Identification and multiply robust estimation of causal effects via instrumental variables from an auxiliary population

Wei Li, Jiapeng Liu, Peng Ding & Zhi Geng

View supplementary material

Accepted author version posted online: 27 Oct 2025.

Submit your article to this journal

Article views: 23

View related articles

View Crossmark data

Identification and multiply robust estimation of causal effects via instrumental variables from an auxiliary population

Wei Li[a], Jiapeng Liu[a], Peng Ding[b], Zhi Geng[c,*]

[a]Center for Applied Statistics and School of Statistics, Renmin University of China

[b]Department of Statistics, University of California, Berkeley

[c]School of Mathematics and Statistics, Beijing Technology and Business University

*zhigeng@pku.edu.cn

Abstract

Estimating causal effects in a target population with unmeasured confounders is challenging, especially when instrumental variables (IVs) are unavailable. However, IVs from auxiliary populations with similar problems can help infer causal effects in the target population. While the homogeneous conditional average treatment effect assumption has been widely used for effect transportability, it has not been explored in IV-based data fusion. We include it as a basic approach, though it may be biased when treatment effect heterogeneity exists. As an alternative approach, we introduce the equi-confounding assumption that the unmeasured confounding bias remains the same after adjusting for observed covariates, while allowing conditional average treatment effects to differ across populations. This allows us to identify the confounding bias in the auxiliary population and remove it from the treatment-outcome association in the target population to recover the causal effect. We develop multiply robust estimators under both approaches and demonstrate them through simulation studies and a real data application.

1 Introduction

It is challenging to estimate causal effect of a treatment on an outcome in observational studies with unmeasured confounding. A commonly employed strategy to address unmeasured confounding is to use instrumental variables (IVs). However, IVs may be unavailable in the target population. In such cases, auxiliary studies with complete IV information can be leveraged to estimate causal effects in the target population. For example, researchers are interested in the effect of education on earnings in a specific industry within a developing country where no suitable IVs are available in the observed data. Similar studies in developed countries have used changes in compulsory schooling laws as IVs to study the same causal problem (Angrist and Krueger, 1991; Oreopoulos, 2006). By appropriately using studies from the developed countries, we can infer the causal effect of education on earnings in the developing country without IVs.

Integrating data from other sources has become increasingly popular for achieving valid causal inference in the target population (Bareinboim and Pearl, 2016; Yang and Ding, 2020; Hu et al., 2022; Li et al., 2024; Colnet et al., 2024). A strand of literature focuses on generalizing or transporting findings from auxiliary randomized trials to the target population based on various versions of the selection exchangeability assumption (Stuart et al., 2011; Pearl and Bareinboim, 2014; Buchanan et al., 2018; Dahabreh et al., 2020; Yang et al., 2023). This assumption typically requires that the conditional average treatment effects (CATEs) given observed covariates are equal across both populations. In the absence of randomized trials, some studies use an auxiliary dataset with supplemental confounding information to mitigate estimation bias in the target population (Yang and Ding, 2020; Luo et al., 2025). In particular, Yang and Ding (2020) proposed a general framework to estimate causal effects in the setting where the target data have unmeasured confounders, but the smaller auxiliary data provide supplementary information on these confounders.

Another line of research leverages auxiliary data with IVs to establish identification and estimation of causal effects when the target data lack some important variables for subsequent analysis (Ridder and Moffitt, 2007; Zhao et al., 2019; Shuai et al., 2024). Specifically, in the classical two-sample IV framework, the target data contain information on the outcome, IV, and covariates, but lack information on the treatment; the auxiliary data provide information on the treatment, IV, and covariates, but lack information on the outcome. This method has been widely used in econometrics and social sciences, and mendelian randomization studies with genetic factors as IVs (Graham et al., 2016; Shu and Tan, 2020; Zhao et al., 2019). Recently, Sun and Miao (2022) developed semiparametric efficient estimation of average treatment effects in the two-sample IV framework. However, this framework requires measurements of IVs in both the target and auxiliary data, which differs from our setting where only the auxiliary data contain IVs.

In this paper, we address the problem of identifying and estimating causal effects in a target population, where unmeasured confounding may be present but IVs are unavailable. To tackle this challenge, we leverage auxiliary data from a population where valid IVs exist. Our approach is built on a set of assumptions that ensure the identifiability of the CATE in the auxiliary population using IVs. In particular, we adopt the no-additive-interaction assumption between unmeasured confounders and treatment (Wang and Tchetgen Tchetgen, 2018), which ensures the identification of the average causal effects with IVs. While Wang and Tchetgen Tchetgen (2018) studied IV-based identification in a single dataset, our work addresses a more complex data integration setting, where IVs are only available in the

auxiliary population. From this perspective, our framework can be viewed as their data integration counterpart. To extend the identification from the auxiliary population to the target population, additional assumptions are required.

We propose two distinct transportability approaches for identification. The homogeneous CATE assumption is canonical for transporting results from randomized trials to observational studies but has not been explored in IV-based data fusion problems. We include it as the basic approach. While it offers a simple way to transport effects, it may be biased when treatment effect heterogeneity exists. As an alternative approach, we introduce a novel equi-confounding assumption, which allows for heterogeneous CATEs while ensuring that after adjusting for observed covariates, the unmeasured confounding bias remains the same across populations. Unlike the homogeneous CATE assumption, which directly sets identical conditional treatment effects across populations, equi-confounding leverages additional information by explicitly estimating both the unmeasured confounding bias and the treatment-outcome association. Specifically, the unmeasured confounding bias can be computed in the auxiliary population and subsequently removed from the observed treatment-outcome association in the target population to identify the average causal effect. This adjustment enables effect transportability while allowing for treatment effect heterogeneity. Under each identifiability approach, we derive the efficient influence function to motivate the corresponding semiparametric estimator. These estimators are multiply robust in that they remain consistent even if the observed data model is partially misspecified. We also show that they are consistent and asymptotically normal, even when machine learning techniques are used for nuisance models estimation.

The rest of this paper is organized as follows. Section 2 introduces the notation, assumptions, and identification results. Section 3 presents the semiparametric multiply robust estimators and their machine learning analogues. Section 4 uses simulation studies to evaluate the finite-sample performance of the proposed estimators. Section 5 applies our approaches to estimate the average causal effect of smoking on physical functional status in the higher-income individuals. We conclude in Section 6 and relegate proofs to the supplementary material.

2 Notation, assumptions, and identification

Consider n individuals merged from two different sources. Let R denote the source indicator, where $R = 0$ if a subject is from the target population, and $R = 1$ if from the auxiliary population. In the target data, we observe a binary treatment X with 0 and 1 for the labels of control and active treatments, respectively. Let Y denote an outcome of interest and V denote a vector of pre-treatment covariates. We assume that U contains all unmeasured common causes of X and Y. We use the potential outcomes framework to define causal effects. Let Y_x denote the potential outcome if a subject, possibly contrary to the fact, were assigned to treatment x. We impose the consistency assumption that the observed outcome is $Y = XY_1 + (1-X)Y_0$. We are interested in the average treatment effect of the target population:

$$\tau = E(Y_1 - Y_0 \mid R = 0).$$

In the auxiliary sample, besides the treatment X, the outcome Y, and observed covariates V, a binary IV Z is also observed. Let $O = (X, Y, Z, V, R)$ if $R = 1$, and $O = (X, Y, V, R)$ if $R = 0$ denote the observed data. For $r = 0, 1$, define $\beta_r(V)$ to be the CATEs in the two populations:

$$\beta_r(V) = E(Y_1 - Y_0 \mid V, R = r).$$

We use $f(\cdot)$ to denote the probability density or mass function of a random variable (vector). Table 1 summarizes some key notation used throughout the paper.

2.1 Preliminary assumptions

In this subsection, we introduce several commonly-used assumptions that ensure the identification of $\beta_1(V)$ using IVs.

Assumption 1 (Latent ignorability and overlap).

(i) $Y_x \perp\!\!\!\perp X \mid V, U, R$ *for* $x = 0, 1$; *(ii)* $f(X = x \mid V, U, R) > 0$ *for* $x = 0, 1$.

Assumption 1(i) means that the treatment X is latent ignorable given observed covariates V and unmeasured variables U in both the target and auxiliary populations. Assumption 1(ii) ensures sufficient overlap between the covariate distributions of the treatment and control groups. In addition, it also implies overlap in treatment assignment between the target and auxiliary populations. To identify $\beta_1(V)$, Assumption 1 is required only for individuals with $R = 1$. We also include latent ignorability and overlap for $R = 0$ in Assumption 1, as it is required by the identification approach introduced in Section 2.3 later.

Assumption 2 (IV for auxiliary population).

(i) $Z \perp\!\!\!\perp U \mid V, R = 1$; *(ii)* $Z \perp\!\!\!\perp Y \mid X, V, U, R = 1$; *(iii)* $Z \perp\!\!\!\perp X \mid V, R = 1$.

Assumption 2 characterizes standard conditions for a valid IV Z in the auxiliary population. Assumption 2(i) requires the IV to be independent of unmeasured confounders U given V. Assumption 2(ii) formalizes no direct effect of the IV on the outcome Y. Assumption 2(iii) ensures that the IV is correlated with the treatment X even after conditioning on V. Although these assumptions are generally untestable without additional restrictions, possible IVs may be obtained through some natural or quasi-experiment in observational studies (Angrist and Krueger, 1991; Angrist et al., 1996; Baiocchi et al., 2014).

Assumptions 1 and 2 are commonly made in the causal inference literature. However, they are insufficient to identify $\beta_1(V)$ in the auxiliary population. To identify $\beta_1(V)$, we introduce the following no-interaction assumption (Wang and Tchetgen Tchetgen, 2018):

Assumption 3 (No additive *U-X* interaction).

$$E(Y_1 - Y_0 \mid V, U, R) = \beta_R(V).$$

The identification of $\beta_1(V)$ requires Assumption 3 to hold only for $R = 1$. Similar to Assumption 1, we also impose Assumption 3 for $R = 0$ to support later discussion.

Assumption 3 allows unmeasured confounders to affect the control potential outcome but assumes they do not modify the treatment effect on the outcome. It is reasonable when key effect modifiers are observed and adjusted for, reducing the role of unmeasured confounders in modifying the treatment effect. Similar no-interaction conditions are widely used in the literature for identifying the average treatment effect with valid or invalid IVs (Wang and Tchetgen Tchetgen, 2018; Sun and Miao, 2022; Tchetgen Tchetgen et al., 2021; Sun et al., 2022). A key implication of Assumption 3 is the need to collect as many predictors of the treatment and outcome as possible to ensure that no residual effect modification involving U remains within strata of the observed covariates V.

Assumptions 1 and 3 with $R = 1$, along with the IV Assumption 2, are sufficient to identify $\beta_1(V)$ as

$$\beta_1(V) = \frac{E(Y \mid Z = 1, V, R = 1) - E(Y \mid Z = 0, V, R = 1)}{E(X \mid Z = 1, V, R = 1) - E(X \mid Z = 0, V, R = 1)}. \quad (1)$$

The nonparametric representation in (1) corresponds to the conditional Wald estimand (Wang and Tchetgen Tchetgen, 2018; Sun and Miao, 2022).

Assumption 4 (Selection positivity).

$$f(R = r \mid V) > 0 \ for \ r = 0, 1.$$

Assumption 4 requires overlap of the observed covariate distribution between the target and auxiliary populations. Similar positivity assumptions have been widely used in the data fusion literature (Dahabreh et al., 2020; Degtiar and Rose, 2023; Shi et al., 2023; Colnet et al., 2024).

2.2 Identification under homogeneous CATE

By definition, the causal parameter is given by $\tau = E\{\beta_0(V) \mid R = 0\}$. To identify τ, additional assumptions are required to establish a connection between $\beta_0(V)$ and $\beta_1(V)$. A straightforward approach involves imposing selection exchangeability or weaker mean exchangeability assumptions, which assumes that the target and auxiliary populations are homogeneous when conditioned on the observed covariates V.

Assumption 5 (Homogeneous CATE).

$$\beta_0(V) = \beta_1(V).$$

Similar assumptions have been widely adopted in the literature for generalizing findings from randomized controlled trials to the target population. For example, Degtiar and Rose (2023), Shi et al. (2023), and Colnet et al. (2024) incorporated these assumptions in their frameworks. Here, we adapt Assumption 5 to the data fusion setting involving IVs. It requires covariates to capture all treatment effect modifiers that are shifted between the two populations. It would be satisfied if the auxiliary sample is randomly drawn from the target population or if all shifting effect modifiers are measured.

Lemma 1 .

Suppose that Assumptions 1 and 3 for $R = 1$, Assumptions 2, 4, and 5 hold. Then τ is identified by $\tau = \tau^h$, where

$$\tau^h = E\{\beta_1(V)| R = 0\}, \quad (2)$$

and the superscript "h" indicates the homogeneous CATE assumption.

2.3 Identification under equi-confounding

The homogeneous CATE assumption has been widely used in existing data fusion methods, but it may be violated when the auxiliary data come from heterogeneous populations. Below we introduce a novel identification approach that allows for different CATEs between the target and auxiliary populations.

Assumption 6 (Equi-confounding).

$$\text{Cov}(X, Y_0 | V, R = 0) = \text{Cov}(X, Y_0 | V, R = 1).$$

The covariance $\text{Cov}(X, Y_0 | V, R)$ in Assumption 6 measures unmeasured confounding bias. Notably, when there is no unmeasured confounding, it is zero. Assumption 6 posits that, conditional on observed covariates, the magnitude of unmeasured confounding bias is the same in the target and auxiliary populations. Under Assumptions 1 and 3, we show in Lemma S2 of the supplementary material that $\text{Cov}(X, Y_1 - Y_0 | V, R) = 0$, which implies $\text{Cov}(X, Y_0 | V, R) = \text{Cov}(X, Y_1 | V, R)$ for both populations. Therefore, requiring the equality of the covariance for Y_0 in Assumption 6 is equivalent to requiring it for Y_1.

Assumption 6 would be testable if IV data were available in the target population. However, we focus on a more challenging and realistic setting where IV data are only observed in the auxiliary population. In our setting, this assumption is not testable. Assumption 6 is reasonable when the key unmeasured confounders operate through similar mechanisms in both populations. This is more likely if the confounders are stable characteristics such as genetic traits, rather than context-dependent factors. Additionally, when rich observed covariates account for major differences between two populations, the remaining confounding bias is less likely to vary significantly.

The following example provides some sufficient conditions for Assumptions 3 and 6. Let $\varepsilon(V, U) = E(X | V, U, R = 1) - E(X | V, U, R = 0)$ denote the difference of the treatment propensity scores given all confounders (V, U) beween the two populations.

Example 1 .

Suppose $U \perp\!\!\!\perp R \mid V$ and the following structural equation model holds:

$$E(Y \mid X, V, U, R) = \zeta_0(V, U) + \zeta(V)R + \beta_R(V)X, \ (3)$$

where $\mathrm{Cov}\{\zeta_0(V,U), \varepsilon(V,U) \mid V, R\} = 0$, and $\{\zeta_0(\cdot), \zeta(\cdot)\}$ are unknown scalar functions. Then under Assumption 1, Assumptions 3 and 6 are satisfied.

Since the variables (X, R) are both binary, model (3) is equivalent to $E(Y \mid X, V, U, R) = \zeta_0(V, U) + g(X, V, R)$ for an arbitrary function $g(\cdot)$. Model (3) essentially requires that there are no additive U-R and U-X interactions. A sufficient condition for the zero-covariance constraint under model (3) is $\varepsilon(V, U) = \varepsilon(V)$, which implies no additive U-R interaction in predicting the treatment conditional on observed covariates. As discussed earlier, similar no-interaction conditions appeared in Wang and Tchetgen Tchetgen (2018) and Tchetgen Tchetgen et al. (2021).

Now we present the identification formula of τ under the equi-confounding assumption.

Theorem 1 .

Suppose that Assumptions 1–3, 4, and 6 hold. Then $\beta_0(V)$ is identified as

$$\beta_0(V) = E\left\{ \frac{\beta_1(V)\sigma_1^2(V)}{\sigma_0^2(V)} - \frac{(2R-1)\eta Y}{f(R \mid V)\sigma_0^2(V)} \bigg| V \right\}, \ (4)$$

where all notation is defined in Table 1. In addition, τ is identified by $\tau = \tau^e$, where

$$\tau^{\mathrm{e}} = E\left[\frac{(1-R)\beta_1(V)\sigma_1^2(V)}{q\sigma_0^2(V)} - \left\{ \frac{R}{f(R \mid V)} - 1 \right\} \frac{\eta Y}{q\sigma_0^2(V)} \right], \ (5)$$

and the superscript "e" indicates the equi-confounding assumption.

Theorem 1 provides a novel identification result about $\beta_0(V)$ under equi-confounding. It provides one identification formula for τ. There are alternative identification formulas based on different components of the observed data likelihood, which are omitted due to space constraints.

The IV Assumption 2 is specifically imposed to identify the CATE in the auxiliary population $\beta_1(V)$, which can be replaced by any alternative conditions as long as $\beta_1(V)$ is guaranteed to be identifiable. For example, if the unmeasured confounders U are fully observable in the auxiliary population, as considered in Yang and Ding (2020), then under Assumption 1, $\beta_1(V)$ can be identified as

$$\beta_1(V) = E\{E(Y \mid X = 1, V, U, R = 1) - E(Y \mid X = 0, V, U, R = 1) \mid V, R = 1\}.$$

Similarly, when negative control variables are available in the auxiliary population, $\beta_1(V)$ can be also identifiable under certain conditions (Miao et al., 2018; Tchetgen Tchetgen et al., 2024). In these scenarios, the identification expressions of τ in Lemma 1 and Theorem 1 remain unchanged except that the expression of $\beta_1(V)$ therein is updated.

In proving Theorem 1, we express $\beta_0(V)$ as:

$$\beta_0(V) = \beta_0^{\text{ols}}(V) - \{\beta_1^{\text{ols}}(V) - \beta_1(V)\} \frac{\sigma_1^2(V)}{\sigma_0^2(V)}, \quad (6)$$

where $\beta_r^{\text{ols}}(V)$ denotes the ordinary least squares (OLS) estimand for the coefficient of X obtained by the simple regression of Y on X, stratified by V and $R = r$, i.e., $\beta_r^{\text{ols}}(V) = \sigma_r^{-2}(V)\text{Cov}(X, Y \mid V, R = r)$ for $r = 0, 1$. If covariates V are omitted and the variances of the treatment are equal across populations such that $\sigma_1^2 = \sigma_0^2$, equation (6) simplifies to:

$$\beta_0 = \beta_0^{\text{ols}} - (\beta_1^{\text{ols}} - \beta_1). \quad (7)$$

This simplification offers a more intuitive interpretation of the identification result in Theorem 1. Specifically, the auxiliary data enable us to express the unmeasured confounding bias using observable quantities (i.e., $\beta_1^{\text{ols}} - \beta_1$), which is then subtracted from the identified treatment-outcome association β_0^{ols} in the target population to yield an expression for the causal effect. This transportability strategy relies on the equi-confounding Assumption 6. Because Assumption 6 parallels the additive equi-confounding bias assumptions used in identification approaches based on difference-in-differences or negative outcome controls (Card and Krueger, 1994; Angrist and Pischke, 2009; Lipsitch et al., 2010; Tchetgen Tchetgen, 2014; Sofer et al., 2016), the identification formula in (7) naturally resembles a difference-in-differences adjustment.

Remark 1 .

In general, the variance ratio term $\sigma_1^2(V) / \sigma_0^2(V)$ in the identification formula (6) cannot be omitted under our current equi-confounding Assumption 6, as doing so would lead to bias. To eliminate this ratio, one might consider replacing Assumption 6 with the following alternative:

$$\beta_0^{\text{ols}}(V) - \beta_0(V) = \beta_1^{\text{ols}}(V) - \beta_1(V). \quad (8)$$

Assumption (8) may appear reasonable and ensures a simpler identification formula. However, it is difficult to find interpretable conditions under which the two sides are equal. In particular, when the homogeneous CATE assumption holds, (8) further requires that $\beta_0^{\text{ols}}(V) = \beta_1^{\text{ols}}(V)$. For example, suppose that Assumption 1 holds, and consider the following model:

$R, V, U \overset{iid}{\sim} \text{Bernoulli}(p),$

$$E(Y|\ X, V, U, R) = \zeta_0 + \zeta_1 R + \zeta_2 V + \zeta_3 U + \beta X, \qquad (9)$$

$$E(X|\ Z, V, U, R) = \iota_0 + \iota_1 R + \iota_2 V + \iota_3 U + \iota_4 RZ.$$

Since the IV Z is only observed when $R = 1$, the model for X includes the interaction term RZ. We assume that R, V, U follow the same Bernoulli distribution for simplicity. Under (9), we have $\beta_0(V) = \beta_1(V) = \beta$. Then (8) requires $\beta_0^{\text{ols}}(V) = \beta_1^{\text{ols}}(V)$. In fact, we have $\beta_r^{\text{ols}}(V) = p(1-p)\zeta_3 \iota_3 \sigma_r^2(V)$ under model (9). Therefore, (8) does not hold if $\zeta_3 \iota_3 \neq 0$ and $\sigma_1^2(V) \neq \sigma_0^2(V)$. The inequality in variance occurs when $\iota_1 \neq 0$ or $\iota_4 \neq 0$. In contrast, Assumption 6 is satisfied under (9). More generally, Assumption 6 is implied by the conditions in Example 1, whereas (8) typically does not hold in that setting. Details are given in Section S3.4 of the supplementary material.

Recently, Li et al. (2025) proposed a method to incorporate weakly aligned data sources that do not satisfy selection exchangeability or the weaker homogeneous CATE assumptions. Their approach, based on selection bias models, requires at least one known auxiliary data source to properly align with the target population, ensuring the identification of the target parameter. Their goal is to improve statistical efficiency by incorporating additional weakly aligned data sources. In contrast, our equi-confounding approach does not require the auxiliary data source to be pre-aligned with the target population. Instead, our approach establishes a novel identification result based on the assumption that the unmeasured confounding bias remains stable across populations. This paper also bears some connection to Li (2024), particularly Example 3 in their work. While Li (2024) assumed that the confounding bias lies in a smooth, finite-dimensional parametric class, we allow the bias to be an arbitrary and possibly nonparametric function.

2.4 Comparison

The identification strategies in Lemma 1 and Theorem 1 offer two distinct approaches for transporting causal effects from the auxiliary population to the target population. The homogeneous CATE assumption directly equates the conditional treatment effects across populations, i.e., $\beta_0(V) = \beta_1(V)$. In contrast, the equi-confounding assumption posits that unmeasured confounding biases are the same in the two populations, conditional on observed covariates. This assumption does not require the CATEs to be equal across populations. As shown in (7) for the simplified setting where covariates are omitted and the variances of the treatment are the same in the two populations, the equi-confounding approach incorporates a difference-in-differences type correction. It leverages additional information from OLS estimates and explicitly accounts for confounding bias using data from the auxiliary population, thereby enabling effect transportability even when treatment effect heterogeneity exists.

The choice between these two approaches depends on the study context. The homogeneous CATE assumption works well when treatment effect heterogeneity is small or explained by observed covariates. The equi-confounding assumption is more appropriate when unmeasured confounders operate similarly in both populations, even if the CATEs differ. Since these assumptions are untestable and not mutually exclusive, they may both hold in certain settings. If domain knowledge strongly supports one assumption, the corresponding

approach should be preferred. Otherwise, researchers should carefully consider which assumption aligns best with the data and study design. When possible, applying both approaches and comparing results, as demonstrated in our real data analysis, can provide additional robustness and strengthen conclusions.

Since both approaches rely on key transportability assumptions, we can conduct sensitivity analyses to evaluate their robustness. To assess the impact of potential violation of the homogeneous CATE Assumption 5, we can introduce the sensitivity parameter as the difference of the two CATEs across populations and modify the identification formula in Lemma 1 accordingly to conduct sensitivity analysis. To assess the impact of potential violations of equi-confounding Assumption 6, we introduce the following sensitivity parameter: $\xi(V) = \text{Cov}(X, Y_0 | V, R = 0) - \text{Cov}(X, Y_0 | V, R = 1)$, which captures the difference in unmeasured confounding bias between the target and auxiliary populations. Under Assumption 6, we have $\xi(V) = 0$, and hence, a large absolute value of $\xi(V)$ indicates a strong deviation from this assumption. Given $\xi(V)$, $\beta_0(V)$ remains identifiable even without Assumption 6, using a formula similar to (4), except that the numerator of the first term is modified to $\beta_1(V)\sigma_1^2(V) - \xi(V)$ to account for the violation. Further details on sensitivity analysis are provided in Section S2 of the supplementary material.

3 Multiply robust estimation

In this section, we focus on the statistical estimands τ^h and τ^e, which can be interpreted as the ATE in the target population under the identification assumptions. While the nuisance models in Lemma 1 and Theorem 1 can be estimated nonparametrically using methods like kernel smoothing or series estimation, these approaches may suffer from poor performance in moderate- or high-dimensional settings due to the curse of dimensionality. To construct estimators with better finite-sample properties, we turn to semiparametric efficiency theory for estimation. Any regular and asymptotically linear estimator has an influence function, and the one with the lowest variance is known as the efficient influence function (EIF) (Bickel et al., 1993). The EIF is central to the semiparametric theory, as its variance defines the semiparametric efficiency bound. It also provides the basis for constructing locally efficient, multiply robust (MR) estimators, which maintain consistency even when some components of the observed data distribution are misspecified, making them particularly useful in practice.

We consider the nonparametric observed data model $\mathcal{M} = \{f(O) : f(O) \text{ is unrestricted}\}$. We refer to $\mathcal{M}$ as the nonparametric model because no parametric models are imposed in $\mathcal{M}$ except some positivity and support overlap assumptions. In the following subsections, we derive the EIFs for the statistical estimands τ^h and τ^e under $\mathcal{M}$, relying only on standard conditions such as positivity and support overlap in the observed data distribution. These conditions do not restrict the model structure and are distinct from the stronger assumptions imposed at the identification stage.

3.1 MR estimation of τ^h

In this subsection, we develop the MR estimator of τ^h expressed in (2). The estimator relies on the results in Wang and Tchetgen Tchetgen (2018). Therefore, this section effectively serves as a review of their MR estimator, adapted to our data fusion setting. We first derive

the EIF of τ^{h} in $\mathcal{M}$, and let $\Delta^{\mathrm{h}} = (\pi, \mu_0^X, \mu_0^Y, \delta^X, \beta_1, \omega)$ denote the set of all nuisance functions associated with this EIF. Among these, the two conditional marginal distributions $Y \mid Z, V, R = 1$ and $X \mid Z, V, R = 1$ are not derived from a natural factorization of the joint distribution, although they still remain variationally independent by Sklar's theorem (Sklar, 1959).

Lemma 2 .

The EIF for τ^{h} in $\mathcal{M}$ is given by $\psi_{\mathrm{eff}}(O; \Delta^{\mathrm{h}}, q) - q^{-1}(1-R)\tau^{\mathrm{h}}$, where

$$\psi_{\mathrm{eff}}(O; \Delta^{\mathrm{h}}, q) = \frac{1-R}{q}\beta_1(V) + \frac{\gamma(V)}{q}\psi_1(O; \Delta^{\mathrm{h}}),$$

and

$$\psi_1(O; \Delta^{\mathrm{h}}) = \frac{R(2Z-1)}{f(Z \mid V, R=1)}\frac{1}{\delta^X(V)}\{Y - \mu_0^Y(V) - X\beta_1(V) + \mu_0^X(V)\beta_1(V)\}.$$

The EIF in Lemma 2 shares a similar structure with Theorem 2 of Li and Luedtke (2023). Under the additional requirement that the target distribution aligns with the observed distributions from both data sources in a specific manner, our result can be viewed as a special case when combined with the EIF derived in Wang and Tchetgen Tchetgen (2018). However, our formulation does not require this alignment. For completeness, we provide a direct and self-contained proof of Lemma 2 in the supplementary material. The EIF must have mean zero and therefore, Lemma 2 implies that $\tau^{\mathrm{h}} = E\{\psi_{\mathrm{eff}}(O; \Delta^h, q)\}$. Let $\mathbb{P}_n$ denote the empirical average, for example, $\mathbb{P}_n\{g(O)\} = n^{-1}\sum_{i=1}^{n}g(O_i)$ for any function $g(\cdot)$. When nuisance functions in Δ^{h} have been estimated by $\hat{\Delta}^{\mathrm{h}}$ and $\hat{q} = n^{-1}\sum_{i=1}^{n}(1-R_i)$, we can construct a MR estimator of τ^{h} as

$$\hat{\tau}_{\mathrm{mr}}^{\mathrm{h}} = \mathbb{P}_n\{\psi_{\mathrm{eff}}(O; \hat{\Delta}^{\mathrm{h}}, \hat{q})\}. \quad (10)$$

The strategy of constructing estimators from estimating equations based on influence functions has been widely used in causal inference literature (Wang and Tchetgen Tchetgen, 2018; Shi et al., 2020; Sun et al., 2022; Sun and Miao, 2022; Jiang et al., 2022).

Following Wang and Tchetgen Tchetgen (2018), we apply a parametric model-based approach to estimate the nuisance functions in Δ^{h}. Specifically, we consider parametric models $\Delta^{\mathrm{h}}(\Theta_1)$ with $\Theta_1 = (\theta_1^{\mathrm{T}}, \ldots, \theta_6^{\mathrm{T}})^{\mathrm{T}}$, where θ_j is a vector of finite-dimensional parameters associated with the jth nuisance model in Δ^{h} for $j = 1, \ldots, 6$; that is, $\Delta^{\mathrm{h}}(\Theta_1) = \{\pi(v; \theta_1), \mu_0^X(v; \theta_2), \mu_0^Y(v; \theta_3), \delta^X(v; \theta_4), \beta_1(v; \theta_5), \omega(v; \theta_6)\}$. We first use maximum likelihood estimation to obtain estimators $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_6$ of the parameters in the nuisance models $\pi, \mu_0^X, \mu_0^Y, \omega$, because these nuisances are either models for binary variables or simply have conditional mean forms. Next, similar to Wang and Tchetgen Tchetgen (2018),

we solve the following estimating equations to obtain estimators $\hat{\theta}_4$ and $\hat{\theta}_5$ of the parameters associated with models δ^X and β_1:

$$\mathbb{P}_n[D_4(V)\{X - \delta^X(V;\theta_4)Z - \mu_0^X(V;\hat{\theta}_2)\}\frac{R(2Z-1)}{f(Z|V,R=1;\hat{\theta}_1)}] = 0, \quad (11)$$

$$\mathbb{P}_n[D_5(V)\{Y - \beta_1(V;\theta_5)X - \mu_0^Y(V;\hat{\theta}_3) + \mu_0^X(V;\hat{\theta}_2)\beta_1(V;\theta_5)\}\frac{R(2Z-1)}{f(Z|V,R=1;\hat{\theta}_1)}] = 0, \quad (12)$$

where $D_j(V)$ is a user-specified vector function of the same dimension as θ_j for $j = 4, 5$. Let $\hat{\Theta}_1 = (\hat{\theta}_1^T, \ldots, \hat{\theta}_6^T)^T$ and $\hat{\Delta}^h = \Delta^h(\hat{\Theta}_1)$. Then the MR estimator $\hat{\tau}_{mr}^h$ under homogeneous CATE is given by (10).

Lemma 3 .

The estimator $\hat{\tau}_{mr}^h$ is consistent and asymptotically normal under standard regularity conditions if one of the following holds:

(i) $\mathcal{M}_1^h$: models for (π, δ^X, ω) are correct;

(ii) $\mathcal{M}_2^h$: models for $(\mu_0^X, \mu_0^Y, \beta_1)$ are correct;

(iii) $\mathcal{M}_3^h$: models for (π, β_1) are correct.

In addition, if all the conditions in (i)–(iii) hold, then $\hat{\tau}_{mr}^h$ attains the semiparametric efficiency bound.

Like Wang and Tchetgen Tchetgen (2018), which considered a single dataset with IVs, Lemma 3 shows that we also obtain a triply robust estimator in the current data fusion setting under the homogeneous CATE assumption. This estimator ensures consistency when one of the three model sets is correctly specified. However, since our focus is on data fusion, our first model set $\mathcal{M}_1^h$ includes an additional model for the data source mechanism ω. The other two model sets are similar to those in Wang and Tchetgen Tchetgen (2018).

3.2 MR estimation of τ^e

In this subsection, we develop the MR estimator of τ^e expressed in (5). Similar to Section 3.1, we first derive the EIF for τ^e in $\mathcal{M}$. Let $\Delta^e = (\Delta^h, \mu_0, \beta_0, \phi_0, \rho)$ denote the set of all nuisance functions associated with this EIF. The nuisance functions in Δ^e are variationally independent. Using the nuisances in Δ^e, we can express the functions $\mu_1(v)$ and $\phi_1(v)$ as

$$\mu_1(v) = \delta^X(v)\pi(v) + \mu_0^X(v), \quad \phi_1(v) = \mu_0^Y(v) - \beta_1(v)\mu_0^X(v). \quad (13)$$

Theorem 2 .

The EIF for τ^e in $\mathcal{M}$ is given by $\varphi_{\mathrm{eff}}(O; \Delta^e, q) - q^{-1}(1-R)\tau^e$, where

$$
\varphi_{\mathrm{eff}}(O; \Delta^e, q) = \frac{1-R}{q}\beta_0(V) - \{\frac{R}{f(R|V)} - 1\}\frac{\eta\{Y - \beta_R(V)X - \phi_R(V)\} - \rho(V)}{q\sigma_0^2(V)}
$$
$$
+ \frac{\gamma(V)\sigma_1^2(V)}{q\sigma_0^2(V)}\psi_1(O; \Delta^h),
$$

and $\psi_1(O, \Delta^h)$ is defined in Lemma 2.

The function $\psi_1(O; \Delta^h)$ involves only the auxiliary sample and also appears in the EIF for τ^h, as shown in Lemma 2. Compared with Lemma 2, the expression of the EIF for τ^e presented here is more complicated. Under homogeneous CATE with $\beta_0(V) = \beta_1(V)$, the first term of the EIF in Theorem 2 coincides with that in Lemma 2. However, because we do not impose the homogeneous CATE assumption when identifying τ via (5), the corresponding EIF includes additional functions, as shown in the second term of $\varphi_{\mathrm{eff}}(O; \Delta^e, q)$. Define the following semiparametric models that impose certain restrictions on the observed data distribution:

$\mathcal{M}_1^e$: models for (π, μ_0^X, δ^X) and (ω, μ_0) are correct;

$\mathcal{M}_2^e$: models for $(\mu_0^X, \mu_0^Y, \beta_1)$ and (ω, μ_0) are correct;

$\mathcal{M}_3^e$: models for (π, μ_0^X, δ^X) and $(\omega, \beta_0, \phi_0)$ are correct;

$\mathcal{M}_4^e$: models for $(\mu_0^X, \mu_0^Y, \beta_1)$ and $(\omega, \beta_0, \phi_0)$ are correct;

$\mathcal{M}_5^e$: models for (π, μ_0^X, δ^X) and (μ_0, β_0, ρ) correct;

$\mathcal{M}_6^e$: models for $(\mu_0^X, \mu_0^Y, \beta_1)$ and (μ_0, β_0, ρ) correct.

Let $\mathcal{M}_{\mathrm{union}}^e = \cup_{k=1}^6 \mathcal{M}_k^e$ denote the semiparametric model that holds if any of $\{\mathcal{M}_k^e\}_{k=1}^6$ is true. We can show that if $\cup_{k \neq 5} \mathcal{M}_k^e$ holds, then $\cup_{k=1}^3 \mathcal{M}_k^h$, as defined in Lemma 3, also holds. However, if only $\mathcal{M}_5^e$ holds, then $\cup_{k=1}^3 \mathcal{M}_k^h$ may not be true. Similar to Section 3.1, we consider modeling Δ^e using standard parametric models $\Delta^e(\Theta)$ with $\Theta = (\Theta_1^T, \theta_7^T, \theta_8^T, \theta_9^T, \theta_{10}^T)^T$, where Θ_1 is defined in Section 3.1 and θ_j is a vector of finite-dimensional parameters associated with the jth nuisance model in Δ^e for $j = 7, \ldots, 10$; that is, $\Delta^e(\Theta) = \{\Delta^h(\Theta_1), \mu_0(v; \theta_7),$

$\beta_0(v; \theta_8), \phi_0(v; \theta_9), \rho(v; \theta_{10})\}$, with $\Delta^h(\cdot)$ defined in Section 3.1.

Below we develop a three-step procedure to estimate the nuisance parameters. In the first step, besides the estimators $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_6$ obtained in Section 3.1, we also use maximum likelihood estimation to obtain the estimator $\hat{\theta}_7$ of the parameter in the nuisance model μ_0. In the second step, we obtain estimators $\hat{\theta}_4$ and $\hat{\theta}_5$ of the parameters associated with models δ^X and β_1 by solving (11) and (12), respectively. We use $\hat{\pi}(v)$ to simply denote $\pi(v; \hat{\theta}_1)$, and other estimated nuisance models are denoted similarly. After obtaining the estimators of the first seven nuisance functions in Δ^e, we then plug them into (13) to obtain the estimators $(\hat{\mu}_1, \hat{\phi}_1)$ of (μ_1, ϕ_1). Similarly, we can obtain the estimators $(\hat{\eta}, \hat{\sigma}_1^2, \hat{\gamma})$ by plugging the estimated nuisance functions into the corresponding expressions of $(\eta, \sigma_1^2, \gamma)$ defined in Table 1; that is,

$$\hat{\eta} = X - R\hat{\mu}_1(V) - (1-R)\hat{\mu}_0(V), \ \hat{\sigma}_1^2(V) = \hat{\mu}_1(V)\{1 - \hat{\mu}_1(V)\}, \ \hat{\gamma}(V) = \frac{1 - \hat{\omega}(V)}{\hat{\omega}(V)}.$$

Then we obtain a plug-in estimator of $\psi_1(O; \Delta^h)$ defined in Lemma 2 and denote it by $\hat{\psi}_1$.

In the third step, we estimate the remaining nuisance models (β_0, ϕ_0, ρ). We first define two functions based on estimates of some nuisance models:

$$\hat{\beta}(\theta_8) = R\hat{\beta}_1(V) + (1-R)\beta_0(V; \theta_8), \qquad \hat{\phi}(\theta_9) = R\hat{\phi}_1(V) + (1-R)\phi_0(V; \theta_9).$$

Then we solve the following equations in (14)–(16) to obtain estimators $\hat{\theta}_8$, $\hat{\theta}_9$, and $\hat{\theta}_{10}$ of the parameters associated with the three nuisance models β_0, ϕ_0, and ρ:

$$\mathbb{P}_n[D_8(V)\{(\frac{R}{\hat{f}(R|V)} - 1)(\hat{\eta}(Y - \hat{\beta}(\theta_8)X - \hat{\phi}(\theta_9)) - \rho(V; \theta_{10})) - \hat{\gamma}(V)\hat{\sigma}_1^2(V)\hat{\psi}_1\}] = 0, \ (14)$$

$$\mathbb{P}_n[D_9(V)(1-R)\{Y - \beta_0(V; \theta_8)X - \phi_0(V; \theta_9)\}] = 0, \ (15)$$

$$\mathbb{P}_n[D_{10}(V)(1-R)\{(X - \hat{\mu}_0(V))(Y - \beta_0(V; \theta_8)X) - \rho(V; \theta_{10})\}] = 0, \ (16)$$

where $\hat{f}(R|V) = R\hat{\omega}(V) + (1-R)\{1 - \hat{\omega}(V)\}$, and $D_j(V)$ is a user-specified vector function of the same dimension as θ_j for $j = 8, 9, 10$. The following lemma shows consistency of the proposed estimators for the nuisance parameters under certain conditions.

Lemma 4 .

Under standard regularity conditions, the estimators obtained from the three-step procedures for the parameters in $\mathcal{M}_j^e$ are consistent if $\mathcal{M}_j^e$ holds, for $j = 1,\dots,6$.

Let $\hat{\Theta} = (\hat{\Theta}_1^T, \hat{\theta}_7^T, \hat{\theta}_8^T, \hat{\theta}_9^T, \hat{\theta}_{10}^T)^T$ and $\hat{\Delta}^e = \Delta^e(\hat{\Theta})$. Since the EIF has mean zero, Theorem 2 implies that $\tau^e = E\{\varphi_{\text{eff}}(O; \Delta^e, q)\}$. Accordingly, the corresponding MR estimator of τ^e is given by $\hat{\tau}_{\text{mr}}^e = \mathbb{P}_n\{\varphi_{\text{eff}}(O; \hat{\Delta}^e, \hat{q})\}$.

Theorem 3 .

The estimator $\hat{\tau}_{\text{mr}}^e$ is a consistent and asymptotically normal estimator of τ^e in $\mathcal{M}_{\text{union}}^e = \cup_{k=1}^6 \mathcal{M}_k^e$ under standard regularity conditions. In addition, if all the models in Δ^e are correct, then $\hat{\tau}_{\text{mr}}^e$ attains the semiparametric efficiency bound.

The estimator $\hat{\tau}_{\text{mr}}^e$ offers six genuine opportunities to estimate τ^e consistently and also has valid inference provided that one of the six strategies holds. Under particular specifications of the nuisance models, $\hat{\tau}_{\text{mr}}^e$ reduces to other different semiparametric estimators. There are six such estimators $\hat{\tau}_1^e, \dots, \hat{\tau}_6^e$, dependent on specific model specifications in $\mathcal{M}_1^e, \dots, \mathcal{M}_6^e$, respectively. Under standard regularity conditions, $\hat{\tau}_j^e$ is consistent and asymptotically normal in $\mathcal{M}_j^e$ for $j = 1,\dots,6$. More details about these semiparametric estimators can be found in Section S1 of the supplementary material.

3.3 Flexible MR estimation

An important advantage of MR estimators constructed using the EIF is that they often yield second-order bias terms, which allow for slow convergence rates for the nuisance parameters involved, thereby enabling the use of flexible regression techniques in estimating these quantities (Chernozhukov et al., 2018; Kennedy, 2022).In this section, we introduce the flexible MR estimation about τ^e using machine learning methods. A similar estimation approach applies to the other estimand τ^h.

To construct our estimator, we use cross-fitting for nuisance estimation. Let K be a fixed positive integer. We divide the data into K groups, train nuisance models on all but one fold, and apply the estimated nuisances only to the hold-out fold. Specifically, we create K-fold random partitions for the auxiliary data $\{i : R_i = 1\}$ and the target data $\{i : R_i = 0\}$, with index sets $\{\mathcal{I}_k^1\}_{k=1}^K$ and $\{\mathcal{I}_k^0\}_{k=1}^K$, respectively. The combined partition $\{\mathcal{I}_k = \mathcal{I}_k^1 \cup \mathcal{I}_k^0\}_{k=1}^K$ divides the entire sample $\{1,\dots,n\}$ into subsets of sizes $\{n_k\}_{k=1}^K$, and we define the complement set $\mathcal{I}_{-k} = \{1,\dots,n\} \setminus \mathcal{I}_k$ with size n_{-k}. The empirical means in the kth partition and the remaining sample are denoted by $\mathbb{P}_{n_k}$ and $\mathbb{P}_{n_{-k}}$, respectively, and we define $\hat{q}_{-k} = \mathbb{P}_{n_{-k}}(1 - R)$.

For each k, we estimate the nuisance functions using the data with index set $\mathcal{I}_{-k}$. Let $\hat{\Delta}_{-k}^e$ denote the machine learning estimators for nuisance models. First, we obtain estimators

$(\hat{\pi}_{-k}, \hat{\mu}_{0,-k}^X, \hat{\mu}_{0,-k}^Y, \hat{\omega}_{-k}, \hat{\mu}_{0,-k})$ for the conditional mean functions $(\pi, \mu_0^X, \mu_0^Y, \omega, \mu_0)$. Next, given $(\hat{\pi}_{-k}, \hat{\mu}_{0,-k}^X, \hat{\mu}_{0,-k}^Y)$, we estimate δ^X and β_1 using machine learning methods based on the following conditional mean relationships:

$$\delta^X(V) = E\left\{\frac{X - \mu_0^X(V)}{\pi(V)} \mid V, R = 1\right\}, \quad \beta_1(V) = E\left\{\frac{Y - \mu_0^Y(V)}{\pi(V)\delta^X(V)} \mid V, R = 1\right\}.$$

Then, using the estimators $(\hat{\pi}_{-k}, \hat{\mu}_{0,-k}^X, \hat{\omega}_{-k}, \hat{\mu}_{0,-k}, \hat{\delta}_{-k}^X, \hat{\beta}_{1,-k})$, we estimate β_0, ϕ_0, and ρ based on the expression in (4) and the definitions of (ϕ_0, ρ) in Table 1. Finally, the MR estimator of τ^e, using cross-fitting, is given by:

$$\hat{\tau}_{cf}^e = \frac{1}{K}\sum_{k=1}^K \mathbb{P}_{n_k}\{\varphi_{eff}(O; \hat{\Delta}_{-k}^e, \hat{q}_{-k})\}.$$

Let $\| g \| = \{\int g^2(o)d\mathbb{P}(o)\}^{1/2}$ denote the $L_2(\mathbb{P})$ norm of any function $g(\cdot)$. In the following theorem, we establish the asymptotic properties of the cross-fitting estimator $\hat{\tau}_{cf}^e$.

Theorem 4 .

For each k, we assume the following conditions hold:

(i) $\hat{\Delta}_{-k}^e(v) \to \Delta^e(v)$ *in probability for all v in the support of V ;*

(ii) $c < \{\hat{\pi}_{-k}(v), \pi(v)\} < 1 - c$, $\{\hat{\mu}_{0,-k}(v), \mu_0(v), \hat{\omega}_{-k}(v), \omega(v), |\hat{\delta}_{-k}^X(v)|, |\delta^X(v)|\} > c$, *and*
$\{|\hat{\beta}_{1,-k}(v)|, |\beta_1(v)|\} < C$ *for some $c \in (0,1)$, $C > 0$, and all v in the support of V ;*

(iii) $n^{1/2}$ *-convergence of second-order terms, i.e.,*

$$(\| \hat{\beta}_{1,-k} - \beta_1 \| + \| \hat{\mu}_{0,-k}^Y - \mu_0^Y \| + \| \hat{\mu}_{0,-k}^X - \mu_0^X \|)(\| \hat{\pi}_{-k} - \pi \| + \| \hat{\delta}_{-k}^X - \delta^X \| + \| \hat{\mu}_{0,-k}^X - \mu_0^X \|)$$
$$+ \| \hat{\mu}_{0,-k} - \mu_0 \| (\| \hat{\beta}_{0,-k} - \beta_0 \| + \| \hat{\phi}_{0,-k} - \phi_0 \|) + \| \hat{\omega}_{-k} - \omega \| \| \hat{\rho}_{-k} - \rho \| = o_\mathbb{P}(n^{-1/2}).$$

Then $\hat{\tau}_{cf}^e$ is asymptotically normal, and achieves the semiparametric efficiency bound.

Conditions (i)-(iii) are similar to those for double machine learning estimation of average treatment effects (Kennedy et al., 2017; Chernozhukov et al., 2018; Kennedy, 2022). These conditions are all common in standard non-parametric regression problems (Van der Vaart, 2000). Among them, condition (iii) is satisfied if all the nuisance parameters are estimated with root mean-squared error rates diminishing faster than $n^{-1/4}$. Such rates are achievable for many highly data-adaptive machine learning methods, including lasso, neural networks or ensembles of these methods.

4 Simulation studies

In this section, we conduct simulation studies to evaluate the finite-sample performance of the proposed estimators. In our simulation setting, baseline covariates $V = (V_1, V_2)$ are independently generated from standard normal distribution $N(0,1)$. The case with four-dimensional covariates is shown in Section S4.2 of the supplementary material. We consider the functional form $V_k^* = \{1 + \exp(1 - 2V_k)\}^{-1}$ for $k = 1, 2$. The unmeasured confounder U is generated from a uniform distribution $\text{Unif}\{-a(V^*), a(V^*)\}$, where $a^2(V^*) = 0.75 + 1.5V_1^* + 1.5V_2^*$. Then the sampling indicator R, instrumental variable Z in the auxiliary sample, treatment X, and outcome Y are separately generated by the following mechanism:

$$
\begin{aligned}
R \mid V \qquad &\sim \text{Bernoulli}\{ p = (1 + \exp(-0.8 - 0.5V_1^* + 2.5V_2^*))^{-1} \}, \\
Z \mid V, R = 1 \quad &\sim \text{Bernoulli}\{ p = (1 + \exp(0.5 + 0.5V_1^* - 3V_2^*))^{-1} \}, \\
X \mid Z, V, U, R \quad &\sim \text{Bernoulli}\{ p = R\alpha_1(Z) + (1 - R)\alpha_0(V) + 0.1U \}, \\
Y \mid X, V, U, R \quad &\sim N\{ (2 - \kappa R + 2V_1^* - 0.5V_2^*)(1 + X) + U, 0.5^2 \},
\end{aligned}
$$

where $\alpha_1(Z) = \{1 + \exp(-1.3 + 2.4Z)\}^{-1}$, $\alpha_0(V) = \{1 + \exp(-V_1^* + V_2^*)\}^{-1}$, and $\kappa = \beta_0(V) - \beta_1(V) \in \{0, 1, 2\}$ denotes the difference of CATEs between two populations. Under the above data generating mechanism, the equi-confounding assumption holds according to Example 1. We also observe that correct models for the nuisance models $(\pi, \mu_0^X, \omega, \mu_0)$ in Δ^e are given by a logistic regression with V^* as linear predictors, and correct models for other models $(\mu_0^Y, \delta^X, \beta_1, \beta_0, \phi_0, \rho)$ in Δ^e are given by a linear regression with V^* as predictors. We are interested in estimating the average causal effect τ, whose true value is 2.43. We evaluate the performance of the proposed estimators in situations where some models may be misspecified. In the misspecification case, the quadratic functional form $V_k^{**} = (V_k - 0.5)^2$ is used in place of V_k^* for $k = 1, 2$ when constructing the nuisance models in Δ^e. To estimate the nuisance models based on (11)-(12) and (14)-(16) given in Section 3, we set $D_j(V) = (1, V^*)^T$ if the model is correct and $D_j(V) = (1, V^{**})^T$ otherwise.

We apply the proposed MR estimators $\hat{\tau}_{mr}^h$, $\hat{\tau}_{mr}^e$ introduced in Section 3 and other semiparametric estimators $\{\hat{\tau}_j^e\}_{j=1}^6$ introduced in Section S1 of the supplementary material to estimate τ. We consider the following eight scenarios:

$\mathcal{S}_0$, all the models are correct;

$\mathcal{S}_j$, only the models in $\mathcal{M}_j^e$ are correct for $j = 1, \ldots, 6$.

$\mathcal{S}_7$, none of the models is correct.

Table 2 summarizes the simulation results for all estimators under the eight scenarios for $\kappa = 0$ and $n = 2000, 4000$, where the bias, standard deviation, and 95% coverage probability

of these estimators averaged across 1000 replications are reported. As expected, the proposed estimator $\hat{\tau}_{\text{mr}}^{\text{e}}$ exhibits small bias and has coverage probability close to the nominal level in $\{\mathcal{S}_j\}_{j=0}^6$, confirming its multiple robustness property. In contrast, the other six estimators $\{\hat{\tau}_j^{\text{e}}\}_{j=1}^6$ generally show substantial bias in scenarios where their respective models are misspecified. For example, the estimator $\hat{\tau}_1^{\text{e}}$ performs well in scenarios $\{\mathcal{S}_j\}_{j=0,1}$, while exhibiting notable bias in scenarios $\{\mathcal{S}_j\}_{j=2,3,4,6}$. We also observe that $\hat{\tau}_1^{\text{e}}$ has comparable performance in some scenarios where models are misspecified, e.g., $\mathcal{S}_5$, to those in scenarios where models are correct. Similar phenomenon may appear in specific scenarios for other estimators, which is subject to the simulation setting. In general, these six estimators can maintain their consistency only when their corresponding models are correct. Since $\kappa = 0$, the homogeneous CATE assumption holds. We observe that the other MR estimator $\hat{\tau}_{\text{mr}}^{\text{h}}$ performs similarly to $\hat{\tau}_{\text{mr}}^{\text{e}}$ in scenarios $\{\mathcal{S}_j\}_{j \neq 5}$. This is because in these scenarios, at least one of the conditions $\{\mathcal{M}_j^{\text{h}}\}_{j=1}^3$ is true. As a result, $\hat{\tau}_{\text{mr}}^{\text{h}}$ is consistent according to Lemma 3. However, it shows significant bias and low coverage probability in scenario $\mathcal{S}_5$, because none of the conditions $\{\mathcal{M}_j^{\text{h}}\}_{j=1}^3$ holds in this scenario. When $\kappa \in \{1, 2\}$, the homogeneous CATE assumption does not hold. The simulation results for $\hat{\tau}_{\text{mr}}^{\text{e}}$ and $\{\hat{\tau}_j^{\text{e}}\}_{j=1}^6$ in these two cases are similar to those when $\kappa = 0$. However, $\hat{\tau}_{\text{mr}}^{\text{h}}$ exhibits significant biases in all scenarios of these two cases. Due to space constraints, the results for $\kappa \in \{1, 2\}$ are provided in Section S4.1 of the supplementary material.

We also implement the proposed MR estimator by using machine learning methods for nuisance model estimation. Specifically, we consider estimators $\hat{\tau}_{\text{cf-la}}^{\text{e}}$ and $\hat{\tau}_{\text{cf-sl}}^{\text{e}}$, whereby the nuisance models are estimated by lasso and super learner, respectively. Figure 1 shows the corresponding results about mean squared error (MSE) and 95% coverage probability of these two MR machine learning-based estimators, averaged across 1000 repeated experiments with sample size from 2000 to 20000. We observe that as sample size increases, the MSEs of both estimators become smaller, and the 95% coverage probabilities are more close to the nominal level. These results corroborate our previous theoretical findings and demonstrate the advantages of the proposed MR estimators. A simulation study for assessing the sensitivity analysis method proposed in Section S2 is given in Section S4.3 of the supplementary material.

5 Application

In this section, we provide an application of the proposed approaches by estimating the average causal effect of smoking on physical functional status, based on data from the household survey portion of Community Tracking Study for the years 1996–1997. This study is designed to provide information of health care system change and its effects on people in the United States (Kemper et al., 1996). Smoking is a risk factor of many diseases including heart disease, stroke, diabetes, and most cancers. These diseases affect physical functioning or physical disability. To mitigate possible unmeasured confounding bias, Leigh and Schembri (2004) used cigarette price as an IV to assess the impact of smoking on physical functional status. We found that cigarette price is a strong IV only for the lower-income group but is a weak IV for the higher-income group. The higher-income group is defined as

those whose incomes are in the top 20%, with the remaining classified as the lower-income group. Below we use the data from the lower-income group with strong IV to help estimate the effect of smoking on physical functional status in the higher-income group.

Let the treatment variable be $X = 1$ if the subject smokes every day or on some days, and $X = 0$ if the subject does not smoke at all. The outcome Y is physical functional status, measured by the Physical Component Summary of the SF-12 (Ware et al., 1995). The baseline covariates include age, gender, race, mental functional status, education, and health condition. However, there may still exist unmeasured variables confounding the relationship between smoking and physical functional status, such as genetic factors. Let the IV be $Z = 1$ if the cigarette price is greater than the median and $Z = 0$ otherwise. Since our purpose is to estimate the average causal effect among higher-income individuals, we treat this group as the target population. The target dataset consists of 4198 subjects, with 1123 having $X = 1$ and 3075 having $X = 0$. The auxiliary dataset includes 18603 lower-income subjects, with 7834 having $X = 1$ and 10769 having $X = 0$. While the IV (i.e., cigarette price) is available in the higher-income group, it shows weak correlation with the treatment. This is because smoking decisions among higher-income individuals may not be significantly influenced by cigarette price. In fact, after conducting a logistic regression of the treatment on the IV in the target data and adjusting for observed covariates, we found that the coefficient of the IV is not statistically significant.

We first apply the classical IV approach (Wang and Tchetgen Tchetgen, 2018) to estimate the effect of smoking on physical functional status using the weak IV available in the target data. As discussed in Section 2.1, the validity of this approach relies on the no-interaction Assumption 3, which posits that unmeasured confounders affect the outcome but do not modify the effect of treatment. This assumption appears plausible in our context. Genetic factors, as noted in prior work (e.g., Fisher (1958)), may confound the relationship between smoking and physical function. However, the biological mechanisms through which smoking impairs physical function such as reduced lung capacity operate relatively independently of genetic traits. Empirical literature also suggests that interaction effects in natural systems are typically smaller and less prevalent than main effects. Furthermore, our analysis adjusts for key effect modifiers, including age, gender, race, education, mental status, and health condition, which are major determinants of both smoking behavior and physical function. By accounting for these variables, the likelihood that remaining unmeasured confounders substantially modify the treatment effect is reduced. Together, these considerations support the plausibility of the no-interaction Assumption 3 in this setting. The point estimate using the weak IV, denoted as $\hat{\tau}_{\text{weak}}$, is -9.54 with a large standard error of 15.69. The corresponding confidence interval $(-40.30, 21.22)$ is very wide. This suggests that the weak IV approach may not yield a reliable estimate due to high variability. Fortunately, the IV is strongly correlated with the treatment in the lower-income group. The IV estimate for this group is -10.90, with a standard error of 3.55, leading to a confidence interval of $(-17.86, -3.94)$. These results suggest that smoking significantly decreases physical functional status in the lower-income population. The strong IV information from the lower-income group can therefore be leveraged below to help transport the causal effect of smoking to the higher-income group.

Next, we apply the two proposed transportability approaches. As a baseline, we first use the MR estimator $\hat{\tau}_{\text{mr}}^{\text{h}}$ under the homogeneous CATE assumption. The point estimate is -7.23, with a standard error of 3.98 and a confidence interval of $(-15.03, 0.56)$. Under the equi-

confounding assumption, we then implement the MR estimator $\hat{\tau}^{e}_{mr}$ along with two cross-fitting based estimators, $\hat{\tau}^{e}_{cf\text{-}la}$ and $\hat{\tau}^{e}_{cf\text{-}sl}$, which were also used in the simulation studies. Table 3 summarizes the results for all estimators of the average causal effect τ in the higher-income group. The three estimators based on the equi-confounding assumption, $\hat{\tau}^{e}_{mr}$, $\hat{\tau}^{e}_{cf\text{-}la}$, and $\hat{\tau}^{e}_{cf\text{-}sl}$, are close to each other. The validity of these estimators depends on Assumption 6, which appears reasonable in this application. First, possible unmeasured confounders such as genetic predisposition to smoking, are likely to influence both smoking behavior and physical function similarly across income groups. These factors are not expected to vary systematically between higher- and lower-income populations in a way that would change their confounding effect. Second, our analysis adjusts for important observed covariates that capture major socioeconomic and health-related differences. By accounting for these variables, the residual confounding bias is more likely to remain comparable across groups. We also conduct a sensitivity analysis for the equi-confounding assumption, with results provided in Section S4.4 of the supplementary material.

Now we compare and interpret the estimates from the above three approaches, including the weak IV approach and two transportability approaches. Table 3 shows that the equi-confounding based estimates align in direction with the weak IV estimate but have much smaller standard errors. Although the weak IV estimate is noisy, it serves as an unbiased benchmark. The close alignment between the equi-confounding based estimates and the weak IV result suggests that the transportability of unmeasured bias may be reasonable. Additionally, the homogeneous CATE assumption yields a similar effect estimate with the equi-confounding based estimators. The two transportability assumptions do not yield conflicting results in this example. The coherence of results across both approaches enhances confidence in the robustness of our conclusions. In summary, all point estimates are negative, and the upper bounds of the confidence intervals from all transportability estimators are slightly above zero. These findings indicate that smoking likely reduces physical functional status in higher-income individuals, with the effect approaching statistical significance.

6 Discussion

Instrumental variables are widely used to identify causal effects in the presence of unmeasured confounding. However, the stringent constraints for a valid IV may render such variables unavailable in the target population. This paper proposes two complementary approaches, referred to as the homogeneous CATE approach and the equi-confounding approach, to identify and estimate the average causal effect in the target population using IVs from an auxiliary population. For each approach, we have developed a semiparametric efficient multiply robust estimator and investigated its asymptotic properties, even when highly data-adaptive machine learning techniques are used for nuisance model estimation.

This paper can be extended or improved in several directions. First, our current analysis focuses on the canonical setting where both the instrumental and treatment variables are binary. Extending our framework to accommodate continuous instrumental and treatment variables poses significant challenges in terms of efficient estimation. In Section S5 of the supplementary material, we have developed a simplified estimation method under linear structural equation models. Second, the proposed framework can be extended to incorporate multiple auxiliary datasets with IVs, which may further enhance the efficiency of the proposed estimators. Finally, although both the homogeneous CATE and equi-confounding

assumptions are introduced for transporting causal effects, either assumption alone is sufficient to ensure identification. If both assumptions are assumed to hold simultaneously, additional constraints are placed on the observed data distribution. Under such constraints, one might explore more efficient estimation strategies within a semiparametric restricted model. We plan to explore these and other related topics in future work.

Supplementary material

The supplementary material includes additional semiparametric estimators under equi-confounding, proofs of all theoretical results, additional simulation studies, sensitivity analysis with implementations in both simulation and real data application, discussions on the case of continuous instrumental and treatment variables, and R code for reproducibility.

Acknowledgements

Conflict of Interest

The authors report there are no competing interests to declare.

References

Angrist, J., Imbens, G., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–455.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Baiocchi, M., Cheng, J., and Small, D. S. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine*, 33(13):2297–2340.

Bareinboim, E. and Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352.

Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press Baltimore.

Buchanan, A. L., Hudgens, M. G., Cole, S. R., et al. (2018). Generalizing evidence from randomized trials using inverse probability of sampling weights. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 181(4):1193–1209.

Card, D. and Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. *American Economic Review*, 84:772–793.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Colnet, B., Mayer, I., Chen, G., Dieng, A., Li, R., Varoquaux, G., Vert, J.-P., Josse, J., and Yang, S. (2024). Causal inference methods for combining randomized trials and observational studies: a review. *Statistical Science*, 39(1):165–191.

Dahabreh, I. J., Robertson, S. E., Steingrimsson, J. A., Stuart, E. A., and Hernan, M. A. (2020). Extending inferences from a randomized trial to a new target population. *Statistics in Medicine*, 39(14):1999–2014.

Degtiar, I. and Rose, S. (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application*, 10:501–524.

Fisher, R. A. (1958). Cancer and smoking. *Nature*, 182(4635):596–596.

Graham, B. S., Pinto, C. C. d. X., and Egel, D. (2016). Efficient estimation of data combination models by the method of auxiliary-to-study tilting (ast). *Journal of Business & Economic Statistics*, 34(2):288–301.

Hu, W., Wang, R., Li, W., and Miao, W. (2022). Paradoxes and resolutions for semiparametric fusion of individual and summary data. *arXiv preprint arXiv:2210.00200*.

Jiang, Z., Yang, S., and Ding, P. (2022). Multiply robust estimation of causal effects under principal ignorability. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1423–1445.

Kemper, P., Blumenthal, D., Corrigan, J. M., et al. (1996). The design of the community tracking study: a longitudinal study of health system change and its effects on people. *Inquiry*, 33:195–206.

Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1229–1245.

Leigh, J. P. and Schembri, M. (2004). Instrumental variables technique: cigarette price provided better estimate of effects of smoking on sf-12. *Journal of Clinical Epidemiology*, 57(3):284–293.

Li, H. H. (2024). Efficient combination of observational and experimental datasets under general restrictions on outcome mean functions. *arXiv preprint arXiv:2406.06941*.

Li, S., Gilbert, P. B., Duan, R., and Luedtke, A. (2025). Data fusion using weakly aligned sources. *Journal of the American Statistical Association*, pages 1–11.

Li, S. and Luedtke, A. (2023). Efficient estimation under data fusion. *Biometrika*, 110(4):1041–1054.

Li, W., Luo, S., and Xu, W. (2024). Calibrated regression estimation using empirical likelihood under data fusion. *Computational Statistics & Data Analysis*, 190:107871.

Lipsitch, M., Tchetgen, E. T., and Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383.

Luo, S., Zhang, Y., Li, W., and Geng, Z. (2025). Multiply robust estimation of causal effects using linked data. *Computational Statistics & Data Analysis*, 209:108175.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1):152–175.

Pearl, J. and Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, 29(4):579–595.

Ridder, G. and Moffitt, R. (2007). The econometrics of data combination. *Handbook of Econometrics*, 6:5469–5547.

Shi, X., Miao, W., Nelson, J. C., and Tchetgen Tchetgen, E. J. (2020). Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(2):521–540.

Shi, X., Pan, Z., and Miao, W. (2023). Data integration in causal inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, 15(1):e1581.

Shu, H. and Tan, Z. (2020). Improved methods for moment restriction models with data combination and an application to two-sample instrumental variable estimation. *Canadian Journal of Statistics*, 48(2):259–284.

Shuai, K., Luo, S., Li, W., and He, Y. (2024). Identifying causal effects using instrumental variables from the auxiliary population. *Statistica Sinica, to appear*.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8:229–231.

Sofer, T., Richardson, D. B., Colicino, E., Schwartz, J., and Tchetgen, E. J. T. (2016). On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Statistical Science*, 31(3):348.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 174(2):369–386.

Sun, B., Cui, Y., and Tchetgen, E. T. (2022). Selective machine learning of the average treatment effect with an invalid instrumental variable. *The Journal of Machine Learning Research*, 23(204):1–40.

Sun, B. and Miao, W. (2022). On semiparametric instrumental variable estimation of average treatment effects through data fusion. *Statistica Sinica*, 32:569–590.

Tchetgen Tchetgen, E. (2014). The control outcome calibration approach for causal inference with unobserved confounding. *American Journal of Epidemiology*, 179(5):633–640.

Tchetgen Tchetgen, E., Sun, B., and Walter, S. (2021). The genius approach to robust mendelian randomization inference. *Statistical Science*, 36(3):443–464.

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., and Miao, W. (2024). An introduction to proximal causal learning. *Statistical Science*, 39(3):375–390.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(3):531–550.

Ware, J. E., Keller, S. D., and Kosinski, M. (1995). *SF-12: How to score the SF-12 physical and mental health summary scales*. Health Institute, New England Medical Center.

Yang, S. and Ding, P. (2020). Combining Multiple Observational Data Sources to Estimate Causal Effects. *Journal of the American Statistical Association*, 115(531):1540–1554.

Yang, S., Gao, C., Zeng, D., and Wang, X. (2023). Elastic integrative analysis of randomised trial and real-world data for treatment heterogeneity estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):575–596.

Zhao, Q., Wang, J., Spiller, W., Bowden, J., and Small, D. S. (2019). Two-sample instrumental variable analyses using heterogeneous samples. *Statistical Science*, 34(2):317–333.
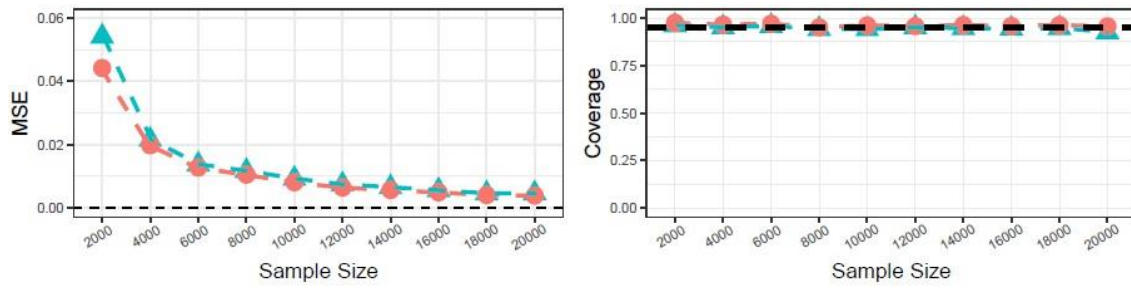
Figure 1: Mean squared error (MSE) and 95% coverage probability of the lasso-based MR estimator $\hat{\tau}^{\mathrm{e}}_{\mathrm{cf\text{-}la}}$ (red circle) and the super learner-based MR estimator $\hat{\tau}^{\mathrm{e}}_{\mathrm{cf\text{-}sl}}$ (green triangle) under $\kappa = 0$ and different sample sizes.

Table 1: Some key notation used throughout the paper.

Notation	Definition	Description
$\omega(V)$	$f(R=1\mid V)$	Sampling mechanism
$\gamma(V)$	$\{1-\omega(V)\}/\omega(V)$	Odds function
q	$f(R=0)$ or $1-E\{\omega(V)\}$	Target data proportion
$\mu_R(V)$	$E(X\mid V,R)$	Treatment propensity score
$\eta(X,V,R)$ or η	$X-\mu_R(V)$	Treatment regression residual
$\sigma_R^2(V)$	$\mathrm{Var}(X\mid V,R)$	Conditional variance of treatment
$\beta_R(V)$	$E(Y_1-Y_0\mid V,R)$	Conditional average treatment effect
$\pi(V)$	$f(Z=1\mid V,R=1)$	Conditional probability of $Z=1$ in auxiliary data
$\mu_0^X(V)$	$E(X\mid Z=0,V,R=1)$	Conditional treatment mean given $(Z=0,V)$ in auxiliary data
$\mu_0^Y(V)$	$E(Y\mid Z=0,V,R=1)$	Conditional outcome mean given $(Z=0,V)$ in auxiliary data
$\delta^X(V)$	$E(X\mid Z=1,V,R=1)-E(X\mid Z=0,V,R=1)$	Conditional treatment mean difference in auxiliary data
$\delta^Y(V)$	$E(Y\mid Z=1,V,R=1)-E(Y\mid Z=0,V,R=1)$	Conditional outcome mean difference in auxiliary data
$\phi_R(V)$	$E\{Y-\beta_R(V)X\mid V,R\}$	Conditional mean function of the residual $Y-\beta_R(V)X$
$\rho(V)$	$E[\eta\{Y-\beta_R(V)X\}\mid V,R=0]$	Conditional mean function of $\eta\{Y-\beta_R(V)X\}$ in target data

Table 2: Bias, standard deviation (SD), and 95% coverage probability (CP) of the estimators under $\kappa = 0$ and $n = 2000, 4000$. All values have been multiplied by 100.

Scenario	$\hat{\tau}_1^{e}$		$\hat{\tau}_2^{e}$		$\hat{\tau}_3^{e}$		$\hat{\tau}_4^{e}$		$\hat{\tau}_5^{e}$		$\hat{\tau}_6^{e}$		$\hat{\tau}_{mr}^{e}$		$\hat{\tau}_{mr}^{h}$	
	2000	4000	2000	4000	2000	4000	2000	4000	2000	4000	2000	4000	2000	4000	2000	4000
Bias																
$\mathcal{S}_0$	-2	-2	0	-1	-2	-2	0	-1	-1	-1	0	0	1	1	1	1
$\mathcal{S}_1$	-2	-2	72	69	25	24	97	94	3	3	30	28	1	0	0	0
$\mathcal{S}_2$	-176	-180	0	-1	-138	-142	26	25	30	28	3	3	0	-1	0	-1
$\mathcal{S}_3$	24	24	97	94	-2	-2	74	72	-13	-14	-35	-37	1	0	0	0
$\mathcal{S}_4$	-138	-143	26	25	-178	-182	0	-1	86	83	-13	-13	0	-1	0	-1
$\mathcal{S}_5$	3	3	21	20	12	12	54	52	-1	-1	75	73	0	0	14	14
$\mathcal{S}_6$	31	28	3	2	18	16	12	12	-143	-148	0	0	0	-1	0	0
$\mathcal{S}_7$	51	50	54	52	52	50	54	52	54	52	54	53	53	51	48	47
SD																
$\mathcal{S}_0$	45	31	19	13	45	31	19	13	36	24	19	13	24	15	25	16
$\mathcal{S}_1$	45	31	25	17	43	29	24	17	20	14	18	13	32	20	32	21
$\mathcal{S}_2$	50	36	19	13	46	33	19	13	23	16	14	10	19	14	20	14
$\mathcal{S}_3$	43	29	24	17	45	31	25	17	20	13	28	19	30	20	32	21
$\mathcal{S}_4$	46	33	19	13	50	36	19	13	38	26	18	12	19	13	20	14
$\mathcal{S}_5$	20	14	19	13	20	14	19	13	36	24	26	18	25	16	16	11
$\mathcal{S}_6$	24	17	14	10	20	14	14	10	43	31	19	13	19	13	19	13
$\mathcal{S}_7$	19	13	18	13	19	13	18	13	19	13	18	13	18	13	18	13
CP																
$\mathcal{S}_0$	97	97	97	96	97	97	97	96	98	98	97	96	98	96	97	96
$\mathcal{S}_1$	97	97	13	1	96	92	0	0	99	99	67	36	95	95	94	95

$\mathcal{S}_2$	2	0	97	96	10	0	73	51	81	62	97	97	97	96	96	96
$\mathcal{S}_3$	96	92	0	0	97	97	10	0	94	87	80	53	95	96	94	95
$\mathcal{S}_4$	9	0	72	52	2	0	97	96	35	5	93	85	97	96	96	96
$\mathcal{S}_5$	99	99	82	69	95	91	15	1	98	98	13	1	96	97	87	76
$\mathcal{S}_6$	81	66	98	97	90	85	89	79	5	0	97	96	97	96	96	96
$\mathcal{S}_7$	20	2	13	1	20	2	14	1	13	0	11	0	14	1	22	2

Table 3: Estimates of the average causal effect τ in the higher-income group.

Estimators	Point estimate	Standard error	95% Confidence interval
$\hat{\tau}_{\text{weak}}$	-9.54	15.69	(-40.30, 21.22)
$\hat{\tau}_{\text{mr}}^{\text{h}}$	-7.23	3.98	(-15.03, 0.56)
$\hat{\tau}_{\text{mr}}^{\text{e}}$	-8.68	4.49	(-17.48, 0.12)
$\hat{\tau}_{\text{cf-la}}^{\text{e}}$	-8.32	4.73	(-17.60, 0.96)
$\hat{\tau}_{\text{cf-sl}}^{\text{e}}$	-8.40	4.53	(-17.28, 0.48)