters and the estimation approach. The performance of

the proposed method is evaluated through simulation studies in Section 4. The application to CHIP data is presented in Section 5. Concluding remarks are provided in Section 6.

2. Notation and Assumptions

Let Y_i denote the individual outcome and let \mathbf{X}_i represent the individual covariate vector, where $Y_i \in \mathbb{R}$, $\mathbf{X} \equiv (1, X_{1,i}, \dots, X_{p,i})^\top \in \mathbb{R}^{p+1}$. Z_i denotes the randomized treatment assignment for the i th unit in the study, where i ranges from 1 to n . We assign $Z_i = 1$ if the i th individual is allocated to the treatment group and $Z_i = 0$ if it is assigned to the control group. Moreover, D_i represents the treatment received by the i th individual, with $D_i = 1$ indicating treatment and $D_i = 0$ indicating no treatment. $D(z)$ and $Y(z)$ denote the potential treatment received and the potential outcome under the assigned treatment $Z = z$, respectively. We define $R(z)$ as the binary response indicator for $Y(z)$, where $R(z) = 1$ if $Y(z)$ is observed and $R(z) = 0$ if $Y(z)$ is missing. Additionally, we define $Y(z, d)$ as representing the potential outcome under the assigned treatment $Z = z$ and actual treatment $D = d$.

Following [6], we define U_i as the compliance status of the i th patient, which is determined as follows:

$$U_i = \begin{cases} c & \text{if } D_i(0) = 0 \text{ and } D_i(1) = 1 \\ n & \text{if } D_i(0) = 0 \text{ and } D_i(1) = 0 \\ a & \text{if } D_i(0) = 1 \text{ and } D_i(1) = 1 \\ d & \text{if } D_i(0) = 1 \text{ and } D_i(1) = 0 \end{cases}$$

where the potential intermediate outcomes c, n, a , and d stand for complier, never-taker, always-taker, and defier, respectively. The complier average causal effect gives the predictors $\mathbf{X} = \mathbf{x}$, which equals

$$\text{CACE}(\mathbf{x}) = E\{Y(1) - Y(0)|U = c, \mathbf{X} = \mathbf{x}\}$$

Next, we provide the necessary assumptions to ensure the identifiability of $\text{CACE}(\mathbf{x})$ under the non-ignorable missingness. These are formalized by the following assumptions [1,11]:

Assumption 1 (Stable unit treatment value assumption, SUTVA). *If $z_i = z'_i$, then $D_i(z) = D_i(z')$. And, if $z_i = z'_i, d_i = d'_i$, then $Y_i(z, d) = Y_i(z', d')$.*

Assumption 2 (Randomization). *The treatment assignment Z is randomization.*

Assumption 3 (Exclusion restrictions). *$P(Y(1, d) = Y(0, d)|\mathbf{X} = \mathbf{x}) = 1$ for $d \in \{0, 1\}$.*

Assumption 4 (First-stage). *Given $\mathbf{X} = \mathbf{x}$, there is a non-zero average causal effect of Z on D , i.e., $E\{D(1)|\mathbf{X} = \mathbf{x}\} \neq E\{D(0)|\mathbf{X} = \mathbf{x}\}$.*

Assumption 5 (Monotonicity). *$D_i(1) \geq D_i(0)$ for all $i = 1, \dots, N$. Defining $\omega_u = P(U = u)$ where $u \in \{c, n, a, d\}$, we then have $\omega_d = 0$.*

Assumption 6 (Compound exclusion restrictions). *$P[\{Y(1, d), R(1)\} = \{Y(0, d), R(0)\}|\mathbf{X} = \mathbf{x}] = 1$ for $d \in \{0, 1\}$.*

Assumption 7 (Non-ignorable missingness). *$E\{R(z)|Y(z) = y, D(z) = d, U = u, \mathbf{X} = \mathbf{x}\} = E\{R(z)|Y(z) = y, D(z) = d, \mathbf{X} = \mathbf{x}\}$ and $E\{R(1)|Y(1) = y, D(1) = d, \mathbf{X} = \mathbf{x}\} = E\{R(0)|Y(0) = y, D(0) = d, \mathbf{X} = \mathbf{x}\}$.*

Assumption 8 (Subgroup). *U is independent of variable \mathbf{X} .*

Assumption 9 (Normal). *The conditional density of the outcome variable Y has the following normal form:*

$$\begin{aligned} p(y|\mathbf{x}; \theta_{zu}) &= p(y|Z=z, U=u, \mathbf{X}=\mathbf{x}; \beta_{zu}, \sigma_{zu}^2) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{zu}} \exp \left\{ -\frac{(y - \beta_{zu}^\top \mathbf{x})^2}{2\sigma_{zu}^2} \right\} \end{aligned}$$

where $u \in \{c, n, a\}$ is the compliance status and $\theta_{zu} = (\beta_{zu}^\top, \sigma_{zu}^2)^\top$.

Assumption 1 posits that $Y_i = D_i Y_i(z, 1) + (1 - D_i) Y_i(z, 0)$ and $D_i = Z_i D(1) + (1 - Z_i) D(0)$, indicating that the observed outcome equals the potential outcome evaluated at the observed treatment value, and the observed treatment equals the potential treatment evaluated at the assigned treatment [27]. This assumption is typically reasonable when dealing with randomly sampled units. Assumption 2 [1] means that Z is as good as randomly assigned. Assumption 3 asserts that, given \mathbf{X} , $Y(d) = Y(1, d) = Y(0, d)$. This concept encapsulates the fundamental principle of instrumental variable procedures, indicating that any influence of Z on Y must be channeled through an impact of Z on D [28]. Under Assumption 3, with \mathbf{X} fixed, the conditional density of Y remains independent of Z for never-takers and always-takers. Assumption 4 stipulates that, for every stratum defined by $\mathbf{X} = \mathbf{x}$, ω_c is strictly positive. Assumption 5 asserts the absence of defiers within the population. Assumptions 1–5 are standard assumptions in causal effect models.

Assumption 6 [2] is credible in a double-blinded clinical trial, as patients are unaware of the treatment assigned to them, and the treatment assignment is performed through randomization. Assumption 7 suggests that the absence of an outcome variable is accounted for by the outcome itself, the treatment received, and the covariates. It implies that the expectation of $R(z)$ given $Y(z) = y, D(z) = d, U = u$, and $\mathbf{X} = \mathbf{x}$ is equal to the expectation of observing R given $Y = y, D = d$, and $X = x$. The missing data mechanism is assumed to follow a logistic model, expressed as $E(R|Y = y, D = d, \mathbf{X} = \mathbf{x}) = E_d(R|Y = y, \mathbf{X} = \mathbf{x}) = \text{logit}_d^{-1}(\alpha_d y + \boldsymbol{\phi}_d^\top \mathbf{x}) = \pi_d(\alpha, \boldsymbol{\phi})$, $d \in \{0, 1\}$, $\text{logit}^{-1}(\cdot) \equiv \exp(\cdot)/\{1 + \exp(\cdot)\}$. Assumption 8 incorporates the independence between X and U into the computational method for estimating model parameters. This inclusion enhances the robustness and validity of the estimation procedure, particularly through a two-step maximum likelihood estimation approach using subgroup analysis. The purpose of Assumption 9 is to establish the identifiability of the models under mild conditions, facilitating the creation of a computational method to estimate model parameters.

3. Identifiability and Estimation

In this section, we initially construct a model for non-compliant data and missing data, outlining the steps to solve the model. Subsequently, we theoretically establish the identifiability of the model involved in each step. Given that the second step entails a system of non-linear equations and integral operations, which can be challenging to handle analytically, we present a numerical calculation method to address this complexity.

We introduce a convenient two-stage estimation process based on the random sample $(R_i, D_i, Z_i, \mathbf{X}_i, Y_i), i = 1, \dots, N$. Given $\mathbf{X} = \mathbf{x}$, the joint density function with respect to variables (y, z, u) can be expressed as follows:

$$f(y, z, u|\mathbf{x}) = f(z|\mathbf{x})f(u|z, \mathbf{x})f(y|z, u, \mathbf{x}) = f(z)f(u|z)f(y|z, u, \mathbf{x})$$

Assumptions 2 and 8 ensure the validity of the equation above. Let $\xi = P(Z = 1), \omega_a = P(U = a) = P(D = 1|Z = 0), \omega_n = P(U = n) = P(D = 0|Z = 1), \omega_c = 1 - \omega_n - \omega_a$, and denote $\boldsymbol{\eta} = (\xi, \omega_a, \omega_n)^\top$. Assumption 3 implies that $f(y|\mathbf{x}; \theta_{1a}) = f(y|\mathbf{x}; \theta_{0a})$, denoted as $f(y|\mathbf{x}; \theta_a)$, and $f(y|\mathbf{x}; \theta_{1n}) = f(y|\mathbf{x}; \theta_{0n})$, and denoted as $f(y|\mathbf{x}; \theta_n)$. Let $n_z =$

$\#\{i : Z_i = z\}, n_{zd} = \#\{i : Z_i = z, D_i = d\}$ for $z \in \{0, 1\}$ and $d \in \{0, 1\}$. As stated in [2], the full likelihood for $(\boldsymbol{\eta}, \alpha, \boldsymbol{\phi}, \boldsymbol{\theta})$ can be expressed as

$$L(\boldsymbol{\eta}, \alpha, \boldsymbol{\phi}, \boldsymbol{\theta}) \propto L(\boldsymbol{\eta}) L(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a, \boldsymbol{\theta}_{1c}) L(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) L(\alpha_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_n, \boldsymbol{\theta}_{0c}) L(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_n) \quad (1)$$

where

$$L(\boldsymbol{\eta}) = \xi^{n_1} (1 - \xi)^{n_0} (1 - \omega_n)^{n_{11}} \omega_n^{n_{10}} \omega_a^{n_{01}} (1 - \omega_a)^{n_{00}} \quad (2)$$

$$\begin{aligned} & L(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a, \boldsymbol{\theta}_{1c}) \\ &= \prod_{i:(Z_i, D_i)=(1,1)} \left[\pi_1(\alpha_1, \boldsymbol{\phi}_1) \left\{ \frac{\omega_a}{\omega_a + \omega_c} p(Y_i | \mathbf{X}_i; \boldsymbol{\theta}_a) + \frac{\omega_c}{\omega_a + \omega_c} p(Y_i | \mathbf{X}_i; \boldsymbol{\theta}_{1c}) \right\} \right]^{R_i} \\ & \quad \left[\int \{1 - \pi_1(\alpha_1, \boldsymbol{\phi}_1)\} \left\{ \frac{\omega_a}{\omega_a + \omega_c} p(y | \mathbf{X}_i; \boldsymbol{\theta}_a) + \frac{\omega_c}{\omega_a + \omega_c} p(y | \mathbf{X}_i; \boldsymbol{\theta}_{1c}) \right\} dy \right]^{1-R_i} \end{aligned} \quad (3)$$

$$\begin{aligned} & L(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) \\ &= \prod_{i:(Z_i, D_i)=(0,1)} \left\{ \pi_1(\alpha_1, \boldsymbol{\phi}_1) p(Y_i | \mathbf{X}_i; \boldsymbol{\theta}_a) \right\}^{R_i} \left\{ \int \{1 - \pi_1(\alpha_1, \boldsymbol{\phi}_1)\} p(y | \mathbf{X}_i; \boldsymbol{\theta}_a) dy \right\}^{1-R_i} \end{aligned} \quad (4)$$

$$\begin{aligned} & L(\alpha_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_n) \\ &= \prod_{i:(Z_i, D_i)=(1,0)} \left\{ \pi_1(\alpha_0, \boldsymbol{\phi}_0) p(Y_i | \mathbf{X}_i; \boldsymbol{\theta}_n) \right\}^{R_i} \left\{ \int \{1 - \pi_1(\alpha_0, \boldsymbol{\phi}_0)\} p(y | \mathbf{X}_i; \boldsymbol{\theta}_n) dy \right\}^{1-R_i} \end{aligned} \quad (5)$$

$$\begin{aligned} & L(\alpha_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_n, \boldsymbol{\theta}_{0c}) \\ &= \prod_{i:(Z_i, D_i)=(0,0)} \left[\pi_1(\alpha_0, \boldsymbol{\phi}_0) \left\{ \frac{\omega_n}{\omega_n + \omega_c} p(Y_i | \mathbf{X}_i; \boldsymbol{\theta}_n) + \frac{\omega_c}{\omega_n + \omega_c} p(Y_i | \mathbf{X}_i; \boldsymbol{\theta}_{0c}) \right\} \right]^{R_i} \\ & \quad \left[\int \{1 - \pi_1(\alpha_0, \boldsymbol{\phi}_0)\} \left\{ \frac{\omega_n}{\omega_n + \omega_c} p(y | \mathbf{X}_i; \boldsymbol{\theta}_n) + \frac{\omega_c}{\omega_n + \omega_c} p(y | \mathbf{X}_i; \boldsymbol{\theta}_{0c}) \right\} dy \right]^{1-R_i} \end{aligned} \quad (6)$$

It is evident from (2) that the maximum likelihood estimator for $\boldsymbol{\eta}$ is $(\hat{\xi}, \hat{\omega}_a, \hat{\omega}_n) = (n_1/n, n_{01}/n_1, n_{10}/n_0)$, which is equivalent to the moment estimator, thus ensuring its identifiability. Next, we propose a two-stage estimation process $(\alpha_d, \boldsymbol{\phi}_d, \boldsymbol{\theta}_{zu})$ and establish the identifiability of the models involved in these two steps.

Step 1. Based on Assumptions 3 and 7, the conditional density of the outcome Y for never-takers and always-takers, as well as the missing data models, are independent of Z . The likelihood function computed on never-takers with $(Z_i, D_i) = (0, 1)$ is (4) and the likelihood function computed on always-takers with $(Z_i, D_i) = (1, 0)$ is (5). We can maximize (4) and (5) to obtain that the maximum likelihood estimators for $(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a)$ and $(\alpha_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_n)$ are $(\hat{\alpha}_1, \hat{\boldsymbol{\phi}}_1, \hat{\boldsymbol{\theta}}_a)$ and $(\hat{\alpha}_0, \hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\theta}}_n)$.

Step 2. We substitute $(\hat{\xi}, \hat{\omega}_a, \hat{\omega}_n)$, $(\hat{\alpha}_1, \hat{\boldsymbol{\phi}}_1, \hat{\boldsymbol{\theta}}_a)$, and $(\hat{\alpha}_0, \hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\theta}}_n)$ into (3) and (6) and consider the numerical characteristics of the mixed normal distribution and the inverse probability weighting method (more details can be found in [29]) to obtain the estimators of $\boldsymbol{\theta}_{1c}$ and $\boldsymbol{\theta}_{0c}$, denoted as $\hat{\boldsymbol{\theta}}_{1c}$ and $\hat{\boldsymbol{\theta}}_{0c}$.

Non-ignorable missingness in Y presents challenges to the identifiability of the observed likelihood function (4) and (5), as emphasized by [12]. The next theorem shows that the parameters are all identifiable under mild conditions.

Theorem 1. Suppose there exists one continuous covariate; if the sign of some element of $(\alpha_d, \boldsymbol{\phi}_d^\top)^\top$ is known, then the vector of parameters $(\alpha_d, \boldsymbol{\phi}_d, \boldsymbol{\theta}_{zu})$ is identifiable, where $d \in \{0, 1\}$ and $zu \in \{1c, 0c, n, a\}$.

Proof. (i) First, we establish the identifiability of parameters $(\alpha_1, \phi_1, \theta_a)$. Suppose there exist two sets of parameters $(\alpha_1, \phi_1, \theta_a)$ and $(\alpha_1^*, \phi_1^*, \theta_a^*)$ such that

$$\log \pi(y, \mathbf{x}; \alpha_1, \phi_1) + \log p(y|\mathbf{x}, \theta_a) = \log \pi(y, \mathbf{x}; \alpha_1^*, \phi_1^*) + \log p(y|\mathbf{x}; \theta_a^*) \quad (7)$$

which holds for all (y, \mathbf{x}) . The authors in [12] accounted that $(\alpha_1, \phi_1, \theta_a)$ is identifiable if (7) implies that

$$\alpha = \alpha^*, \quad \phi_1 = \phi_1^*, \quad \theta_a = \theta_a^*$$

Without a loss of generality, suppose X_1 is a continuous variable and can take any real values, and let covariates X_2, \dots, X_p be omitted since we can consider X_2, \dots, X_p as fixed, while X_1 varies. Due to the fact that the proof for the case $\phi_1 = 0$ can be obtained by mimicking the following process, here, we consider the case $\phi_1 \neq 0$. Let $g(\cdot) = \log \pi(\cdot)$; then, $g(\alpha_1 y + \phi_1^\top \mathbf{x}) + \log p(y|\mathbf{x}; \theta_a) = g(\alpha_1^* y + \phi_1^{*\top} \mathbf{x}) + \log p(y|\mathbf{x}; \theta_a^*)$ can be transformed into the following form:

$$\begin{aligned} & g(\alpha_1 y + \phi_{1,0} + \phi_{1,1} x_1) + \log p(y|x_1; \beta_{a,0}, \beta_{a,1}, \sigma_a^2) \\ &= g(\alpha_1^* y + \phi_{1,0}^* + \phi_{1,1}^* x_1) + \log p(y|x_1; \beta_{a,0}^*, \beta_{a,1}^*, \sigma_a^{2*}) \end{aligned} \quad (8)$$

Applying operation $\partial/\partial x_1$ on both sides of (8) yields

$$\begin{aligned} & g'(\alpha_1 y + \phi_{1,0} + \phi_{1,1} x_1) \phi_1 + \frac{y - \beta_{a,0} - \beta_{a,1} x_1}{\sigma^2} \beta_{a,1} \\ &= g'(\alpha_1^* y + \phi_{1,0}^* + \phi_{1,1}^* x_1) \phi_1^* + \frac{y - \beta_{a,0}^* - \beta_{a,1}^* x_1}{\sigma^{2*}} \beta_{a,1}^* \end{aligned} \quad (9)$$

Applying operation $\partial^2/\partial y^2$ on both sides of (9) yields

$$g^{(3)}(\alpha_1 y + \phi_{1,0} + \phi_{1,1} x_1) \phi_{1,1} \alpha_1^2 = g^{(3)}(\alpha_1^* y + \phi_{1,0}^* + \phi_{1,1}^* x_1) \phi_{a,1}^* \alpha_1^{*2}. \quad (10)$$

If $\phi_{1,1}^* = 0$ or $\alpha_1^* = 0$, then (10) reduces to $g_i^{(3)}(\alpha_1 y + \phi_{1,0} + \phi_{1,1} x_1) \phi_{1,1} \alpha_1^2 = 0$ and then $\phi_{1,1} \alpha_1^2 = 0$, which contradicts the assumption $\phi_{1,1} \neq 0$ and $\alpha_1 \neq 0$. Now, we consider the case $\phi_{1,1}^* \neq 0$ and $\alpha_1^* \neq 0$. Due to the fact that $g(\cdot) = \log \text{logit}^{-1}(\cdot)$, the roots of the derivatives of $g(t)$ are as follows:

$$\left\{ \begin{array}{l} g'(t) \text{ has no roots} \\ g''(t) \text{ has no roots} \\ g^{(3)}(t) \text{ has one root } 0 \\ g^{(4)}(t) \text{ has two roots } \log(2 + \sqrt{3}), \log(2 - \sqrt{3}) \\ g^{(5)}(t) \text{ has three roots } 0, \log(5 + 2\sqrt{6}), \log(5 - 2\sqrt{6}) \end{array} \right.$$

If $\phi_{1,1} \alpha_1^* \neq \phi_{1,1}^* \alpha_1$, then this leads to a contradiction. Assume the line $\alpha_1 y + \phi_{1,0} + \phi_{1,1} x_1$ intersects with $\alpha_1^* y + \phi_{1,0}^* + \phi_{1,1}^* x_1$ at (\dot{x}_1, \dot{y}) ; then, we have $\alpha_1 \dot{y} + \phi_{1,0} + \phi_{1,1} \dot{x}_1 = \alpha_1^* \dot{y} + \phi_{1,0}^* + \phi_{1,1}^* \dot{x}_1 = t$. If $t = 0$, by applying operation $\partial/\partial y$ and $\partial/\partial x_1$ on both sides of (10) at (\dot{x}_1, \dot{y}) , we have $\phi_{1,1} \alpha_1^3 = \phi_{1,1}^* \alpha_1^{*3}$ and $\phi_{1,1}^2 \alpha_1^2 = \phi_{1,1}^{*2} \alpha_1^{*2}$, then $\phi_{1,1} \alpha_1^* = \phi_{1,1}^* \alpha_1$, which is a contradiction. If t is equal to $\log(2 + \sqrt{3})$ or $\log(2 - \sqrt{3})$, by (10) and by applying operation $\partial^2/\partial x_1 \partial y$ on both sides of (10) at (\dot{x}_1, \dot{y}) and $\phi_{1,1} \alpha_1^2 = \phi_{1,1}^* \alpha_1^{*2}$ and $\phi_{1,1}^2 \alpha_1^3 = \phi_{1,1}^{*2} \alpha_1^{*3}$, then $\phi_{1,1} \alpha_1^* = \phi_{1,1}^* \alpha_1$, which is a contradiction. If $t \neq b, b \in \{0, \log(2 + \sqrt{3}), \log(2 - \sqrt{3})\}$, by (10) and by applying operation $\partial/\partial y$ on both sides of (10) at (\dot{x}_1, \dot{y}) , then $\phi_{1,1} \alpha_1^2 = \phi_{1,1}^* \alpha_1^{*2}$ and $\phi_{1,1} \alpha_1^3 = \phi_{1,1}^* \alpha_1^{*3}$, which means that $\{\phi_{1,1} = \phi_{1,1}^*, \alpha_1 = \alpha_1^*\}$; then, $\phi_{1,1} \alpha_1^* = \phi_{1,1}^* \alpha_1$, which is a contradiction. Thus, $\phi_{1,1} \alpha_1^* = \phi_{1,1}^* \alpha_1$.

Let $\phi_{1,1}/\phi_{1,1}^* = \alpha/\alpha^* = k$, and (10) reduces to

$$g^{(3)}(ks + \phi_{1,0})k^3 = g^{(3)}(s + \phi_{1,0}^*) \quad (11)$$

with $s = \alpha_1^*y + \phi_{1,1}^*x_1$. If $k \neq 1$, assume the line $ks + \phi_{1,0}$ intersects with $s + \phi_{1,0}^*$ at \dot{s} ; then, $k\dot{s} + \phi_{1,0} = \dot{s} + \phi_{1,0}^* = t$. If $t \neq 0$, let $s = \dot{s}$ in (11); then, $k = 1$, which is a contradiction. If $t = 0$, by applying operation $\partial/\partial s$ on both sides of (11) at \dot{s} , we have $k^4 = 1$, that is, $k = 1$ or $k = -1$. If $k = -1$, then $\alpha_1 = -\alpha_1^*, \phi_{1,1} = -\phi_{1,1}^*$, and $-\dot{s} + \phi_{1,0} = \dot{s} + \phi_{1,0}^* = 0$, which means that $\phi_{1,0} = -\phi_{1,0}^*$, that is, $(\alpha_1^*, \phi_{1,0}^*, \phi_{1,1}^*) = -(\alpha_1, \phi_{1,0}, \phi_{1,1})$. Recall the condition that the sign of any element of $(\alpha_1, \phi_{1,0}, \phi_{1,1})$ is assumed to be known. Therefore, the case according to which $t = 0$ and $k = -1$ is impossible. Therefore, we have $k = 1$; then, (11) reduces to

$$g^{(3)}(t + \phi_{1,0}) = g^{(3)}(t + \phi_{1,0}^*)$$

and we have $\phi_{1,0} = \phi_{1,0}^*$ because $g^{(3)}(\cdot)$ has only one maximum point. Now, we have $(\alpha_1^*, \phi_{1,0}^*, \phi_{1,1}^*) = (\alpha_1, \phi_{1,0}, \phi_{1,1})$, and the condition in (8) reduces to

$$\log p(y|x_1; \beta_{a,0}, \beta_{a,1}, \sigma_a^2) = \log p(y|x_1; \beta_{a,0}^*, \beta_{a,1}^*, \sigma_a^{2*}) \quad (12)$$

We can readily obtain $(\beta_{a,0}, \beta_{a,1}, \sigma_a^2) = (\beta_{a,0}^*, \beta_{a,0}^*, \sigma_a^{2*})$ from (12). Thus, $(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a)$ is identifiable. In an analogous manner to the above, $(\alpha_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_n)$ is identifiable.

(ii) Similar to the proof of the above, we now only need to demonstrate the following: if

$$g(\alpha_1, \boldsymbol{\phi}_1) + \log \left\{ \frac{\omega_a}{\omega_a + \omega_c} p(y|\mathbf{x}; \boldsymbol{\theta}_a) + \frac{\omega_c}{\omega_a + \omega_c} p(y|\mathbf{x}; \boldsymbol{\theta}_{1c}) \right\} \quad (13)$$

$$= g(\alpha_1^*, \boldsymbol{\phi}_1^*) + \log \left\{ \frac{\omega_a}{\omega_a + \omega_c} p(y|\mathbf{x}; \boldsymbol{\theta}_a^*) + \frac{\omega_c}{\omega_a + \omega_c} p(y|\mathbf{x}; \boldsymbol{\theta}_{1c}^*) \right\} \quad (14)$$

we can obtain $\boldsymbol{\theta}_{1c} = \boldsymbol{\theta}_{1c}^*$, and then the $\boldsymbol{\theta}_{1c}$ is identifiable. We now substitute $(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) = (\alpha_1^*, \boldsymbol{\phi}_1^*, \boldsymbol{\theta}_a^*)$ into (13); then, (13) reduces to

$$p(y|\mathbf{x}; \boldsymbol{\theta}_{1c}) = p(y|\mathbf{x}; \boldsymbol{\theta}_{1c}^*) \quad (15)$$

and we can readily obtain $\boldsymbol{\theta}_{1c} = \boldsymbol{\theta}_{1c}^*$; then, $\boldsymbol{\theta}_{1c}$ is identifiable. Likewise, $\boldsymbol{\theta}_{0c}$ is identifiable. Therefore, the conclusion of Theorem 1 follows. \square

The sign of some element of $(\alpha_d, \boldsymbol{\phi}_d^\top)^\top$ is easy to verify. According to [30], factors such as respondents' cognitive level, motivation, and social status influence non-response probability. Leveraging this insight, we can speculate on the trend of non-response probability and infer the sign of the parameters in the missing mechanism model. For example, in a household income survey, low-income individuals may be less inclined to disclose their true income, implying $\alpha > 0$.

Given the similarity between the methods for estimating $(\alpha_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_n)$ and $\boldsymbol{\theta}_{0c}$ and those for estimating $\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_{1c}$, we will delineate the estimation methods for estimating $(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a)$ and $\boldsymbol{\theta}_{1c}$ separately in two steps. The likelihood function computed on always-takers with $(Z_i, D_i) = (1, 0)$ is (4). Thus, the maximization problem in (4) with respect to parameters $(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a)$ is equivalent to finding $\alpha_1, \boldsymbol{\phi}_1$, and $\boldsymbol{\theta}_a$ to maximize

$$l_n(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) = \sum_{i=1}^n l_{ni}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) \quad (16)$$

with

$$\begin{aligned} l_{ni}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) &= r_i \log \pi(Y_i, \mathbf{X}_i; \alpha_1, \boldsymbol{\phi}_1) + r_i \log p_a(Y_i|\mathbf{X}_i; \boldsymbol{\theta}_a) \\ &\quad + (1 - r_i) \log \int \{1 - \pi(y, \mathbf{X}_i; \alpha_1, \boldsymbol{\phi}_1)\} p_a(y|\mathbf{X}_i; \boldsymbol{\theta}_a) dy \end{aligned}$$

and

$$p_a(Y_i|\mathbf{X}_i; \boldsymbol{\theta}_a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp \left\{ -\frac{(Y_i - \boldsymbol{\beta}_a^\top \mathbf{X}_i)^2}{2\sigma_a^2} \right\}$$

where $\boldsymbol{\theta}_a = (\boldsymbol{\beta}_a^\top, \sigma_a^2)^\top$.

For convenience, we denote $\pi(y, \mathbf{X}_i; \alpha_1, \boldsymbol{\phi}_1)$ by $\pi_{1,i}$. The score equations can be derived by means of taking the derivatives of (16) with respect to $\alpha_1, \boldsymbol{\phi}_1$ and $\boldsymbol{\theta}_a$, yielding

$$\sum_{i=1}^n \mathbf{U}_i(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) = \begin{pmatrix} \sum_{i=1}^n \mathbf{U}_{\alpha_1,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) \\ \sum_{i=1}^n \mathbf{U}_{\boldsymbol{\phi}_1,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) \\ \sum_{i=1}^n \mathbf{U}_{\boldsymbol{\beta}_a,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) \\ \sum_{i=1}^n \mathbf{U}_{\sigma_a^2,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) \end{pmatrix} = 0 \quad (17)$$

where

$$\begin{aligned} \mathbf{U}_{\alpha_1,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) &= R_i(1 - \pi_{1,i})Y_i - (1 - R_i) \frac{E\{\pi_{1,i}(1 - \pi_{1,i})Y_i\}|\mathbf{X}_i}{E(1 - \pi_{1,i}|\mathbf{X}_i)} \\ \mathbf{U}_{\boldsymbol{\phi}_1,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) &= \left[R_i(1 - \pi_{1,i}) - (1 - R_i) \frac{E\{\pi_{1,i}(1 - \pi_{1,i})\}|\mathbf{X}_i}{E(1 - \pi_{1,i}|\mathbf{X}_i)} \right] \mathbf{X}_i \\ \mathbf{U}_{\boldsymbol{\beta}_a,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) &= \frac{1}{\sigma_a^2} \left[\mathbf{X}_i \left\{ R_i Y_i - (1 - R_i) \frac{E\{(1 - \pi_{1,i})Y_i\}|\mathbf{X}_i}{E(1 - \pi_{1,i}|\mathbf{X}_i)} \right\} - \mathbf{X}_i \mathbf{X}_i^\top \boldsymbol{\beta}_a \right] \end{aligned}$$

and

$$\begin{aligned} \mathbf{U}_{\sigma_a^2,i}(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a) &= \frac{1}{2\sigma_a^4} \left[\left\{ R_i(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}_a)^2 + \right. \right. \\ &\quad \left. \left. (1 - R_i) \frac{E\{(1 - \pi_{1,i})(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}_a)^2|\mathbf{X}_i\}}{E(1 - \pi_{1,i}|\mathbf{X}_i)} \right\} - \sigma_a^2 \right] \end{aligned}$$

However, computing the integral involved in (17) is challenging. To circumvent this difficulty, we propose utilizing Monte Carlo approximation. Given $\boldsymbol{\theta}_a$, we can straightforwardly sample Y directly from the conditional distribution with density function $p(y|\mathbf{X}_i; \boldsymbol{\theta}_a)$. Let $Y_{i1}, Y_{i2}, \dots, Y_{im}$ denote a sample of size m . For convenience, we denote

$$\begin{aligned} \frac{E\{(1 - \pi_{1,i})Y_i|\mathbf{X}_i\}}{E(1 - \pi_{1,i}|\mathbf{X}_i)} &\approx \frac{\sum_{j=1}^m (1 - \pi_{1,i})Y_{ij}}{\sum_{j=1}^m (1 - \pi_{1,i})} \equiv I_{1,i}(\alpha_1, \boldsymbol{\phi}_1) \\ \frac{E\{\pi_{1,i}(1 - \pi_{1,i})|\mathbf{X}_i\}}{E(1 - \pi_{1,i}|\mathbf{X}_i)} &\approx \frac{\sum_{j=1}^m \pi_{1,ij}(1 - \pi_{1,ij})}{\sum_{j=1}^m (1 - \pi_{1,i})} \equiv I_{2,i}(\alpha_1, \boldsymbol{\phi}_1) \\ \frac{E\{(1 - \pi_{1,i})Y_i^2|\mathbf{X}_i\}}{E(1 - \pi_{1,i}|\mathbf{X}_i)} &\approx \frac{\sum_{j=1}^m (1 - \pi_{1,ij})Y_{ij}^2}{\sum_{j=1}^m (1 - \pi_{1,i})} \equiv I_{3,i}(\alpha_1, \boldsymbol{\phi}_1) \end{aligned}$$

As stated in [17], the introduced additional variability becomes ignorable when the resample size m is sufficiently large; for example, $m = O(n^{1+\epsilon})$, where ϵ is a small positive constant, so we use \approx here. We record the parameters of t iterations as $(\alpha_1^t, \boldsymbol{\phi}_1^t, \boldsymbol{\theta}_a^t)$ and then obtain the estimated value of $(\alpha_1^{t+1}, \boldsymbol{\phi}_1^{t+1}, \boldsymbol{\theta}_a^{t+1})$ times as follows:

$$\begin{aligned} \sum_{i=1}^n [R_i(1 - \pi_{1,i})Y_i - (1 - R_i)I_{i,1}(\alpha_1^t, \boldsymbol{\phi}_1^t)] &= 0 \\ \sum_{i=1}^n [R_i(1 - \pi_{1,i}) - (1 - R_i)I_{i,2}(\alpha_1^t, \boldsymbol{\phi}_1^t)]\mathbf{X}_i &= 0 \\ \sum_{i=1}^n [\mathbf{X}_i\{R_iY_i - (1 - R_i)I_{i,1}(\alpha_1^t, \boldsymbol{\phi}_1^t)\} - \mathbf{X}_i\mathbf{X}_i^\top\boldsymbol{\beta}_a] &= 0 \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^n [R_i(Y_i - \mathbf{X}_i^\top\boldsymbol{\beta}_a)^2 + (1 - R_i)\{I_{i,3}(\alpha_1^t, \boldsymbol{\phi}_1^t) \\ - 2I_{i,1}(\alpha_1^t, \boldsymbol{\phi}_1^t)\mathbf{X}_i^\top\boldsymbol{\beta}_a + (\mathbf{X}_i^\top\boldsymbol{\beta}_a)^2\} - \sigma_a^2] &= 0 \end{aligned}$$

The above formula is iterated through until the following is satisfied:

$$\|\{\alpha_1^t - \alpha_1^{t+1}, (\boldsymbol{\phi}_1^t - \boldsymbol{\phi}_1^{t+1})^\top, (\boldsymbol{\theta}_a^t - \boldsymbol{\theta}_a^{t+1})^\top\}^\top\| < tol$$

where tol is a given constant that can be arbitrarily small. Ultimately, we obtain $(\hat{\alpha}_1, \hat{\boldsymbol{\phi}}_1, \hat{\boldsymbol{\theta}}_a)$ for the parameters $(\alpha_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_a)$. In the same manner, we can obtain $(\hat{\alpha}_0, \hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\theta}}_n)$, where $\hat{\boldsymbol{\theta}}_a = (\hat{\boldsymbol{\beta}}_a^\top, \hat{\sigma}_a^2)^\top$, $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n^\top, \hat{\sigma}_n^2)^\top$. In practice, the above computation procedure is straightforward to implement, as it relies on the use of the empirical distribution. Moreover, it is attractive because the introduced additional variability becomes ignorable when the resample size m is sufficiently large; for instance, $m = O(n^{1+\epsilon})$, where ϵ is a small positive constant.

According to the assumption of Condition 7, the absence of Y is unrelated to Z . Then, we can focus on the individuals having $(Z_i, D_i) = (1, 1)$ and, using the inverse probability weighting method, the estimated equation is as follows:

$$\begin{aligned} \sum_{i=1}^n \frac{Y_i - \mathbf{X}_i^\top\boldsymbol{\beta}_M}{\pi(y, \mathbf{X}_i; \hat{\alpha}_1, \hat{\boldsymbol{\phi}}_1)} &= 0 \\ \sum_{i=1}^n \frac{(Y_i - \mathbf{X}_i^\top\boldsymbol{\beta}_M)^2}{\pi(y, \mathbf{X}_i; \hat{\alpha}_1, \hat{\boldsymbol{\phi}}_1)} - \sigma_M^2 &= 0 \end{aligned}$$

We can readily obtain that the estimators for $\boldsymbol{\beta}_M$ and σ_M^2 are $\hat{\boldsymbol{\beta}}_M$ and $\hat{\sigma}_M^2$ from the above estimation equation, and considering the numerical characteristics of the mixed normal distribution, we have

$$\begin{aligned} \hat{\boldsymbol{\beta}}_M &= \frac{\hat{\omega}_a}{\hat{\omega}_a + \hat{\omega}_{1c}}\hat{\boldsymbol{\beta}}_a + \frac{\hat{\omega}_{1c}}{\hat{\omega}_a + \hat{\omega}_{1c}}\boldsymbol{\beta}_{1c} \\ \hat{\sigma}_M^2 &= \frac{\hat{\omega}_a}{\hat{\omega}_a + \hat{\omega}_{1c}}\hat{\sigma}_a^2 + \frac{\hat{\omega}_{1c}}{\hat{\omega}_a + \hat{\omega}_{1c}}\sigma_{1c}^2 + \frac{\hat{\omega}_a\hat{\omega}_{1c}}{\hat{\omega}_a + \hat{\omega}_{1c}}\{(\hat{\boldsymbol{\beta}}_a - \boldsymbol{\beta}_{1c})^\top E\mathbf{X}\}^2 \end{aligned}$$

By solving the two equations above, we can derive estimators for $\boldsymbol{\beta}_{1c}$ and σ_{1c}^2 , denoted as $\hat{\boldsymbol{\beta}}_{1c}$ and $\hat{\sigma}_{1c}^2$. In the same manner, we can obtain $\hat{\boldsymbol{\theta}}_{0c}$. Here, $\hat{\boldsymbol{\theta}}_{1c} = (\hat{\boldsymbol{\beta}}_{1c}^\top, \hat{\sigma}_{1c}^2)^\top$, $\hat{\boldsymbol{\theta}}_{0c} = (\hat{\boldsymbol{\beta}}_{0c}^\top, \hat{\sigma}_{0c}^2)^\top$. By denoting $\boldsymbol{\beta}_c = \boldsymbol{\beta}_{1c} - \boldsymbol{\beta}_{0c}$, we arrive at the complier average causal effect given the predictors $\mathbf{X} = \mathbf{x}$ as

$$\text{CACE}(\mathbf{x}) = E\{Y(1) - Y(0)|U = c, \mathbf{X} = \mathbf{x}\} = (\hat{\boldsymbol{\beta}}_{1c} - \hat{\boldsymbol{\beta}}_{0c})^\top \mathbf{x} = \hat{\boldsymbol{\beta}}_c^\top \mathbf{x} \quad (18)$$

Considering the variance of the mixed normal distribution, the estimator for the corresponding variance $\sigma_c^2 = \kappa\sigma_{1c}^2 + (1 - \kappa)\sigma_{0c}^2 + \kappa(1 - \kappa)(\mu_{1c} - \mu_{0c})^2$, where $\kappa = 0.5$, $\mu_{1c} = \boldsymbol{\beta}_{1c}^\top \mathbf{X}$, $\mu_{0c} = \boldsymbol{\beta}_{0c}^\top \mathbf{X}$. Thus,

$$\hat{\sigma}_c^2 = 0.5\hat{\sigma}_{1c}^2 + 0.5\hat{\sigma}_{0c}^2 + 0.25(\hat{\beta}_c^\top E\mathbf{X})^2$$

In practice, we can utilize the bootstrap method to approximate the sampling variance of the estimator of CACE.

4. Simulations

In this section, we perform simulations to assess the effectiveness of the outlined estimation procedure. Assuming $P(Z = 1) = \xi = 0.5$ and $P(U = a) = \omega_a = 0.3, P(U = n) = \omega_n = 0.3$, we divide all samples into six groups based on $(Z = z, U = u)$, with the number in each group denoted as n_{zu} . The covariates \mathbf{X} in all groups are two-dimensional normal distributions, with the mean vector and the covariance matrix as follows, respectively:

$$\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

Conditional on \mathbf{X} , if $(Z_i, U_i) = (1, a)$ or $(0, a)$, the response Y is generated from

$$Y|\mathbf{X} \sim N(\beta_{a,0} + \beta_{a,1}X_1 + \beta_{a,2}X_2, \sigma_a^2)$$

where $(\beta_{a,0}, \beta_{a,1}, \beta_{a,2}, \sigma_a^2)^\top = (1, 1, 2, 1)^\top$. Instead, if $(Z_i, U_i) = (1, n)$ or $(0, n)$, we use

$$Y|\mathbf{X} \sim N(\beta_{n,0} + \beta_{n,1}X_1 + \beta_{n,2}X_2, \sigma_n^2)$$

where $(\beta_{n,0}, \beta_{n,1}, \beta_{n,2}, \sigma_n^2)^\top = (-1, -1, 2, 1)^\top$. If $(Z_i, U_i) = (1, c)$, we simulate Y via the conditional distribution

$$Y|\mathbf{X} \sim N(\beta_{1c,0} + \beta_{1c,1}X_1 + \beta_{1c,2}X_2, \sigma_{1c}^2)$$

where $(\beta_{1c,0}, \beta_{1c,1}, \beta_{1c,2}, \sigma_{1c}^2)^\top = (0.5, 1.5, 2, 1)^\top$, while if $(Z_i, U_i) = (0, c)$, we have

$$Y|\mathbf{X} \sim N(\beta_{0c,0} + \beta_{0c,1}X_1 + \beta_{0c,2}X_2, \sigma_{0c}^2)$$

where $(\beta_{0c,0}, \beta_{0c,1}, \beta_{0c,2}, \sigma_{0c}^2)^\top = (-0.5, -1.5, 2, 1)^\top$. The missing mechanism model in the group with $D = 1$ is as follows:

$$P(R = 1|Y, \mathbf{X}) = \text{logit}^{-1}\{\alpha_1 Y + \phi_{a,0} + \phi_{a,1}X_1\}$$

where $\text{logit}^{-1}(\cdot) = \exp(\cdot)/\{1 + \exp(\cdot)\}$ and $(\alpha_1, \phi_{1,0}, \phi_{1,1})^\top = (0.2, 1.5, 0.3)^\top$. Instead, if $D_i = 0$, we use

$$P(R = 1|Y, \mathbf{X}) = \text{logit}^{-1}\{\alpha_0 Y + \phi_{0,0} + \phi_{0,1}X_1\}$$

where $(\alpha_0, \phi_{0,0}, \phi_{0,1})^\top = (0.2, 2, 0.2)^\top$. The missing rate for the outcome is 13.4% in group $D = 1$ and 15.7% in group $D = 0$.

The Monte Carlo samples have a size of $n = 3000$. Tables 1 and 2 report the empirical bias, standard deviation obtained by means of non-parametric bootstrap with $B = 1000$ replications, and coverage of 95% Wald-type confidence intervals. The numerical results of Table 1 demonstrate a highly accurate estimation of η , the missing mechanism models parameters, the always-taker parameters, and the never-taker parameters, which means that the proposed method can still accurately estimate the parameters in the causal model under the non-ignorable missingness. The numerical results in Table 2 demonstrate a highly accurate estimation of the complier parameters and CACE components. However, there is some under-coverage in the Wald confidence intervals for $\beta_{1c,0}$ and $\beta_{c,0}$.

Table 1. Empirical bias (Bias), bootstrap standard deviation (Std. dev.) and 95% Wald coverage probability (95% Cover) of the estimators for η , the missing mechanism models parameters, the always-taker parameters, and the never-taker parameters.

	Parameter	Bias	Std. dev.	95% Cover
η	ξ	0.0004	0.009	0.983
	ω_a	0.0001	0.012	0.915
	ω_n	0.0001	0.012	0.906
$D = 1$	α_1	-0.0009	0.074	0.972
	$\phi_{1,0}$	0.0115	0.205	0.978
	$\phi_{1,1}$	0.0231	0.270	0.992
$D = 0$	α_0	0.0035	0.073	0.957
	$\phi_{0,0}$	0.0329	0.311	0.991
	$\phi_{0,1}$	-0.0038	0.148	0.985
$U = a$	$\beta_{a,0}$	0.0035	0.093	0.926
	$\beta_{a,1}$	-0.0014	0.073	0.944
	$\beta_{a,2}$	-0.0006	0.050	0.960
	σ_a^2	-0.0084	0.071	0.933
$U = n$	$\beta_{n,0}$	-0.0020	0.089	0.937
	$\beta_{n,1}$	0.0012	0.072	0.923
	$\beta_{n,2}$	0.0003	0.054	0.952
	σ_n^2	-0.0047	0.074	0.918

Table 2. Empirical bias (Bias), bootstrap standard deviation (Std. dev.) and 95% Wald coverage probability (95% Cover) of the estimators for the compliers parameters and the CACE components.

	Parameter	Bias	Std. dev.	95% Cover
$(Z, U) = (1, c)$	$\beta_{1c,0}$	-0.059	0.258	0.947
	$\beta_{1c,1}$	-0.0026	0.139	0.881
	$\beta_{1c,2}$	0.0014	0.107	0.937
	σ_{1c}^2	0.0025	0.074	0.987
$(Z, U) = (0, c)$	$\beta_{0c,0}$	0.0039	0.134	0.949
	$\beta_{0c,1}$	-0.0040	0.110	0.969
	$\beta_{0c,2}$	-0.0001	0.077	0.989
	σ_{0c}^2	-0.0031	0.124	0.907
CACE	$\beta_{c,0}$	-0.0065	0.193	0.883
	$\beta_{c,1}$	0.0054	0.155	0.936
	$\beta_{c,2}$	0.0025	0.106	0.991
	σ_c^2	-0.0559	0.166	0.908

The estimators for η exhibit significantly smaller bias and variance compared to other parameters, attributed to the larger sample size used for their estimation. Conversely, the estimators for the always-taker and never-taker parameters show similar bias and variance due to their comparable missing rates. The subpar estimation results for the compliers parameters arise from various factors, including the numerical characteristics of the normal distribution, which we aim to optimize in future iterations.

5. Real Data Analysis

The CHIP study monitors income distribution and economic factors among rural, rural-to-urban migrant, and urban households in China [31]. The purpose of this study is to investigate whether there is a significant change in the return rate of education when the male population transitions from rural to migrant status. The survey captures details such as employment status, education level, age, income, and other relevant information. Data collection involves systematic random sampling from various regions to ensure geographic representation. The survey includes data from cities and towns in fifteen

provinces, representing different regions across the country. These provinces include Liaoning, Shanxi, Jiangsu, Shandong, Guangdong, Anhui, Henan, Sichuan, Hunan, Hubei, Gansu, Xinjiang, Yunnan, Beijing, and Chongqing, covering the north, eastern coastal areas, interior regions, and western regions of China.

Each respondent's actual annual income serves as a proxy for the individual's earnings E , utilized in the response of the Mincer earnings function [32]. We compute the number of years since leaving school as age – years of schooling – 6, as Chinese children typically start school at 7 years old. The years since the onset of labor market experience are calculated as age – 16, as Chinese individuals who have reached the age of 16 can legally participate in the labor market.

The rural population's migration for work primarily depends on two factors: the economic development status of the region and the distance between their home and work. If the region's economic development is favorable, migration may be less likely. Conversely, shorter distances between the home and work place increase the likelihood of migration. Hence, leveraging Chinese administrative division data, we assign $Z_i = 1$ if the individual's household head's registered residence is Beijing, Shanxi, Guangdong, Hubei Chongqing, or Liaoning, and $Z_i = 0$ otherwise. The sample sizes of the observed data are shown in Table 3. In the 2013 wave of the study, the rural and migrant sub-samples accounted for 62.97% and 4.54% of the dataset, respectively, including urban individuals. However, in China in 2013, the percentages of rural and migrant households were of 45.77% and 13.30%, respectively [33]. To align with this distribution, we adjust the sample weights. Specifically, each migrant respondent is weighted four times as much as a rural one, since $(62.97/4.54) / (45.77/13.30) \approx 4$.

Table 3. Sample sizes of different groups. The initials in brackets identify never-takers (n), always-takers (a), and compliers (c).

	$(Z, D) = (1, 1)$	$(Z, D) = (1, 0)$	$(Z, D) = (0, 1)$	$(Z, D) = (0, 0)$
$R = 1$	472 (a, c)	229 (n)	4052 (a)	5464 (c, n)
$R = 0$	37 (a, c)	13 (n)	1059 (a)	1857 (c, n)

The Mincer earnings function is a single-equation model that explains wage income as a function of education and work experience, which can be expressed as follows:

$$\log(1 + E) = \beta_0 + \beta_1 S + \beta_2 Exper + \beta_3 Exper^2 + \varepsilon \quad (19)$$

In the specification, E represents earnings (in CNY), S indicates years of schooling, and $Exper$ stands for potential work experience. ε is an unobserved normal random error with mean 0 and variance $\sigma^2 > 0$. Without considering the cost of schooling, the returns to education are given by $\partial \log(1 + E) / \partial S = \beta_1$. The missing data mechanism is modeled using the following model:

$$\begin{aligned} P(R = 1 | E, S, Exper) = & \text{logit}^{-1}(\alpha \log(1 + E) \\ & + \theta_0 + \theta_1 \log S + \theta_2 \log Exper + \theta_3 \log Exper^2) \end{aligned} \quad (20)$$

Using the Assumption 9 described in Section 2 in this setting, model (19) implies a normal distribution of the outcome $Y = \log(1 + E)$:

$$p(y | \mathbf{x}; \boldsymbol{\theta}_{zu}) = \frac{1}{\sqrt{2\pi}\sigma_{zu}} \exp \left[-\frac{(y - \beta_{zu,0} - \beta_{zu,1}s - \beta_{zu,2}exper - \beta_{zu,3}exper^2)^2}{2\sigma_{zu}^2} \right] \quad (21)$$

where $\mathbf{x} = (s, exper)^\top$ is the realization of the covariates vector $\mathbf{X} = (S, Exper)^\top$. Thus, the complier average causal effect which captures the causal effect of migrant work on the earnings of compliers is estimated by

$$\text{CACE}(s, exper; \hat{\beta}_c) = \hat{\beta}_{c,0} + \hat{\beta}_{c,1}s + \hat{\beta}_{c,3}exper + \hat{\beta}_{c,3}exper^2$$

with $\hat{\beta}_c = (\hat{\beta}_{c,0}, \hat{\beta}_{c,1}, \hat{\beta}_{c,2}, \hat{\beta}_{c,3})^\top = (\hat{\beta}_{1c,0} - \hat{\beta}_{0c,0}, \hat{\beta}_{1c,1} - \hat{\beta}_{0c,1}, \hat{\beta}_{1c,2} - \hat{\beta}_{0c,2}, \hat{\beta}_{1c,3} - \hat{\beta}_{0c,3})^\top$. Specifically, the coefficient $\hat{\beta}_{c,1}$ gives the estimated causal effect of migrant work on the returns to education of a complier.

The analysis of the 2013 CHIP data conducted in R version 4.3.2, which invented in August 1993 by statisticians Ross Ihaka and Robert Jetman of the University of Auckland, New Zealand. Utilizing the estimation methods outlined in Section 3 of this paper, yielded results presented in Tables 4 and 5. The results from Table 4 show that log-income has a significant positive effect on the probability of missingness, as evidenced by $\alpha_1 = 0.4130$ and $\alpha_0 = 0.5284$. Table 5 illustrates that the estimated returns to education are $\hat{\beta}_{1c,1} = 2.81\%$ for migrant compliers and $\hat{\beta}_{0c,1} = -0.85\%$ for rural compliers. The difference, $\hat{\beta}_{c,1} = 3.66\%$, indicates that migrant work enhances the returns to education for compliers. Additionally, based on the observation that $\beta_0 > 0$, we infer that the initial wages of migrant compliers are notably higher than those of rural compliers. This suggests that migrant work also contributes to higher initial wages for individuals. In conclusion, the estimated returns to education vary significantly among different target groups. However, it is noteworthy that the estimated returns to education of migrant workers and rural residents are consistently lower than those of urban residents across the board, with an average difference of 8.4%, as documented by [34]. This finding scientifically validates the social and economic significance of migrant work from a human capital perspective, offering a basis for decision-making aimed at enhancing the condition of Chinese rural labor.

Table 4. Parameter estimators, bootstrap standard deviation (Std. dev.), and empirical 95% Wald confidence intervals (95% CI) along with the estimators for η , the missing mechanism model parameters, the always-taker parameters, and the never-taker parameters.

	Parameter	Estimator	Std. dev.	95% CI
η	ξ	0.1946	0.0031	[0.1886, 0.2006]
	ω_a	0.1169	0.0027	[0.1117, 0.1222]
	ω_n	0.7154	0.0060	[0.7036, 0.7271]
$D = 1$	α_1	0.4130	0.1521	[0.1150, 0.7111]
	$\phi_{1,0}$	-7.0218	1.5790	[10.1167, -3.9269]
	$\phi_{1,1}$	0.3597	0.0627	[0.2369, 0.4826]
	$\phi_{1,2}$	0.3003	0.0391	[0.2238, 0.3769]
	$\phi_{1,3}$	-0.0060	0.0008	[-0.0076, -0.0045]
$D = 0$	α_0	0.5284	0.2365	[0.0647, 0.9920]
	$\phi_{0,0}$	-5.2175	2.2725	[-9.6715, -0.7635]
	$\phi_{0,1}$	0.2046	0.0707	[0.0659, 0.3432]
	$\phi_{0,2}$	0.0624	0.0335	[-0.0032, 0.1281]
	$\phi_{0,3}$	-0.0018	0.0006	[-0.0031, -0.0006]
$U = a$	$\beta_{a,0}$	9.1899	0.1464	[8.9029, 9.4771]
	$\beta_{a,1}$	0.0356	0.0084	[0.0191, 0.0521]
	$\beta_{a,2}$	0.0894	0.0079	[0.0739, 0.1050]
	$\beta_{a,3}$	-0.0020	0.0002	[-0.0023, -0.0016]
	σ_a^2	0.5331	0.0282	[0.4777, 0.5884]
$U = n$	$\beta_{n,0}$	9.3902	0.1468	[9.1024, 9.6780]
	$\beta_{n,1}$	0.0498	0.1011	[0.0300, 0.0696]
	$\beta_{n,2}$	0.0393	0.0072	[0.0252, 0.0534]
	$\beta_{n,3}$	-0.0009	0.0001	[-0.0012, -0.0006]
	σ_n^2	0.5513	0.0382	[0.4764, 0.6262]

Table 5. Parameter estimators, bootstrap standard deviation (Std. dev.), and empirical 95% Wald confidence intervals (95% CI) along with the estimators for the complier parameters and the CACE components.

	Parameter	Estimator	Std. dev.	95% CI
$(Z, U) = (1, c)$	$\beta_{1c,0}$	10.2684	0.1981	[9.8800, 10.6567]
	$\beta_{1c,1}$	0.0281	0.0131	[0.0023, 0.0537]
	$\beta_{1c,2}$	0.0149	0.0132	[-0.0111, 0.0408]
	$\beta_{1c,3}$	-0.0005	0.0003	[-0.0011, 0.0001]
	σ_{1c}^2	0.5972	0.0518	[0.4958, 0.6987]
$(Z, U) = (0, c)$	$\beta_{0c,0}$	9.0075	1.1509	[6.7518, 11.2633]
	$\beta_{0c,1}$	-0.0085	0.1009	[-0.2063, 0.1893]
	$\beta_{0c,2}$	0.0310	0.0472	[-0.0616, 0.1236]
	$\beta_{0c,3}$	-0.0012	0.0011	[-0.0033, 0.0009]
	σ_{0c}^2	2.2993	1.0110	[0.3177, 4.2809]
CACE	$\beta_{c,0}$	1.2609	1.1994	[-1.0899, 3.6116]
	$\beta_{c,1}$	0.0366	0.1026	[-0.1646, 0.2378]
	$\beta_{c,2}$	0.0162	0.0490	[-0.1122, 0.0799]
	$\beta_{c,3}$	0.0007	0.0011	[-0.0015, 0.0029]
	σ_c^2	2.2835	0.5075	[1.2888, 3.2782]

6. Conclusions

In this study, the challenge of identifying and estimating complier average causal effect parameters under non-ignorable missingness is tackled by increasing covariates to mitigate the sensitivity to the violation of specific identification assumptions. The missing data mechanism is assumed to follow a logistic model, wherein the absence of the outcome is explained by the outcome itself, the treatment received and the covariates, giving it an advantage over the assumptions proposed by [2,3]. The identifiability of the models is established under mild conditions by assuming that the outcome follows a normal distribution. A computational method is developed to estimate model parameters through a two-step likelihood estimation approach, utilizing subgroup analysis.

Some studies in the literature discuss the consistency of parameter estimation in the presence of non-ignorable missing and non-compliant data. The authors of [23] demonstrated that parameter estimators are consistent under the assumption of a correct missing mechanism model. The authors of [26] obtained the consistency of parameters of interest even when confounders are missing not at random. The authors of [3] established the asymptotic results of the estimators when the missing outcome depends only on itself. Estimating the parameters of interest based on subgroup analysis poses significant challenges when the absence of the outcome is explained by the outcome itself, the treatment received, and the covariates. We plan to address this aspect in future research endeavors to enhance our methodologies.

There are many directions worthy of further research. A possible extension in this research area involves utilizing instrumental variables to transform the identifiability of the observation likelihood into the identifiability of the parameters of interest. Indeed, we propose a relaxation of the exclusion restriction based on likelihood analysis, resulting in a parametric model characterized by mixtures of distributions. Furthermore, we can adopt a semi-parametric model for theoretical modeling, incorporating more sophisticated structures into the missing mechanism models and regression models.

Author Contributions: J.D.: methodology, software, writing—original draft preparation; G.W.: data curation, software, validation; X.L.: methodology, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Social Science Foundation of Guangxi Province in China, grant number 23BTJ001.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We thank the Editor, Associate Editor, and referees, as well as our financial sponsors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, H.; Geng, Z.; Zhou, X. Identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data. *Biometrics* **2009**, *109*, 142–149. [[CrossRef](#)] [[PubMed](#)]
2. Chen, H.; Ding, P.; Geng, Z.; Zhou, X. Semiparametric Inference of the Complier Average Causal Effect with Nonignorable Missing Outcomes. *ACM Trans. Intell. Syst. Technol. (TIST)* **2015**, *7*, 1–15. [[CrossRef](#)]
3. Zheng, R. Causal Inference with Unmeasured Confounding from Nonignorable Missing Outcomes. *arXiv* **2023**, arXiv:2305.07226.
4. Zhao, J.; Ma, Y. A versatile estimation procedure without estimating the nonignorable missingness mechanism. *J. Am. Stat. Assoc.* **2022**, *117*, 1916–1930. [[CrossRef](#)]
5. Du, J.; Cui, X. Semiparametric estimation in generalized additive partial linear models with nonignorable nonresponse data. *Stat. Pap.* **2023**, *1*–25. [[CrossRef](#)]
6. Imbens, G.W.; Rubin, D.B. Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann. Stat.* **1997**, *25*, 305–327. [[CrossRef](#)]
7. Zhou, J.; Hodges, J.S.; Suri, M.F.K.; Chu, H. A Bayesian hierarchical model estimating CACE in meta-analysis of randomized clinical trials with noncompliance. *Biometrics* **2019**, *75*, 978–987. [[CrossRef](#)] [[PubMed](#)]
8. Park, S.; Kürüm, E. A two-stage joint modeling method for causal mediation analysis in the presence of treatment noncompliance. *J. Causal Inference* **2020**, *8*, 131–149. [[CrossRef](#)]
9. Dempster, A.P.; Laird, N.P.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–22. [[CrossRef](#)]
10. Mercatanti, A. A Likelihood-Based Analysis for Relaxing the Exclusion Restriction in Randomized Experiments with Noncompliance. *Aust. N. Z. J. Stat.* **2013**, *55*, 129–153. [[CrossRef](#)]
11. Abadie, A. Semiparametric instrumental variable estimation of treatment response models. *J. Econom.* **2003**, *55*, 231–263. [[CrossRef](#)]
12. Wang, S.; Shao, J.; Kim, J.K. An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Stat. Sin.* **2014**, *24*, 1097–1116. [[CrossRef](#)]
13. Tchetgen, T.E.; Eric, J.; Wirth, K.E. A general instrumental variable framework for regression analysis with outcome missing not at random. *Biometrics* **2017**, *73*, 1123–1131. [[CrossRef](#)] [[PubMed](#)]
14. Wang, L.; Shao, J.; Fang, F. Propensity model selection with nonignorable nonresponse and instrument variable. *Stat. Sin.* **2021**, *31*, 647–672. [[CrossRef](#)]
15. Morikawa, K.; Kim, J.K. Semiparametric optimal estimation with nonignorable nonresponse data. *Ann. Stat.* **2021**, *49*, 2991–3014. [[CrossRef](#)]
16. Cui, X.; Guo, J.; Yang, G. On the identifiability and estimation of generalized linear models with parametric nonignorable missing data mechanism. *Comput. Stat. Data Anal.* **2017**, *107*, 64–80. [[CrossRef](#)]
17. Du, J.; Li, Y.; Cui, X. Identification and Estimation of Generalized Additive Partial Linear Models with Nonignorable Missing Response. *Commun. Math. Stat.* **2024**, *12*, 113–156. [[CrossRef](#)]
18. Ding, P.; Li, F. Causal inference. *Stat. Sci.* **2018**, *33*, 214–237. [[CrossRef](#)]
19. Yang, S.; Wang, L.; Ding, P. Causal inference with confounders missing not at random. *Biometrika* **2018**, *106*, 875–888. [[CrossRef](#)]
20. Frangakis, C.E.; Rubin, D.P. Principal stratification in causal inference. *Biometrics* **2002**, *58*, 21–29. [[CrossRef](#)]
21. Taylor, L.; Zhou, X. Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics* **2009**, *65*, 88–95. [[CrossRef](#)] [[PubMed](#)]
22. Nguyen, T.Q.; Carlson, M.C.; Stuart, E.A. Identification of complier and noncomplier average causal effects in the presence of latent missing-at-random (LMAR) outcomes: A unifying view and choices of assumptions. *arXiv* **2023**, arXiv:2312.11136.
23. Li, W.; Zhou, X. Identifiability and estimation of causal mediation effects with missing data. *Stat. Med.* **2017**, *36*, 3948–3965. [[CrossRef](#)] [[PubMed](#)]
24. Sun, Z.; Liu, L. Semiparametric inference of causal effect with nonignorable missing confounders. *Stat. Sin.* **2021**, *31*, 1669–1688. [[CrossRef](#)]
25. He, Y.; Zheng, L.; Luo, P. Treatment Benefit and Treatment Harm Rates with Nonignorable Missing Covariate, Endpoint, or Treatment. *Mathematics* **2023**, *11*, 4459. [[CrossRef](#)]
26. Sun, J.; Fu, B. Identification and Estimation of Causal Effects with Confounders Missing Not at Random. *arXiv* **2023**, arXiv:2303.05878.
27. Rubin, D.B. Randomization analysis of experimental data: The Fisher randomization test comment. *J. Am. Stat. Assoc.* **1980**, *75*, 591–593. [[CrossRef](#)]
28. Ogburn, E.L.; Rotnitzky, A.; Robins, J.M. Doubly robust estimation of the local average treatment effect curve. *J. R. Stat. Soc. Ser. B* **2015**, *77*, 373–396. [[CrossRef](#)] [[PubMed](#)]

29. Seaman, S.R.; White, I.R. Review of inverse probability weighting for dealing with missing data. *Stat. Methods Med. Res.* **2013**, *22*, 278–295. [[CrossRef](#)]
30. Ogburn, E.L.; Rotnitzky, A.; Robins, J.M. The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opin. Q.* **2002**, *66*, 371–403.
31. Terry, S.; Li, S.; Yue, X.; Hiroshi, S. *Changing Trends in China’s Inequality: Evidence, Analysis, and Prospects*; Oxford University Press: New York, NY, USA, 2020.
32. Mincer, J. *Schooling, Experience and Earnings*; Columbia University Press: New York, NY, USA, 1974.
33. Luo, C.; Li, S.; Yue, X. An analysis of changes in the extent of income disparity in China (2013–2018). *Soc. Sci. China* **2021**, *1*, 33–54.
34. Li, H.; Liu, P.W.; Zhang, J. Estimating returns to education using twins in urban China. *J. Dev. Econ.* **2012**, *97*, 494–504. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.