Submitted to the Annals of Applied Statistics

EFFICIENT SEMIPARAMETRIC ESTIMATION OF MARGINAL TREATMENT EFFECTS WITH GENETIC INSTRUMENTAL VARIABLES

BY ASHISH PATEL[1,a], FRANCIS J. DITRAGLIA[2,b], AND STEPHEN BURGESS[1,3,c]

[1]*MRC Biostatistics Unit, University of Cambridge,* [a]*ashish.patel@mrc-bsu.cam.ac.uk*

[2]*Department of Economics, University of Oxford,* [b]*francis.ditraglia@economics.ox.ac.uk*

[3]*Cardiovascular Epidemiology Unit, University of Cambridge,* [c]*sb452@cam.ac.uk*

Alcohol misuse is a key target of public health strategies aimed at reducing cardiovascular risk. The effect of excessive alcohol consumption on blood pressure may vary systematically with individuals' unobserved propensity to engage in heavy drinking, complicating causal inference with observational data. The marginal treatment effects framework uses an instrumental variable for treatment choice (excessive alcohol consumption) to study how selection into treatment is linked with the treatment effect. We explore the use of a genetic instrument within this framework, which is challenging because genetic compliers (individuals for whom a change in the instrument changes their treatment choice) are likely to be a small proportion of the overall sample. This can lead to greater sampling uncertainty in the tails of the propensity score distribution, i.e., the conditional probability of choosing treatment, and in turn poor estimation of causal estimands that measure heterogeneous treatment effects. We show that the use of efficient influence functions of target estimands improves estimation in terms of robustness to sampling uncertainty in nonparametrically estimated propensity scores. We find evidence of *reverse selection on gains*: individuals most prone to excessive alcohol consumption experience larger adverse effects on blood pressure.

1. Introduction. Excessive alcohol consumption is a modifiable behaviour that is responsible for an estimated 2.6 million deaths per year globally according to World Health Organization estimates (Anderson et al., 2023). Alcohol misuse features prominently in public health recommendations aimed at reducing cardiovascular risk, with clinical trial evidence establishing a causal link between alcohol consumption and higher blood pressure levels (Tasnim et al., 2020).

Individuals differ substantially in their alcohol drinking habits and responsiveness to health information (DiMatteo, 2004), suggesting the effects of excessive alcohol consumption could also be heterogeneous and closely tied to drinking behaviour. For example, those with high blood pressure may be advised to lower alcohol intake which may attenuate any positive association. Understanding this relationship between the effects of excessive alcohol consumption (the *treatment effect*) and the choice to drink excessively (the *treatment choice*) is crucial for the design of effective public health interventions (Heckman and Vytlacil, 2001).

In practice, we cannot randomly assign individuals to drink alcohol excessively, which leaves the concern that an observed association between treatment choice and the outcome (blood pressure) does not reflect only the causal effect of treatment but may instead be distorted by unmeasured factors that correlate the two, such as dietary, smoking, and exercise habits (Di Federico et al., 2023). The method of instrumental variables (IVs) is an important tool in applied research to infer causal effects by leveraging sources of variation in treatment choice that are unrelated to unobserved confounders.

Keywords and phrases: Marginal Treatment Effects, Semiparametric Inference, Genetic Instruments.

In genetic epidemiology, Mendelian randomization uses genetic variants as IVs, where a genetic variant can be regarded as a variable taking values 0, 1, or 2, corresponding to the number of treatment-increasing alleles inherited by an individual. This approach aims to exploit the quasi-random allocation of genetic variants at conception. In particular, genetic variants may be less prone to issues of reverse causation and confounding (Davey Smith and Ebrahim, 2003). The approach is most plausible when the biological mechanism by which the genetic instrument influences treatment choice is known (Larsson, Butterworth and Burgess, 2023). Our instrument is a genetic score for alcohol consumption: a linear combination of genetic variants, calculated as a weighted sum of consumption-increasing alleles. A variant from the *ADH1B* gene region receives the largest weight in our genetic score, and is known to affect the ability to metabolize alcohol and thus the risk of alcohol dependence (Rietschel and Treutlein, 2013).

Recent years have seen a sharp rise in the number of Mendelian randomization studies being performed (Hemani et al., 2025), with the vast majority of studies estimating a simple linear IV model, even when the treatment choice is binary (Wu and Wang, 2024). A large literature has focused on clarifying the interpretation of IV estimators in simple linear models that assume a constant treatment effect, when in fact there is unobserved heterogeneity linking treatment effects with treatment choice. Imbens and Angrist (1994) discussed conditions under which the IV estimand can be interpreted as an average treatment effect for *compliers*, a subgroup of the population for whom a change in the instrument changes their treatment choice.

There are two reasons why the IV estimand is unlikely to be useful for Mendelian randomization analyses under unobserved heterogeneity in treatment effects. First, genetics variants are likely to explain a low proportion of variability in treatment choice, and so genetic compliers may be only a small, and potentially unrepresentative, proportion of the population of interest (Burgess and Labrecque, 2018). Second, the IV estimand may not be interesting from a policy perspective because it is not possible to directly manipulate the instrument (genetic variants).

A closely related strategy that directly estimates the relationship between treatment effects and treatment choice is the marginal treatment effects (MTE, Heckman and Vytlacil 1999, 2005) framework. The MTE estimand measures how treatment effects vary with a continuous latent variable representing individual-specific reluctance to take treatment, and it can also be reweighted to derive conventional estimands such as the average treatment effect.

In this paper, we apply the MTE framework to study treatment effect heterogeneity with a genetic instrumental variable. MTE estimation relies on first stage estimation of a propensity score (defined as the conditional probability of treatment given observed covariates and instrument). Due to the potentially complex nature of how genetic variants influence excessive alcohol consumption (Edenberg and Foroud, 2013), we will not impose a parametric assumption on the propensity score. Instead, we will study the general case where the propensity score is estimated nonparametrically.

The conventional approach to estimation in the MTE framework uses a simple linear regression and does not account for uncertainty in first-stage propensity score estimation. When the instrument induces only limited variation in treatment, the estimated propensity score is concentrated over a narrow range, leading to poor estimation of the marginal treatment effects curve. These conventional estimates of target estimands will *not* be robust nor efficient.

Thus, for improved estimation in the MTE framework with genetic instruments, we derive semiparametrically efficient estimates of target estimands under plug-in nonparametric propensity scores. Importantly, these efficient estimates have low variance, and are robust to sampling uncertainty in the nonparametrically estimated propensity score. We illustrate the benefits of this robustness property in desensitizing estimation to both weak instruments and

the choice of tuning parameters for nonparametric propensity score estimation. In our linear model, calculating efficient estimates and their standard errors is computationally straightforward.

When using genetic instruments, the exclusion restriction assumption may fail due to horizontal pleiotropy, where variants affect the outcome through pathways other than treatment, which can bias causal effect estimates. We therefore build on the "plausibly exogenous" sensitivity analysis of Conley, Hansen and Rossi (2012) for our MTE framework, relaxing exact instrument validity to allow a direct instrument effect on potential outcomes. This allows an assessment of how robust estimates of target estimands are to plausible violations of the exclusion restriction assumption.

Proofs of technical results are provided in Supplementary Material. An R package to implement the methods discussed in this paper is provided at github.com/ash-res/efficient.mte/.

2. Model and Estimands.

2.1. *Background.* The combination of rich phenotypic information, genetic data, and linked electronic health records has made UK Biobank a cornerstone dataset for recent epidemiological research on causal mechanisms of diseases (Sudlow et al., 2015). UK Biobank recruited over 500,000 participants between the ages of 40 and 69 across the United Kingdom between 2006 and 2010, collecting detailed self-reported data on health and lifestyle behaviours, alongside physical measurements such as blood pressure, and genome-wide genotyping. A large literature using UK Biobank data has studied the health consequences of excessive alcohol consumption, including cardiovascular disease (Larsson et al., 2020), cancer risk (Larsson et al., 2025), brain health (Topiwala et al., 2022a), and mortality (Schaefer et al., 2023; Kassaw et al., 2024).

Some of these studies have established the strong adverse associations of even moderate alcohol consumption (Topiwala et al., 2022a; Schaefer et al., 2023), while others have used Mendelian randomization designs for causal estimation of population-average effects (Larsson et al., 2020, 2025). The Mendelian randomization study of Kassaw et al. (2024) moved beyond estimating a simple linear model by using a stratification technique to estimate average causal effects at different consumption levels.

However, policy conclusions based solely on average treatment effects may overlook economically and clinically important patterns of self-selection into treatment. In many health applications, individuals who select into a treatment are either those who benefit most (selection on gains) or those who benefit least (reverse selection on gains); this has important implications for policy design and targeting, as well as for the interpretation of conventional IV estimates (Heckman and Vytlacil, 2001). Previous applications have shown that considering selection on gains can alter substantive conclusions in many different health contexts, including breast cancer treatment decisions (Basu et al., 2007), living arrangements and elderly health (Johar and Maruyama, 2014), medication adherence (Depalo, 2020), health insurance coverage (Kowalski, 2023), and social wellbeing (Wilding, Munford and Sutton, 2023).

Despite extensive research in genetic epidemiology examining the average effects of alcohol consumption, including studies using Mendelian randomization designs, little is known about whether individuals who are predisposed to engage in excessive drinking are systematically more or less susceptible to its adverse health effects. The MTE framework allows us to investigate the extent to which selection into excessive alcohol consumption is related to underlying heterogeneity in blood pressure.

2.2. *Model.* To estimate the causal effects of a binary treatment, the majority of Mendelian randomization applications continue to estimate a simple linear instrumental variable (IV) model (Wu and Wang, 2024). However, when unobserved heterogeneity in treatment choice is related to treatment effects, the standard IV estimand can be difficult to interpret (Imbens and Angrist, 1994). In particular, the IV estimand reflects an average treatment effect for compliers, a small and potentially unrepresentative subgroup of the overall population (Burgess and Labrecque, 2018). Moreover, the IV estimand may be of limited policy relevance since the genetic instrument cannot be altered.

The marginal treatment effects framework offers a way to move beyond this complier average effect, and to directly estimate how treatment effects vary with individuals' propensity to choose treatment, as well as to recover conventional estimands (such as the average treatment effect) as simple weighted averages.

We start by introducing the generalized Roy (1951) model which captures unobserved heterogeneity linking treatment effects with the choice to take treatment. Let A denote a binary indicator for treatment, which equals 1 if treated, and 0 if untreated. Let Z denote an instrumental variable for treatment, and $\boldsymbol{X}$ denote a vector of observed covariates. To relate unobserved heterogeneity in treatment effects with treatment choice, it is useful to start by modelling the selection process. We assume treatment choice is described by the model

$$A = I\{\pi_{\mathcal{P}}(\boldsymbol{X}, Z) > V\} \tag{1}$$

where $\pi_{\mathcal{P}}(\boldsymbol{X}, Z)$ is an *unknown* function of $\boldsymbol{X}$ and Z, but is monotone in Z conditional on $\boldsymbol{X}$. The subscript $\mathcal{P}$ denotes that it is the true or population average, rather than an estimate or sample average. The *unobserved* variable V represents unobserved reluctance to choosing treatment that is unrelated to the genetic instrument given covariates. For any fixed level of covariates $\boldsymbol{X}$ and instrument Z, those individuals with high levels of V are less likely to choose treatment than those with low levels of V. Thus, V may be the composite of various factors such as individual preferences, motivation, environment, and physiology. Since the function $\pi_{\mathcal{P}}(\boldsymbol{X}, Z)$ is left unspecified, Equation (1) is a quite general specification.

For each individual, let Y_0 denote their untreated potential outcome, and Y_1 their treated potential outcome. The generalized Roy model allows the individual-specific treatment effect $Y_1 - Y_0$ to vary with V, which captures the dynamic that unobserved heterogeneity in treatment effects and treatment choice are correlated. To estimate treatment effects in this setting, we will make use of a valid instrument Z such that

$$Z \text{ is conditionally independent of } (Y_0, Y_1, V) \text{ given } \boldsymbol{X}. \tag{2}$$

In the language of IVs, Equation (2) is an *exclusion restriction* assumption: for any fixed level of covariates, the instrument is independent of the potential outcomes, and so the only way it may influence the outcome is through its effect on treatment choice. The instrument can influence the propensity to take treatment by shifting $\pi_{\mathcal{P}}(\boldsymbol{X}, Z)$ without changing unobserved heterogeneity V, and thus can be used to capture variation in treatment choice that is unrelated to unobserved confounding.

To analyze treatment effect heterogeneity, it is easier to think about how the treatment effect varies with quantiles of unobserved heterogeneity V, rather than its cardinal scale. To this end, it is standard to use the probability integral transform to normalize V as a standard uniform random variable conditional on $\boldsymbol{X}$, so that $\pi_{\mathcal{P}}(\boldsymbol{X}, Z) = \mathcal{P}(A = 1|\boldsymbol{X}, Z)$ in Equation (1) becomes the propensity score of treatment, which we will estimate nonparametrically. We emphasize that this transformation is not a modelling assumption, but a normalization given Equations (1) and (2) if the conditional distribution of V given covariates $\boldsymbol{X}$ is continuous. Given this, V is a rank that orders individuals by their unobserved propensity to select

treatment; only this ordering matters for the treatment selection model, so no substantive characterization of V is required beyond "unobserved reluctance to take treatment".

Next, we model the dependence of the potential outcome under each treatment choice on V. We assume a simple linear model for conditional potential outcome means such that

$$E[Y_a|\boldsymbol{X},V] = \underbrace{\alpha_a - 0.5\zeta_a}_{\text{constant term}} + \boldsymbol{\beta_a}'\boldsymbol{X} + \zeta_a V \tag{3}$$

for $a = 0, 1$, where α_a, $\boldsymbol{\beta_a}$, and ζ_a are unknown parameters. The specific form of the constant term is without loss of generality: since V is uniform on $[0,1]$ given $\boldsymbol{X}$, the $-0.5\zeta_a$ centres the heterogeneity so that $E[U_a|\boldsymbol{X}] = 0$ for $U_a = Y_a - \alpha_a - \boldsymbol{\beta_a}'\boldsymbol{X}$, which simplifies derivations (Brinch, Mogstad and Wiswall, 2017).

The so-called marginal treatment effect (MTE; Heckman and Vytlacil, 1999) is a function of observed covariates $\boldsymbol{X}$ and individual-specific reluctance to take treatment V,

$$\begin{aligned}\theta_{\text{MTE}}(\boldsymbol{X},V) &= E[Y_1 - Y_0|\boldsymbol{X},V] \\ &= \underbrace{\alpha_1 - \alpha_0 - 0.5(\zeta_1 - \zeta_0)}_{\text{constant term}} + (\boldsymbol{\beta_1} - \boldsymbol{\beta_0})'\boldsymbol{X} + (\zeta_1 - \zeta_0)V.\end{aligned}$$

Within this MTE function, determining the sign of $\zeta_1 - \zeta_0$, if any, is typically a main focus of MTE estimation since it reveals the direction of correlation between treatment effects and unobserved reluctance to take treatment.

To understand how this model of potential outcomes relates to the standard linear IV model used in the majority of Mendelian randomization applications, note that since the observed outcome is $Y = (1-A)Y_0 + AY_1$, we can write $E[Y|\boldsymbol{X},V] = \delta(\boldsymbol{X}) + \theta_{\text{MTE}}(\boldsymbol{X},V)\cdot A + \zeta_0 V$ where the term $\zeta_0 V$ is independent with the instrument Z by Equation (2), and $\delta(\boldsymbol{X})$ is a function of covariates. In the standard linear IV model under constant treatment effects, the IV estimator would be consistent for $\alpha_1 - \alpha_0$, the average treatment effect over everyone in the population. However, more generally this IV estimate will not be useful for inferring how treatment effects vary with unobserved heterogeneity in treatment choice.

Covariates play three roles in our MTE model, which together guide their selection in practice. First, since only conditional independence of the instrument is required in Equation (2), covariates can be chosen to make this assumption more credible. Unfortunately, few covariates are likely to be available for this purpose when using a genetic instrument since most measured characteristics are post-randomization variables, and conditioning on a variable lying on the pathway from the instrument to the outcome may induce bias. Moreover, parsimonious covariate selection is also preferable because including many covariates can result in slow convergence rates of first stage nonparametric propensity score estimates. Second, covariates can provide identifying information on treatment effect heterogeneity. In our model, the MTE is identified using the joint variation in covariates and the propensity score (Brinch, Mogstad and Wiswall, 2017), and therefore covariates that shift the propensity score can be particularly valuable when the genetic instrument induces only limited variation in treatment choice. Third, and most simply, covariates may describe treatment effect heterogeneity that is itself of scientific or policy interest.

2.3. *Marginal Treatment Effects and Target Estimands.* One issue with directly estimating the MTE function $\theta_{\text{MTE}}(\boldsymbol{X},V)$ is that it depends on the unobserved variable V. However, an observable quantity is the conditional mean of the observed outcome $m(\boldsymbol{x},p) = E_{\mathcal{P}}[Y|\boldsymbol{X} = \boldsymbol{x}, \pi_{\mathcal{P}}(\boldsymbol{X},Z) = p]$, which under Equations (1)–(3) can be written as

$$m(\boldsymbol{x},p) = \boldsymbol{r}(\boldsymbol{x},p)'\boldsymbol{\gamma} \tag{4}$$

where $\boldsymbol{r}(\boldsymbol{x}, p)$ is a known vector, and $\boldsymbol{\gamma}$ is an unknown vector such that

$$\boldsymbol{r}(\boldsymbol{x}, p) = \begin{bmatrix} 1 \\ \boldsymbol{x} \\ p \\ \boldsymbol{x}p \\ p^2 \end{bmatrix}, \text{ and } \boldsymbol{\gamma} = \begin{bmatrix} \alpha_0 \\ \boldsymbol{\beta_0} \\ \alpha_1 - \alpha_0 - 0.5(\zeta_1 - \zeta_0) \\ \boldsymbol{\beta_1} - \boldsymbol{\beta_0} \\ 0.5(\zeta_1 - \zeta_0) \end{bmatrix}.$$

Heckman and Vytlacil (1999) showed that the MTE function is identified as the derivative

$$\begin{aligned} \theta_{\text{MTE}}(\boldsymbol{x}, v) &= \partial m(\boldsymbol{x}, p)/\partial p|_{p=v} \\ &= \underbrace{\alpha_1 - \alpha_0 - 0.5(\zeta_1 - \zeta_0)}_{\text{constant term}} + (\boldsymbol{\beta_1} - \boldsymbol{\beta_0})'\boldsymbol{x} + (\zeta_1 - \zeta_0)v \end{aligned}$$

and it represents the treatment effect for an individual who is indifferent between being treated and untreated, at their given value of instrument and covariates. This specification of the MTE means that the treatment effect can vary linearly with covariates $\boldsymbol{X}$ and unobserved heterogeneity V, and the intercept of the linear-in-v MTE slope can vary with covariates.

By taking weighted averages of the vector of parameters $\boldsymbol{\gamma}$, we can also calculate several scalar treatment effect estimands that may be of interest; a detailed list of possible target estimands is discussed in Heckman and Vytlacil (2005) and Mogstad and Torgovitsky (2018). In our analysis, we focus on the following four target estimands. First, the *average treatment effect* (ATE) measures the average effect of treatment across the whole population,

$$\theta_{\text{ATE}}(\boldsymbol{\gamma}, \mathcal{P}) = E_{\mathcal{P}}[Y_1 - Y_0] = E_{\mathcal{P}}[\boldsymbol{r}_{\text{ATE}}(\boldsymbol{X})]'\boldsymbol{\gamma} \tag{5}$$

where $\boldsymbol{r}_{\text{ATE}}(\boldsymbol{X}) = (0, \mathbf{0}', 1, \boldsymbol{X}', 1)'$.

The *average treatment effect on the treated* (ATT) measures the average treatment effect for those who choose treatment,

$$\theta_{\text{ATT}}(\boldsymbol{\gamma}, \mathcal{P}) = E_{\mathcal{P}}[Y_1 - Y_0|A = 1] = \mathcal{P}(A = 1)^{-1} E_{\mathcal{P}}[\boldsymbol{r}_{\text{ATT}}(\boldsymbol{X}, Z, \mathcal{P})]'\boldsymbol{\gamma} \tag{6}$$

where $\boldsymbol{r}_{\text{ATT}}(\boldsymbol{X}, Z, \mathcal{P}) = (0, \mathbf{0}', \pi_{\mathcal{P}}(\boldsymbol{X}, Z), \pi_{\mathcal{P}}(\boldsymbol{X}, Z)\boldsymbol{X}', \pi_{\mathcal{P}}(\boldsymbol{X}, Z)^2)'$. Analogously, the *average treatment effect on the untreated* (ATU) measures the average effect of treatment for those who choose to be untreated,

$$\theta_{ATU}(\boldsymbol{\gamma}, \mathcal{P}) = E_{\mathcal{P}}[Y_1 - Y_0|A = 0] = \mathcal{P}(A = 0)^{-1} E_{\mathcal{P}}[\boldsymbol{r}_{\text{ATU}}(\boldsymbol{X}, Z, \mathcal{P})]'\boldsymbol{\gamma} \tag{7}$$

where $\boldsymbol{r}_{ATU}(\boldsymbol{X}, Z, \mathcal{P}) = (0, \mathbf{0}', 1 - \pi_{\mathcal{P}}(\boldsymbol{X}, Z), (1 - \pi_{\mathcal{P}}(\boldsymbol{X}, Z))\boldsymbol{X}', 1 - \pi_{\mathcal{P}}(\boldsymbol{X}, Z)^2)'$.

The *average selection on gains* (ASG) estimand simply finds the difference between the ATT and the ATU, and summarizes how selection into treatment is influenced by the treatment effect on average,

$$\begin{aligned} \theta_{\text{ASG}}(\boldsymbol{\gamma}, \mathcal{P}) &= E_{\mathcal{P}}[Y_1 - Y_0|A = 1] - E_{\mathcal{P}}[Y_1 - Y_0|A = 0] \\ &= (E_{\mathcal{P}}[\boldsymbol{r}_{\text{ATT}}(\boldsymbol{X}, Z, \mathcal{P})]\mathcal{P}(A = 1)^{-1} - E_{\mathcal{P}}[\boldsymbol{r}_{\text{ATU}}(\boldsymbol{X}, Z, \mathcal{P})]\mathcal{P}(A = 0)^{-1})'\boldsymbol{\gamma}. \end{aligned} \tag{8}$$

In our application, if θ_{ASG} is negative, then this would be consistent with health conscious behaviour where individuals may act on private knowledge of their pre-existing health conditions, and so those choosing not to consume alcohol excessively would be more likely to be adversely affected in terms of higher blood pressure if they did.

It is easy to extend the potential outcomes model in Equation (3) to allow for a more complex specification of the MTE function $\theta_{\text{MTE}}(\boldsymbol{X}, V)$. Although this would see the dimension of $\boldsymbol{\gamma}$ increase, the target estimands in Equations (5)–(8) would still be weighted averages of $\boldsymbol{\gamma}$. We derive our results under a more general model that allows for higher-order polynomial terms and interactions between observed covariates $\boldsymbol{X}$ and unobserved heterogeneity V (see Supplementary Material for further details).

3. Efficient Estimation and Inference. This section outlines two approaches to estimate the MTE parameters $\boldsymbol{\gamma}$ and the target estimands in Section 2.2. First, the "conventional" approach, and the one that is most widely used in current practice, estimates the MTE parameters $\boldsymbol{\gamma}$ through a simple linear regression, and then calculates the sample counterparts of Equations (5)–(8) to estimate the target estimands.

Second, we propose an alternative "efficient" approach by calculating efficient influence functions of $\boldsymbol{\gamma}$ and target estimands to derive semiparametrically efficient estimates. Compared with the conventional approach, our proposed efficient estimates of target estimands will have a lower asymptotic variance and be robust to sampling uncertainty in first stage propensity score estimation. This robustness is particularly important when genetic instruments only weakly influence treatment choice.

3.1. *Estimation of Individual MTE Parameters.* The previous section shows that the estimands we are interested in are weighted averages of the parameters in $\boldsymbol{\gamma}$. Given the observed mean in Equation (4), $\boldsymbol{\gamma}$ is equal to the coefficient of the *population-level* linear regression of Y on $\boldsymbol{r}(\boldsymbol{X}, \pi_{\mathcal{P}}(\boldsymbol{X}, Z))$. Therefore, a feasible and intuitive method to estimate $\boldsymbol{\gamma}$ simply runs a linear regression of Y on $\boldsymbol{r}(\boldsymbol{X}, \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z))$ where $\pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z)$ is an estimate of the propensity score. We denote this estimate of $\boldsymbol{\gamma}$ as $\widehat{\boldsymbol{\gamma}}$, and refer to this as the "conventional" estimator because it is commonly used in applied work. For inference, usual standard errors of $\widehat{\boldsymbol{\gamma}}$ from a typical linear regression output will not be correct because they should be corrected to account for additional variability from first stage propensity score estimation.

Is this conventional estimator $\widehat{\boldsymbol{\gamma}}$ the best estimator of $\boldsymbol{\gamma}$? In terms of achieving a low variance, it is efficient; its asymptotic variance is equal to the semiparametric variance lower bound for the regression estimand $\boldsymbol{\gamma}$. However, $\widehat{\boldsymbol{\gamma}}$ can be sensitive to misspecification of the propensity score $\pi_{\mathcal{P}}(\boldsymbol{X}, Z)$, or to the choice of tuning parameters if it is nonparametrically estimated as in our case. More broadly, $\widehat{\boldsymbol{\gamma}}$ can be sensitive to sampling uncertainty in first stage propensity score estimates, which we demonstrate is more likely under weak instruments. Given that genetic variants will typically explain only a low proportion of variability in treatment choice, it is important in practice that our estimates of MTE parameters are robust to weak instruments.

One way to achieve this is by constructing an estimator $\widetilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ based on the efficient influence function of the regression estimand $\boldsymbol{\gamma}$, which we call the "efficient" estimator. This estimator is semiparametrically efficient for the regression estimand $\boldsymbol{\gamma}$. In our linear model, $\widetilde{\boldsymbol{\gamma}}$ has a convenient closed form expression

$$\widetilde{\boldsymbol{\gamma}} = (\boldsymbol{\Omega}_{\widehat{\mathcal{P}}} + \boldsymbol{\Gamma}_{\widehat{\mathcal{P}}})^{-1} \boldsymbol{\Upsilon}_{\widehat{\mathcal{P}}}$$

where $\boldsymbol{\Omega}_{\widehat{\mathcal{P}}}$, $\boldsymbol{\Gamma}_{\widehat{\mathcal{P}}}$, and $\boldsymbol{\Upsilon}_{\widehat{\mathcal{P}}}$ are the sample counterparts, with plug-in propensity score estimates, of the population means

$$\begin{aligned}
\boldsymbol{\Omega}_{\mathcal{P}} &= E_{\mathcal{P}}[\boldsymbol{r}(\boldsymbol{X}, \pi_{\mathcal{P}}(\boldsymbol{X}, Z))\boldsymbol{r}(\boldsymbol{X}, \pi_{\mathcal{P}}(\boldsymbol{X}, Z))'] \\
\boldsymbol{\Gamma}_{\mathcal{P}} &= E_{\mathcal{P}}\big[\boldsymbol{R}(\boldsymbol{X}, \pi_{\mathcal{P}}(\boldsymbol{X}, Z))(A - \pi_{\mathcal{P}}(\boldsymbol{X}, Z))\boldsymbol{r}(\boldsymbol{X}, \pi_{\mathcal{P}}(\boldsymbol{X}, Z))'\big] \\
\boldsymbol{\Upsilon}_{\mathcal{P}} &= E_{\mathcal{P}}[\boldsymbol{r}(\boldsymbol{X}, \pi_{\mathcal{P}}(\boldsymbol{X}, Z))Y]
\end{aligned}$$

where $\boldsymbol{R}(\boldsymbol{X}, \pi_{\mathcal{P}}(\boldsymbol{X}, Z)) = (1, \boldsymbol{X}', 2\pi_{\mathcal{P}}(\boldsymbol{X}, Z))'$.

THEOREM 3.1 (Efficient MTE parameter estimates). *$\widetilde{\boldsymbol{\gamma}}$ is based on the efficient influence function of the regression estimand $\boldsymbol{\gamma}$, and both $\widehat{\boldsymbol{\gamma}}$ and $\widetilde{\boldsymbol{\gamma}}$ have the same asymptotic variance.*

While both $\widehat{\boldsymbol{\gamma}}$ and $\widetilde{\boldsymbol{\gamma}}$ have the same asymptotic variance, $\widetilde{\boldsymbol{\gamma}}$ has the advantage of being robust to uncertainty in first stage propensity score estimation. One property that feeds into

TABLE 1

The conventional estimator $\widehat{\boldsymbol{\omega}}'(.)\widehat{\gamma}$ and the efficient estimator $\widetilde{\boldsymbol{\omega}}'(.)\widehat{\gamma}$ for each treatment effect estimand, where $\widehat{p}_0 = \widehat{\mathcal{P}}(A=0)^{-1}$ and $\widehat{p}_1 = \widehat{\mathcal{P}}(A=1)^{-1}$.

Estimand	Conventional (plug-in) weights: $\widehat{\boldsymbol{\omega}}(.)$	Efficient weights: $\widetilde{\boldsymbol{\omega}}(.)$
ATE	$E_{\widehat{\mathcal{P}}}[\boldsymbol{r}_{\text{ATE}}(\boldsymbol{X})]$	$\widehat{\boldsymbol{\omega}}_{\text{ATE}} - \widehat{\boldsymbol{\omega}}_{\text{ATE}}\boldsymbol{\Omega}_{\widehat{\mathcal{P}}}^{-1}\boldsymbol{\Gamma}_{\widehat{\mathcal{P}}}$
ATT	$\widehat{p}_1 E_{\widehat{\mathcal{P}}}[\boldsymbol{r}_{\text{ATT}}(\boldsymbol{X}, \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z))]$	$\widehat{\boldsymbol{\omega}}_{\text{ATT}} - \widehat{p}_1 E_{\widehat{\mathcal{P}}}[\boldsymbol{r}_{\text{ATT}}(\boldsymbol{X}, \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z))]'\boldsymbol{\Omega}_{\widehat{\mathcal{P}}}^{-1}\boldsymbol{\Gamma}_{\widehat{\mathcal{P}}} +$ $\widehat{p}_1 E_{\widehat{\mathcal{P}}}\left[\frac{\partial}{\partial \pi}\boldsymbol{r}_{\text{ATT}}(\boldsymbol{X}, \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z)) \cdot (A - \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z))\right]$
ATU	$\widehat{p}_0 E_{\widehat{\mathcal{P}}}[\boldsymbol{r}_{\text{ATU}}(\boldsymbol{X}, \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z))]$	$\widehat{\boldsymbol{\omega}}_{\text{ATU}} - \widehat{p}_0 E_{\widehat{\mathcal{P}}}[\boldsymbol{r}_{\text{ATU}}(\boldsymbol{X}, \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z))]'\boldsymbol{\Omega}_{\widehat{\mathcal{P}}}^{-1}\boldsymbol{\Gamma}_{\widehat{\mathcal{P}}} +$ $\widehat{p}_0 E_{\widehat{\mathcal{P}}}\left[\frac{\partial}{\partial \pi}\boldsymbol{r}_{\text{ATU}}(\boldsymbol{X}, \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z)) \cdot (A - \pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z))\right]$
ASG	$\widehat{\boldsymbol{\omega}}_{\text{ATT}} - \widehat{\boldsymbol{\omega}}_{\text{ATU}}$	$\widetilde{\boldsymbol{\omega}}_{\text{ATT}} - \widetilde{\boldsymbol{\omega}}_{\text{ATU}}$

the robustness of the efficient estimator $\widetilde{\gamma}$ is that it is asymptotically equivalent to an estimator that would be obtained if the true propensity score function $\pi_{\mathcal{P}}(\boldsymbol{X}, Z)$ was known (Hines et al., 2022). In contrast the conventional estimator $\widehat{\gamma}$ would not be efficient if the propensity score was known, even though it is when the propensity score is estimated.

For inference, it is also easy to calculate the standard error of $\widetilde{\gamma}$ because a consistent estimator of the asymptotic variance of $\widetilde{\gamma}$ is the sample counterpart of the variance of the efficient influence function divided by sample size. In particular, unlike for $\widehat{\gamma}$, there is no need to calculate the additional uncertainty related to first stage propensity score estimation because this term is negligible by design (see, for example, Hines et al., 2022).

3.2. *Estimation of Target Estimands.* For the four target estimands in Equations (5)–(8), we analogously describe the "conventional" estimates as those that simply replace the unknown γ with the conventional ordinary least squares estimator $\widehat{\gamma}$, the unknown propensity scores $\pi_{\mathcal{P}}(\boldsymbol{X}, Z)$ with parametric or nonparametric estimates $\pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z)$, and population observed means with their sample counterparts. These conventional estimates of the target estimands will *not be robust nor efficient.*

In Table 1, we derive efficient estimates of target estimands. Both the conventional and efficient estimates of the target estimands are weighted averages of the ordinary least squares estimator $\widehat{\gamma}$; Table 1 shows the weights for the efficient estimates involve only a couple more terms compared to the conventional estimates, and therefore the efficient estimates are also easy to compute.

THEOREM 3.2 (Efficient target parameter estimates). *For target estimands in Equations (5)–(8), which are defined as functions of the regression estimand γ, the semiparametrically efficient estimates are given in Table 1.*

Using efficient estimators of target estimands ensures low variance in estimation, and they also have the robustness property discussed above where the first order asymptotic distribution of the efficient estimators are equivalent to the oracle case where the true propensity score is known. The benefit of robustness to first stage nuisance estimation has been extensively explored, especially in the literature on "doubly robust" estimation of treatment effects (Scharfstein, Rotnitzky and Robins, 1999), where asymptotic normality holds even under

slow convergence rates from machine learning estimators of first stage nuisance functions (Chernozhukov et al., 2018).

For our applied focus, this robustness property is useful when estimating MTEs using a genetic score as an instrument. The propensity score can be precisely measured where the distribution of genetic scores are heavily concentrated, but estimating the tails of the propensity scores are based on many fewer observations (see Figure 4 below). As a result, sampling uncertainty in propensity score estimation will be much greater in the tails, providing a challenge to reliably estimate the MTE curve over the full range of observed propensity scores. Importantly, efficient estimators are less sensitive to this type of weak instrument setting.

The asymptotic variances of our *efficient* treatment effect estimators are unaffected by whether propensity score estimation is nonparametric or parametric, provided the propensity score estimates converge at the usual nonparametric rate of $o_p(n^{-1/4})$ required in locally robust generalized method of moments estimation (Hines et al., 2022; Chernozhukov et al., 2022). We suggest nonparametric estimation of the propensity score since misspecification may be particularly harmful in the MTE framework. By Equation (4) above, the propensity score enters nonlinearly in multiple regressors in the outcome regression model used to estimate MTE parameters. Compared with some treatment effects settings, there is no double robustness property that can rescue a misspecified propensity score model. Moreover, we expect in typical applications using genetic instrumental variables that there may be no strong biological or statistical priors to justify a specific parametric model of treatment choice. For example, in our empirical investigation, the genetics of excessive alcohol consumption is complex and polygenic (Edenberg and Foroud, 2013).

However, a parametric first stage may offer greater stability in applications with smaller sample sizes, or where there is very limited variation in treatment choice caused by the genetic instrument. Our accompanying software allows users to supply either parametric or nonparametric propensity score estimates.

3.3. *Sensitivity Analysis with an Invalid Instrument.* In Mendelian randomization, a central concern is that the exclusion restriction may be violated due to *horizontal pleiotropy*, where genetic variants affect the outcome through pathways other than the treatment (Verbanck et al., 2018). Consequently, a substantial literature has focused on developing methods that remain valid under restricted forms of pleiotropy. Examples include estimators that assume a majority of variants are valid instruments (Kang et al., 2016), random-effects approaches that model pleiotropic effects as having mean zero (Burgess et al., 2020), and Bayesian methods that accommodate correlated pleiotropy between instrument strength and direct effects (Morrison et al., 2020); see Kang et al. (2025) for a recent review.

In this section, we adopt a different perspective based on the "plausibly exogenous" framework of Conley, Hansen and Rossi (2012). Rather than imposing identifying assumptions for consistent estimation in the presence of invalid instruments, this approach allows uncertainty over the exclusion restriction assumption, and conducts a sensitivity analysis over a range of plausible direct instrument effects on the outcome. Researchers specify a set of values for an exclusion restriction violation parameter and then consider how the resulting causal estimates vary across that range. Thus, rather than trying to eliminate concerns about instrument invalidity, the objective is to assess whether substantive conclusions are robust to plausible departures from exact instrument validity. We emphasize that while this sensitivity analysis relaxes the exclusion restriction assumption, the monotonicity assumption, that selection into excessive alcohol consumption is monotone in the instrument given covariates, is maintained throughout and remains necessary for a marginal treatment effects interpretation.

Following Conley, Hansen and Rossi (2012), we allow the instrument to have a direct effect on the potential outcomes. Our selection model in Equation (1) is unchanged, and we

retain the exogeneity of the instrument for unobserved selection, that V is conditionally independent of Z given $\boldsymbol{X}$ as in Equation (2). However, we now relax the exclusion restriction assumption by generalizing the potential outcome model in Equation (3) to

$$Y_a = (\alpha_a - 0.5\zeta_a) + \boldsymbol{\beta_a}'\boldsymbol{X} + \tau_i Z + \zeta_a V + \epsilon_a, \qquad a = 0, 1,$$

where $E[\epsilon_a|\boldsymbol{X}, V, Z] = 0$, and τ_i is the direct effect of the instrument on individual i's potential outcomes, with average $E[\tau_i|\boldsymbol{X}, Z] = \tau$. Equation (3) is the special case of no direct effect, $\tau_i = 0$, under which the exclusion restriction in Equation (2) holds. The direct instrument effect is the same for both potential outcomes, meaning the causal effects of treatment are unchanged even when $\tau_i \neq 0$. In particular, although the instrument may affect outcome *levels* directly, it does not affect marginal treatment *gains*. Nevertheless, when $\tau \neq 0$ the instrument has a non-zero average direct effect τ on the potential outcome means, and this leads to biased estimation in general.

For our empirical investigation, our genetic score for alcohol consumption is now permitted to have a direct effect on blood pressure that is the same for a given individual whether or not they drink alcohol excessively, but which may differ across individuals. This is a meaningful relaxation of the exclusion restriction assumption: variants in the genetic score instrument can affect blood pressure through pathways other than excessive drinking, which may involve factors such as sleep, BMI, smoking, diet, or socioeconomics.

If we continue to estimate treatment effects using Heckman and Vytlacil (1999)'s approach, it can be shown that the observed mean $m(\boldsymbol{x}, p) = E[Y|\boldsymbol{X} = \boldsymbol{x}, \pi(\boldsymbol{X}, Z) = p]$ now has an additional term compared with Equation (4),

$$m(\boldsymbol{x}, p) = \boldsymbol{r}(\boldsymbol{x}, p)'\boldsymbol{\gamma} + \tau z(\boldsymbol{x}, p)$$

where $z(\boldsymbol{x}, p)$ solves $\pi(\boldsymbol{x}, z) = p$; see Supplementary Material for further details. Therefore, the usual strategy of estimating $\boldsymbol{\gamma}$ by a linear regression of Y on $\boldsymbol{r}(\boldsymbol{X}, \pi(\boldsymbol{X}, Z))$ may lead to biased estimation due to omitted variable bias.

Differentiating $m(\boldsymbol{x}, p)$ with respect to p no longer identifies the MTE function, but rather the MTE function plus a bias term,

$$\frac{\partial}{\partial p} m(\boldsymbol{x}, p) = \text{MTE}(\boldsymbol{x}, p) + \underbrace{\left(\partial \pi(\boldsymbol{x}, z)/\partial z\Big|_{z=z(\boldsymbol{x},p)}\right)^{-1} \tau}_{\text{bias}}.$$

Intuitively, this bias is larger when the average direct instrument effect τ on the potential outcomes is larger, and is amplified when $\partial \pi(\boldsymbol{x}, z)/\partial z$ is small, i.e., when the instrument is weak. In contrast to the homogeneous treatment effects setting of Conley, Hansen and Rossi (2012), here the bias varies with observed heterogeneity in the slope of the propensity score function.

Under knowledge of the true average direct effect τ, this bias is easily removed à la Conley, Hansen and Rossi (2012): we simply subtract the direct instrument variation τZ from the observed outcome Y. Since $m(\boldsymbol{x}, p) = \boldsymbol{r}(\boldsymbol{x}, p)'\boldsymbol{\gamma} + \tau z(\boldsymbol{x}, p)$, running a linear regression of $Y - \tau Z$ on $\boldsymbol{r}(\boldsymbol{X}, \pi(\boldsymbol{X}, Z))$ gives unbiased estimates of $\boldsymbol{\gamma}$ in Equation (4). We can then take the same weighted averages of $\boldsymbol{\gamma}$ discussed in Section 3.2 for unbiased estimation and inference on target estimands.

In practice, we do not know the true τ, and it is left as an unknown sensitivity parameter. We can consider a bound on τ such that it is assumed that it lies in a closed interval $[\underline{\tau}, \bar{\tau}]$. Then, for each $\tau^\star$ in $[\underline{\tau}, \bar{\tau}]$, we calculate a confidence interval for a target estimand assuming that $\tau = \tau^\star$, i.e., using $Y - \tau^\star Z$ as the outcome variable instead of Y leads to a collection of confidence intervals each indexed by $\tau^\star$. Taking the union of all such confidence intervals over $\tau^\star \in [\underline{\tau}, \bar{\tau}]$ provides a widened confidence interval for our target estimand that aims

to account for plausible violations of the exclusion restriction assumption. This confidence interval, asymptotically, will achieve nominal coverage under the assumption that the true τ lies in the interval $[\underline{\tau},\bar{\tau}]$.

The choice of a plausible bound $[\underline{\tau},\bar{\tau}]$ for the direct instrument effect necessarily relies on substantive judgment. One strategy is to benchmark τ against the effects of known risk factors that could generate pleiotropic pathways from the instrument to the outcome. For example, if variants within the genetic score are associated with obesity risk, and obesity is associated with at most roughly 5 mmHg higher systolic blood pressure, then this may provide a reasonable upper bound on the magnitude of any direct instrument effect. An alternative, more data-driven approach is to examine the observed associations between individual variants in the genetic score and the outcome. The largest such association can be used to inform a plausible upper bound for τ, under the assumption that any direct instrument effect operates in the same direction as the observed variant–outcome association. Together, these approaches provide a basis for conducting sensitivity analyses over a plausible range of values for τ.

4. Simulation Study. We performed a simulation study to illustrate how efficient methods can provide a more robust analysis under varying levels of instrument strength. We start by describing the simulation design, and then present results showing how the estimation performance of MTE parameters and target estimands varies with instrument strength under a nonparametrically estimated propensity score.

4.1. *Design.* We generated a binary covariate $X \sim \text{Binomial}(1, 0.5)$, and an instrument $Z \sim N(0, 1)$ as exogenous variables. Let (W_0, W_1) be bivariate normal random variables with variance 0.2 and correlation 0.2, and let V be a standard uniform random variable. We generated the potential outcomes as

$$Y_0 = 0.3 + 0.1X - 0.3(V - 1/2) + \sqrt{0.2}W_0$$
$$Y_1 = 0.5 + 0.2X + 0.3(V - 1/2) + \sqrt{0.2}W_1$$

so that Equation (3) was satisfied with $\alpha_0 = 0.3$, $\alpha_1 = 0.5$, $\beta_0 = 0.1$, $\beta_1 = 0.2$, $\zeta_0 = -0.3$, and $\zeta_1 = 0.3$.

The treatment indicator was generated as $A = 1$ only if $\Phi(\boldsymbol{q}(X, Z)'\boldsymbol{\eta})) > V$, where Φ is the cumulative distribution function of a standard normal random variable. We set $\boldsymbol{q}(X, Z) = (1, X, Z, XZ)'$ and the true parameter values as $\boldsymbol{\eta} = (0, -0.2, \bar{\eta}, -0.2\bar{\eta})'$ for varying $\bar{\eta} > 0$. The further $\bar{\eta}$ is away from zero, the greater the instrument strength and the greater the variation in the propensity score for a given covariate level.

Given this simulation design, the correct specification of the observable mean function $m(x, p)$ in Equation (4) was

$$m(x, p) = 0.3 + 0.1x - 0.1p + 0.1xp + 0.3p^2.$$

Differentiating $m(x, p)$ with respect to p and evaluating at $p = v$ gives the true MTE function

$$\theta_{\text{MTE}}(x, v) = -0.1 + 0.1x + 0.6v.$$

From this, we see that the MTE curve is increasing in unobserved heterogeneity V. If *lower* values of the outcome variable are favourable, then this describes a setting of "selection on gains", i.e. those who benefit most from treatment are those who choose to be treated. By averaging the MTE function over covariate X and unobserved heterogeneity V, we find that the true ATE was 0.25, and the true values of other target estimands varied with instrument strength $\bar{\eta}$. As we increase instrument strength from $\bar{\eta} = 0.2$ to $\bar{\eta} = 1$, the ATT ranged from 0.088 to 0.135, the ATU from 0.388 to 0.352, and the ASG from -0.300 to -0.217.

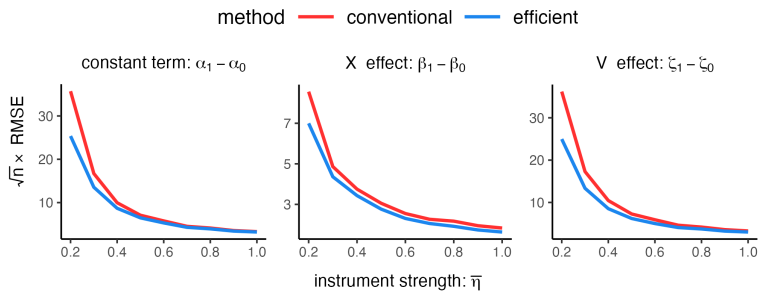

FIG 1. *Root mean squared error (RMSE) of MTE parameter estimates over 1000 simulation experiments for each value of instrument strength* $\bar{\eta}$.

4.2. *Estimation of MTE Parameters.* The propensity score was estimated by a nonparametric kernel regression using the `np` R package (Hayfield and Racine, 2008). One advantage of efficient estimation is that it is designed to be insensitive as the propensity score estimates $\pi_{\widehat{\mathcal{P}}}(\boldsymbol{X}, Z)$ move slightly away from the truth $\pi_{\mathcal{P}}(\boldsymbol{X}, Z)$. What this means in practice is that the efficient estimator of MTE parameters $\widetilde{\boldsymbol{\gamma}}$ should be more accurate than the conventional estimator $\widehat{\boldsymbol{\gamma}}$ when instruments are relatively weak, where some regions of the propensity score are estimated with fewer observations. This advantage of the efficient approach over the conventional approach is what we demonstrate in this section.

For bandwidth selection, for each simulation run we computed bandwidth constants using cross-validation on 3 random samples of size $1,000$ from a full sample of $n = 10,000$, and the bandwidths for the full sample size were scaled down from the constant according to the optimal rate; for further details, see the discussion of the `bwscaling` argument in Hayfield and Racine (2008). The robustness property of the efficient approach means that its estimates should be relatively insensitive to the choice of bandwidth. Supplementary Figures S4 and S5 show that efficient estimators of MTE and target parameters continue to have low bias when the chosen bandwidth is away from the optimal level, especially when oversmoothing.

Figure 1 shows that the efficient estimates of MTE parameters are better than the conventional estimates in terms of achieving a lower mean squared error (MSE). The improvement from using the efficient approach is greater when instruments are weaker. As instrument strength increases, the MSE from both methods converge toward zero as expected. Since the efficient and conventional approach have the same asymptotic variance, we might expect any differences in MSE to be driven by bias. Supplementary Figure S1 shows that the bias is indeed lower for efficient estimates when instruments are sufficiently strong ($\bar{\eta} \geq 0.4$), and especially for the V effect, the conventional estimates were considerably more biased even in the strongest instrument case ($\bar{\eta} = 1$).

4.3. *Estimation of Target Estimands.* Given that the target estimands ATE, ATT, ATU, and ASG are linear functions of the MTE parameters $\boldsymbol{\gamma}$, we would expect the estimation performance of MTE parameters is closely linked with the estimation performance of our target estimands. However, the asymptotic variances of the conventional and efficient approaches are now no longer the same, and hence we now expect a greater advantage from efficient estimation in terms of MSE performance.

Figure 2 shows this is the case for all four estimands, ATE, ATT, ATU, and ASG. For example, under very weak instruments ($\bar{\eta} = 0.2$), the average root-MSE (RMSE) of the efficient estimates was around 40% lower than that of the conventional ones for all target estimands. Supplementary Figure S2 shows the bias of efficient estimates hovered closely around zero

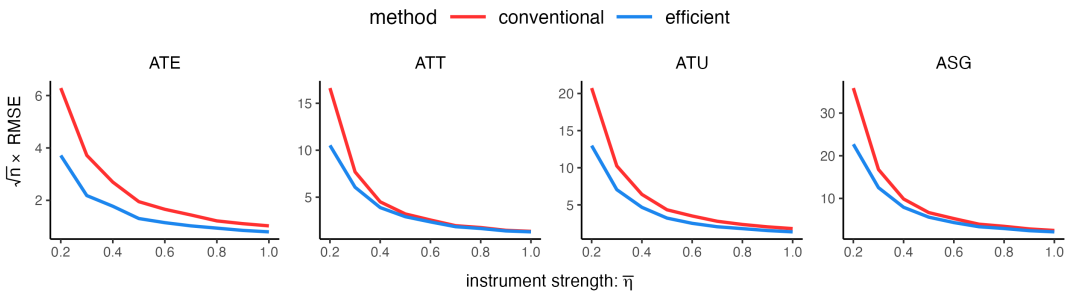

FIG 2. *Root mean squared error (RMSE) of target estimand point estimates over 1000 simulation experiments for each value of instrument strength* $\bar{\eta}$.

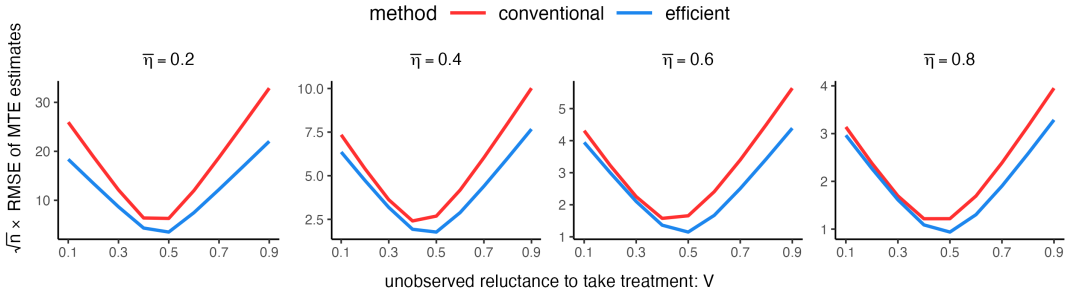

FIG 3. *Root mean squared error (RMSE) of MTE estimates varying with* V *and averaged over covariates,* $E[Y_1 - Y_0|V]$.

for all estimands when instruments were sufficiently strong ($\bar{\eta} \geq 0.6$). In certain weak instrument cases, the bias of the conventional approach was lower than that of the efficient approach, especially for the ATT estimand. Thus, given its superior MSE performance in Figure 2, the variance of the efficient estimates was considerably lower under weak instruments.

To compare estimation performance on the MTE estimand itself, Figure 3 plots the average RMSE of estimates of the MTE estimand averaged over covariates X, i.e. $E_{\mathcal{P}}[Y_1 - Y_0|V]$.

Figure 3 shows that the MTE estimates had a lower MSE than the conventional estimates. Both estimators are more precise where most observations are concentrated (where V is around 0.5), and are less precise when estimating the treatment effect for those with extreme values of V. For the case of $\bar{\eta} = 0.2$, observed propensity scores, on average, ranged from 0.211 to 0.769, and under this greater sampling uncertainty in the tails of propensity scores, the gap between the MSE of the two approaches is greater.

The observed propensity scores had close to full support over the 0 to 1 range when instruments were sufficiently strong ($\bar{\eta} \geq 0.8$); when $\bar{\eta} = 0.8$, 1.6% of samples had a propensity score less than 0.05 and 1.1% of samples had a propensity score over 0.95. In this setting, not only is the efficient estimator better in terms of MSE, but it performs better in terms of bias (Supplementary Figure S3).

Overall, our simulation results highlight the benefits of efficient estimation of MTE parameters and related target estimands. In particular, efficient estimates have a lower MSE in finite samples compared with the conventional plug-in propensity score approach, and in most cases a lower bias. The efficient estimates are also more robust to weak instrument settings, and to the choice of tuning parameters in nonparametric propensity score estimation.

5. Empirical Application: Estimating Effects of Excessive Alcohol Consumption on Blood Pressure. Estimating the causal effects of excessive alcohol consumption is complicated by self-selection into drinking and heterogeneity in responses to health advice. A generalized Roy model can capture the dynamic that alcohol consumption behaviour may influence, and be influenced by, worse health outcomes. The UK National Health Service advises adults to drink no more than 14 units of alcohol per week (where 1 unit is 8g of pure alcohol). Therefore, we work with a binary treatment indicator for "excessive alcohol consumption" defined as self-reported alcohol consumption of at least 14 units of alcohol per week. Systolic blood pressure measurements were used as the outcome variable. Here, unobserved heterogeneity V can be thought as describing heterogeneity in "health consciousness" levels: within groups sharing similar observable covariate values or characteristics, those with high values of V are less likely to excessively consume alcohol compared with those with low values of V.

We use a genetic score for alcohol consumption as a continuous instrument for excessive alcohol consumption, and estimate MTE parameters and other target estimands to study the potentially heterogeneous effects of excessive alcohol consumption on systolic blood pressure. In Supplementary Material, we present results under a more general model for the MTE that allows the effects of unobserved heterogeneity V to differ for women and men, however we did not find evidence for a sex difference in the MTE slope-in-V.

5.1. *Data.*

5.1.1. *UK Biobank and Alcohol Consumption.* We considered data on 367,703 genotyped participants in the UK Biobank of European ancestry. After removing individuals that provided incomplete data on alcohol consumption, systolic blood pressure, and antihypertensive medication use, we were left with complete data on 294,563 individuals. The treatment indicator of excessive alcohol consumption was defined as a binary variable that equalled 1 when an individual's self-reported alcohol consumption was more than 14 units of alcohol per week on average, and zero otherwise. Under this definition, 53.5% of the 294,563 participants were consuming alcohol excessively.

5.1.2. *Instrument Selection.* The motivation of using genetic instruments is that genetic variants, which are fixed at conception, may be less prone to unmeasured confounding and reverse causation, and may provide a source of exogenous variation in treatment choice (Davey Smith and Ebrahim, 2003). Our analysis used a genetic score for alcohol consumption composed of 93 genetic variants, as an instrumental variable for excessive alcohol consumption, as previously used in Topiwala et al. (2022b). The 93 genetic variants were each marginally associated with alcohol consumption (log drinks per week) at a p-value less than the genome-wide significance threshold 5×10^{-8} in a large published genome-wide association study (Liu et al., 2019). The individual variants used in the genetic score were not strongly correlated (the correlation between any two variants was $R^2 \leq 0.1$). The genetic score was robustly associated with a continuous measure of alcohol consumption in UK Biobank participants on average, with an F-statistic of 1862.

5.1.3. *Outcome Selection.* We considered systolic blood pressure (SBP) as the outcome variable. For individuals on antihypertensive medication, we added 10mmHg to their SBP measurements since most antihypertensive drugs lower systolic blood pressure by around that amount (Dimmitt et al., 2019). The average SBP over the 294,563 UK Biobank participants considered for our analysis was 139.98 mmHg. Approximately 45.5% of the sample had SBP over 140 mmHg, a commonly used indicator for hypertension. There were sex differences in

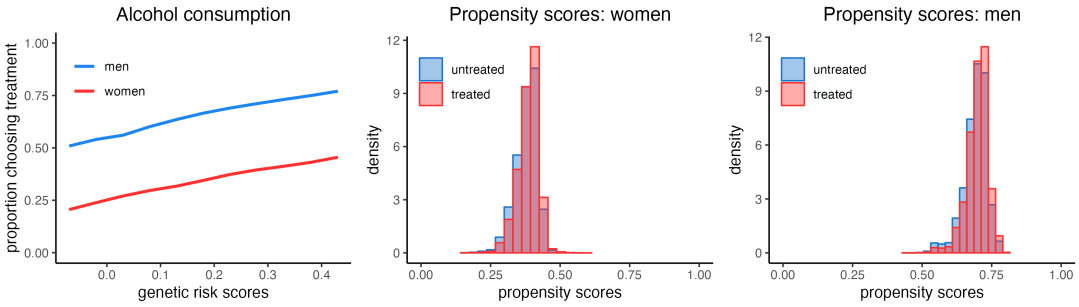

FIG 4. *Excessive alcohol consumption (treatment) varying with the genetic score and sex (left), and the common support of observed propensity scores for women (middle) and for men (right).*

SBP measurements; the average SBP estimate was 136.60 mmHg among female participants, and 143.51 mmHg among male participants. We standardized the outcome so that each unit change in the outcome represents a standard deviation change in SBP, where 1 standard deviation of SBP equalled 19.759 mmHg.

5.1.4. *Covariate Selection.* We considered sex as a single covariate $\boldsymbol{X} = X_{woman}$, a binary indicator for female participants. We also considered age and geographical location (a binary indicator if they lived north of Stoke in the UK) as potential covariates in addition to sex, but their impact on alcohol consumption behaviour was not estimated to be relevant.

Sex is a candidate covariate since it is determined at conception and so it is not a post-randomization variable. Moreover, as a single binary covariate it leads to no curse-of-dimensionality issues for nonparametric estimation of the propensity score. We may expect sex to explain treatment effect heterogeneity because women and men metabolize alcohol differently due to biological differences (Graham et al., 1998; Briasoulis, Agarwal and Messerli, 2012). Finally, from a policy perspective, sex differences in the harms of excessive alcohol consumption can inform more targeted public-health guidance. Around 51.1% of participants were women, and around 48.9% were men.

5.2. *Predictors of Excessive Alcohol Consumption.* Estimation of MTEs is more precise when there is large variation in the propensity score attributable to the genetic instrument and covariates. Figure 4 plots how the propensity to drink excessively varies with sex and the genetic score.

Figure 4 shows that for both men and women, a higher genetic score is associated with a greater propensity for excessive alcohol consumption, which is unsurprising given that the genetic score was calculated for alcohol consumption. For a given genetic score, women were considerably less likely to drink excessively than men. We conclude that sex and the weighted genetic score are conditionally relevant predictors of excessive alcohol consumption. While there is an even match in the distributions of treated and untreated groups, the variation in the propensity score conditional on sex is quite limited. Therefore estimating a more general MTE curve that allows for $V \times X$ interactions could be challenging (see Supplementary Material for results under this more general model).

5.3. *Main Results.* The propensity score was estimated using a nonparametric kernel regression (of excessive alcohol consumption on sex and the genetic score) using the `np` R package (Hayfield and Racine, 2008). For bandwidth selection, we computed bandwidth constants using cross-validation on 20 random samples (without replacement) of size $10,000$ from a full sample of $n = 294,563$, and the bandwidths for the full sample size were scaled

TABLE 2
Efficient estimates of MTE parameters from the model
$\theta_{MTE}(X, V) \sim$ *constant term* $+ X_{woman} + V_{health\text{-}consciousness}$

MTE Parameter	Constant term	Woman (X)	Health consciousness (V)
Estimate	1.905	−0.815	−2.215
Standard Error	0.728	0.340	1.080
95% CI-Lower	0.477	−1.482	−4.331
95% CI-Upper	3.332	−0.148	−0.098

TABLE 3
Efficient estimates of target estimands from the model
$\theta_{MTE}(X, V) \sim$ *constant term* $+ X_{woman} + V_{health\text{-}consciousness}$

Estimand	ATE	ATT	ATU	ASG
Estimate	0.379	0.952	−0.282	1.235
Standard Error	0.051	0.341	0.349	0.684
95% CI-Lower	0.279	0.283	−0.966	−0.105
95% CI-Upper	0.479	1.621	0.401	2.575

down from the constant according to the optimal rate; for further details, see the discussion of the `bwscaling` argument in Hayfield and Racine (2008).

From Table 2, the MTE of excessive alcohol consumption on systolic blood pressure is estimated to be positive for most individuals, apart from those with high values of health consciousness V. The MTE curve is estimated to be decreasing in V, so that any adverse effect of excessive alcohol consumption on systolic blood pressure is lower for health conscious individuals. This suggests evidence of "reverse selection on gains": individuals that are less health conscious may experience worse health outcomes (greater changes in systolic blood pressure) due to excessive alcohol consumption than more health conscious individuals.

We also estimate sex differences in the treatment effect, where the adverse effect of excessive alcohol consumption on systolic blood pressure is estimated to be worse for men than for women. One potential explanation of this finding is that excessive alcohol drinking may involve episodic drinking (i.e. binge drinking) in a higher proportion of men and those with low health consciousness, and such drinking has a greater effect on blood pressure compared to steady levels of high alcohol consumption.

Table S1 in Supplementary Material presents MTE parameter estimates from a more general model allowing for interaction effects between sex (X) and unobserved health consciousness (V). The effects of sex and health consciousness are directionally the same as those estimated in Table 2, although there is larger uncertainty in the estimates. We estimate no interaction effect between X and V, so we are unable to conclude the slope of the MTE curve in terms of health consciousness levels V is different for female participants compared to male participants.

Table 3 presents estimates of target estimands. For the ATE, genetically-predicted excessive alcohol consumption is associated with higher systolic blood pressure measurements by 0.379 standard deviations (p-value < 0.001), which is 7.489 mmHg. This is consistent with experimental evidence that excessive alcohol consumption leads to higher blood pressure levels (Tasnim et al., 2020). We also note that the observational association between excessive alcohol consumption and systolic blood pressure is $E_{\widehat{\mathcal{P}}}[Y|A=1] - E_{\widehat{\mathcal{P}}}[Y|A=0] = 0.186$, or 3.675 mmHg.

The ATT was estimated to be positive (p-value: 0.005), and while there is greater uncertainty regarding the ATU estimate, results broadly suggest a positive estimate of the ASG

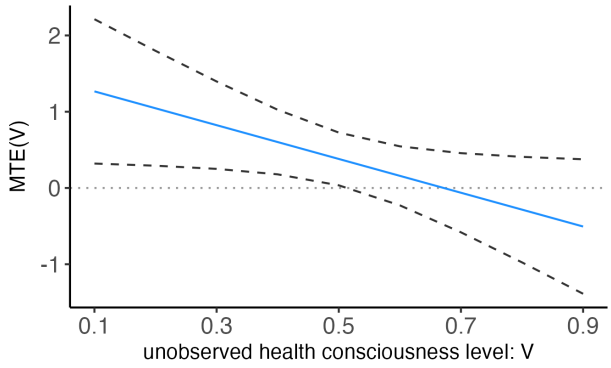

FIG 5. *Efficient MTE estimates varying with V and averaged over covariates, $E[Y_1 - Y_0|V]$ (solid blue line). The dashed black curves indicate 95% asymptotic confidence intervals.*

(p-value: 0.071), which is consistent with the reverse selection on gains results inferred directly from MTE parameters in Table 2. Table S2 in Supplementary Material shows that in an MTE model allowing for $X \times V$ interaction effects, the estimates of the target estimands ATE, ATT, ATU, and ASG were similar to those estimated in Table 2. This supports the results of Table S1 that suggested no evidence for sex differences in the MTE slope-in-V.

Figure 5 illustrates that the harmful effects of excessive alcohol consumption are estimated to be decreasing in health consciousness levels. The 95% asymptotic confidence intervals (dashed black lines) show how the MTE is estimated with less precision at the tail ends of V which again reflects the lack of support in the propensity score at those regions. Supplementary Figure S6 plots the MTE curve under a more general model allowing for $X \times V$ interaction effects, and this estimated curve is very close to that of Figure 5 which supports our MTE specification with separable X and V effects.

To compare our results to a parametric first stage, we re-estimated the propensity score using logit and probit specifications. The resulting treatment effect estimates are similar to those obtained under our nonparametric baseline, both in magnitude and precision, with parametric specifications suggesting slightly stronger evidence of reverse selection on gains (Supplementary Figure S7).

5.4. *Sensitivity Analysis.* Our empirical results so far have assumed that our genetic instrument does not affect systolic blood pressure other than through its effect on the choice to drink alcohol excessively. This is more credible when the mechanistic or biological effects of a genetic instrument are better known (Gill et al., 2021), as with the genetic variant rs1229984 from the *ADH1B* gene region, which received the largest weight in our genetic score.

Nevertheless, as in most Mendelian randomization analyses, horizontal pleiotropy remains a concern, and it is possible that some variants in the genetic score may affect systolic blood pressure through their effects on traits other than alcohol consumption. We present the sensitivity of treatment effect estimates to a setting in which the instrument has a direct effect τ on systolic blood pressure, in a way that is the same whether or not someone chooses to drink alcohol excessively, as described in Section 3.3. The practical implication of this relaxation of the exclusion restriction assumption is that individual variants in the genetic score can affect blood pressure through factors unrelated to the effects of excessive drinking. We allow τ to lie in the interval $[-\bar{\tau}, \bar{\tau}]$ for some positive constant $\bar{\tau}$. To inform a choice of $\bar{\tau}$, we combine two perspectives.

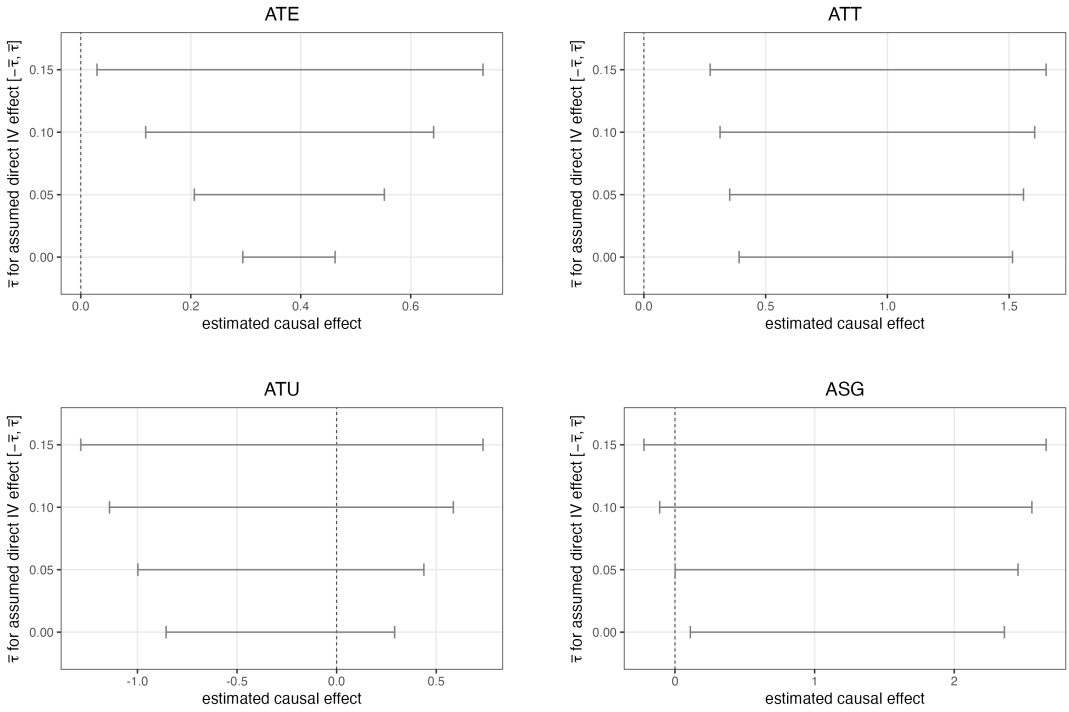

FIG 6. *Sensitivity analysis of ATE, ATT, ATU, and ASG estimates under the assumption that the direct instrument effect τ is the same for both potential outcomes, and lies in the interval $[-\bar{\tau}, \bar{\tau}]$ for choices $\bar{\tau} = 0.00, 0.05, 0.10,$ and 0.15. Each panel shows 90% confidence intervals for that estimand.*

First, we consider potential pleiotropic pathways by which a genetic score for alcohol consumption could affect systolic blood pressure levels. A unit increase in the genetic score is associated with higher BMI levels (estimate: 0.93 kg/m²; p-value: < 0.001), and is associated with fewer years of education (estimate: -0.30; p-value: < 0.001). A unit higher BMI level (kg/m²) is associated with higher systolic blood pressure levels by 1.02 mmHg (p-value: < 0.001), and an extra year of education is associated with lower systolic blood pressure levels by 0.66 mmHg (p-value: < 0.001). Overall, these estimates roughly suggest a unit change in the genetic score may lead to at most 1 mmHg change in systolic blood pressure levels through each of these pathways.

Second, we note that a linear regression of systolic blood pressure on the genetic score estimates that a unit increase in the genetic score is associated with higher blood pressure levels by 3.03 mmHg, which suggests a natural limit of the direct effect $\bar{\tau}$ of around 3 mmHg, if we further assume any causal effects of genetically-predicted excessive alcohol consumption are in the same direction as our observed association between the genetic score and systolic blood pressure.

Translating these associations from mmHg to our standardized systolic blood pressure measurements, our sensitivity analysis sets $\bar{\tau}$ to at most 0.15. Setting $\bar{\tau} = 0.05$ allows a unit change in the genetic score to directly affect systolic blood pressure levels by roughly 1 mmHg in either direction. Similarly, setting $\bar{\tau} = 0.10$ allows the instrument to have a direct effect on systolic blood pressure of roughly 2 mmHg, while setting $\bar{\tau} = 0.15$ allows a roughly 3 mmHg change.

Figure 6 presents 90% confidence intervals of our target estimands under different choices of $\bar{\tau}$; since these intervals take a union over $|\tau| \leq \bar{\tau}$, they are conservative under the maintained assumption, and we report them at the 90% rather than the 95% level. The confidence intervals are increasing in width with $\bar{\tau}$; for example, a 90% confidence interval for

the ATE effect widens from $[0.295, 0.463]$ when $\bar{\tau} = 0$ to $[0.029, 0.731]$ when $\bar{\tau} = 0.15$. These widened confidence intervals reflect greater uncertainty over the exclusion restriction assumption. The conclusions of our main analyses in Table 3 continue to hold, although there is more uncertainty in ASG estimates for larger values of $\bar{\tau}$.

5.5. *Summary.* Our empirical findings have implications both for the applied MTE literature and public health policies aimed at curbing excessive alcohol consumption. While our positive ATE estimate supports existing evidence that excessive alcohol consumption raises blood pressure levels, our more novel finding suggests heterogeneity in how this harm is distributed. In our setting, the unobserved "health consciousness" variable V orders individuals by their reluctance to engage in excessive alcohol consumption, from those who drink excessively almost regardless of low genetic propensity (low V) to those who almost never do even with high genetic propensity (high V). The economically interpretable object is the slope-in-V of the MTE estimand. We estimate a downward-sloping MTE curve in V, which we can interpret as "reverse selection on gains", i.e. selection that is adverse for health, since the individuals most prone to drink excessively are those for whom excessive drinking is most damaging. Our model implies the ordering ATT $\geq$ ATE $\geq$ ATU, and our point estimates are consistent with this ranking.

In this case, population-average effects, such as the ATE, understate the harm actually experienced by some excessive drinkers. Policies that reduce excessive drinking in this group would therefore prevent more harm per person than an ATE calculation alone would suggest. However, this group may also be the hardest to target through public health messaging or economic policy; for example, if less health-conscious individuals (those with low V) are also less responsive to alcohol price increases. Our results indicate where the greatest gains lie, but not which interventions would best achieve them. The estimated sex difference, with larger adverse effects among men, also provides an observable characteristic on which guidance or screening might be prioritized.

Finally, we caution that our treatment effect estimates compared the difference between genetically defined subgroups that may reflect a more long-term, sustained effect of high alcohol consumption on blood pressure (Burgess et al., 2023). As such, these estimates are not easily comparable to those from a short-term intervention of a clinical trial (Tasnim et al., 2020).

6. Conclusion. The MTE framework (Heckman and Vytlacil, 1999, 2005) is able to model the plausible scenario that the effect of a treatment may vary with both observed and unobserved heterogeneity that influences selection into treatment. It is a popular method to estimate causal effects in observational studies, and has been applied to study a wide range of topics including education and healthcare; see Shea and Torgovitsky (2023) for a detailed list of applications.

As with any instrumental variable method, the credibility of an applied analysis relies on the instrument being valid and relevant. In epidemiology, the Mendelian randomization paradigm has popularized the use of genetic variants as instrumental variables. However, the vast majority of Mendelian randomization applications have estimated causal effects in a *homogeneous* effects model where the treatment effect is assumed to be the same for all individuals, due in part to easy-to-use software that can be applied to readily available summary data from genome-wide association studies.

The contribution of our work was four-fold. First, we demonstrated the use of genetic instruments to estimate *heterogeneous* causal effects in the MTE framework, which studies the relationship between treatment choice and treatment effects. This introduces a credible approach to study effect heterogeneity in Mendelian randomization studies with a binary exposure.

Second, genetic instruments will typically be weak in the sense that they are unlikely to cause large shifts in the propensity scores of treatment, which can lead to issues for estimating the MTE function. Thus, to reduce sensitivity to the first stage uncertainty at the tails of propensity score estimates, we derived the efficient influence functions of target estimands in linear-in-parameter models of MTEs. Our work has clarified that this efficient approach can be more robust for estimation of target estimands, while being computationally straightforward. We have also developed an R package (github.com/ash-res/efficient.mte/) for straightforward application of the methods we discussed. While this work was motivated by the specific case of a genetic instrument, the methods proposed in this paper are applicable to a wider class of instrumental variable studies.

Third, we have developed a sensitivity analysis based on Conley, Hansen and Rossi (2012) that allows researchers to assess the robustness of their treatment effects estimates to plausible violations of the exclusion restriction assumption due to pleiotropy. Wherever possible, biologically-plausible genetic variants, where their mechanistic effects on treatment are well-understood, should be prioritized for instrument selection (Gill et al., 2021). However, we expect that in many applications this will not be possible when using genetic instruments. Hence, we recommend that researchers use this approach to report a sensitivity analysis in MTE applications when using a genetic instrument.

Fourth, our empirical analysis suggests genetic evidence for heterogeneity in the harmful effects of excessive alcohol consumption, such that individuals that are prone to drink alcohol excessively may experience greater changes in systolic blood pressure compared to individuals that are less prone to drink excessively. Given that high alcohol consumption is known to be significantly related with many adverse health conditions globally (Anderson et al., 2023), this highlights the importance of policies to target those more vulnerable to excessively consume alcohol.

Acknowledgments. We thank Jack Bowden for helpful discussions. We thank the editor, associate editor, and two anonymous referees for detailed comments.

Funding. SB was supported by the Wellcome Trust (225790/Z/22/Z). This research was funded by the United Kingdom Research and Innovation Medical Research Council (MC-UU-00002/7).

SUPPLEMENTARY MATERIAL

Supplementary Material A
This contains proofs of theoretical results, further details on our sensitivity analysis, further simulation results, and further empirical results under a more general MTE model.

Supplementary Material B
This contains R code to apply the efficient MTE estimation method proposed in this paper.

REFERENCES

ANDERSON, B. O., BERDZULI, N., ILBAWI, A., KESTEL, D., KLUGE, H. P., KRECH, R., MIKKELSEN, B., NEUFELD, M., POZNYAK, V. and REKVE, D. (2023). Health and cancer risks associated with low levels of alcohol consumption. *The Lancet Public Health* **8** e6–e7.

BASU, A., HECKMAN, J. J., NAVARRO-LOZANO, S. and URZUA, S. (2007). Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics* **16** 1133–1157.

BRIASOULIS, A., AGARWAL, V. and MESSERLI, F. H. (2012). Alcohol consumption and the risk of hypertension in men and women: a systematic review and meta-analysis. *The Journal of Clinical Hypertension* **14** 792–798.

BRINCH, C. N., MOGSTAD, M. and WISWALL, M. (2017). Beyond LATE with a discrete instrument. *Journal of Political Economy* **125** 985–1039. https://doi.org/10.1086/692712

BURGESS, S. and LABRECQUE, J. A. (2018). Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *European Journal of Epidemiology* **33** 947–952.

BURGESS, S., FOLEY, C. N., ALLARA, E., STALEY, J. R. and HOWSON, J. M. M. (2020). A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications* **11**.

BURGESS, S., MASON, A. M., GRANT, A., SLOB, E. A. W., GKATZIONIS, A., ZUBER, V., PATEL, A., TIAN, H., LIU, C., HAYNES, W. G., HOVINGH, G. K., KNUDSEN, L. B., WHITTAKER, J. C. and GILL, D. (2023). Using genetic association data to guide drug discovery and development: Review of methods and applications. *American Journal of Human Genetics* **110** 195–214.

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. *Econometrics Journal* **21** C1–C68. https://doi.org/10.1111/ectj.12097

CHERNOZHUKOV, V., ESCANCIANO, J. C., ICHIMURA, H., NEWEY, W. K. and ROBINS, J. M. (2022). Locally robust semiparametric estimation. *Econometrica* **90** 1501–1535. https://doi.org/10.3982/ecta16294

CONLEY, T. G., HANSEN, C. B. and ROSSI, P. E. (2012). Plausibly exogenous. *The Review of Economics and Statistics* **94** 260–272.

DAVEY SMITH, G. and EBRAHIM, S. (2003). Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32** 1–22. https://doi.org/10.1093/ije/dyg070

DEPALO, D. (2020). Explaining the causal effect of adherence to medication on cholesterol through the marginal patient. *Health Economics* **29** 110–126.

DI FEDERICO, S., FILIPPINI, T., WHELTON, P. K., CECCHINI, M., IAMANDII, I., BORIANI, G. and VINCETI, M. (2023). Alcohol intake and blood pressure levels: a dose-response meta-analysis of nonexperimental cohort studies. *Hypertension* **80** 1961–1969.

DIMATTEO, M. R. (2004). Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Medical Care* **42** 200–209.

DIMMITT, S. B., STAMPFER, H. G., MARTIN, J. H. and FERNER, R. E. (2019). Efficacy and toxicity of antihypertensive pharmacotherapy relative to effective dose 50. *British Journal of Clinical Pharmacology* **85** 2218–2227.

EDENBERG, H. J. and FOROUD, T. (2013). Genetics and alcoholism. *Nature Reviews Gastroenterology and Hepatology* **10** 487–494.

GILL, D., GEORGAKIS, M. K., WALKER, V. M., SCHMIDT, A. F., GKATZIONIS, A. and DAVIES ET AL., N. M. (2021). Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Research* **6** 1–19.

GRAHAM, K., WILSNACK, R., DAWSON, D. and VOGELTANZ, N. (1998). Should alcohol consumption measures be adjusted for gender differences? *Addiction* **93** 1137–1147.

HAYFIELD, T. and RACINE, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software* **27** 1–32.

HECKMAN, J. J. and VYTLACIL, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences of the United States of America* **96** 4730–4734. https://doi.org/10.1073/pnas.96.8.4730

HECKMAN, J. J. and VYTLACIL, E. (2001). Policy-relevant treatment effects. *American Economic Review* **91** 107–111.

HECKMAN, J. J. and VYTLACIL, E. J. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* **73** 669–738. https://doi.org/10.1111/j.1468-0262.2005.00594.x

HEMANI, G., STENDER, S., WOLTERS, F. J., HOFMAN, A. and DAVEY SMITH, G. (2025). The rapid growth in Mendelian randomization studies. *European Journal of Epidemiology* 1–7.

HINES, O., DUKES, O., DIAZ-ORDAZ, K. and VANSTEELANDT, S. (2022). Demystifying statistical learning based on efficient influence functions. *American Statistician* **76** 292–304. https://doi.org/10.1080/00031305.2021.2021984

IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62** 467–475. https://doi.org/10.2307/2951620

JOHAR, M. and MARUYAMA, S. (2014). Does coresidence improve an elderly parent's health? *Journal of Applied Econometrics* **29** 965–983.

KANG, H., ZHANG, A., CAI, T. T. and SMALL, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American Statistical Association* **111** 132–144.

KANG, H., GUO, Z., LIU, Z. and SMALL, D. (2025). Identification and inference with invalid instruments. *Annual review of statistics and its application* **12** 385–405.

KASSAW, N. A., ZHOU, A., MULUGETA, A., LEE, S. H., BURGESS, S. and HYPPÖNEN, E. (2024). Alcohol consumption and the risk of all-cause and cause-specific mortality: a linear and nonlinear Mendelian randomization study. *International Journal of Epidemiology* **53** dyae046.

KOWALSKI, A. E. (2023). Reconciling seemingly contradictory results from the Oregon health insurance experiment and the Massachusetts health reform. *Review of Economics and Statistics* **105** 646–664.

LARSSON, S. C., BUTTERWORTH, A. S. and BURGESS, S. (2023). Mendelian randomization for cardiovascular diseases: principles and applications. *European Heart Journal* **44** 4913–4924.

LARSSON, S. C., BURGESS, S., MASON, A. M. and MICHAËLSSON, K. (2020). Alcohol consumption and cardiovascular disease: a Mendelian randomization study. *Circulation: Genomic and Precision Medicine* **13** e002814.

LARSSON, S. C., MASON, A. M., CRONJÉ, H. T., BASSETT, E., HORTA, G., KAR, S. and BURGESS, S. (2025). Alcohol consumption and risk of cancer: a Mendelian randomization analysis of four biobanks and consortium data. *BMC medicine* **23** 1–15.

LIU, M., JIANG, Y., WEDOW, R., LI, Y., BRAZEL, D. M., CHEN, F., DATTA, G., DAVILA-VELDERRAIN, J., MCGUIRE, D. and TIAN, C. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics* **51** 237–244.

MOGSTAD, M. and TORGOVITSKY, A. (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics* **10** 577–613. https://doi.org/10.1146/annurev-economics-101617-041813

MORRISON, J., KNOBLAUCH, N., MARCUS, J. H., STEPHENS, M. and HE, X. (2020). Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics* **52** 740–747. https://doi.org/10.1038/s41588-020-0631-4

RIETSCHEL, M. and TREUTLEIN, J. (2013). The genetics of alcohol dependence. *Annals of the New York Academy of Sciences* **1282** 39–70.

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

SCHAEFER, S. M., KAISER, A., BEHRENDT, I., EICHNER, G. and FASSHAUER, M. (2023). Association of alcohol types, coffee and tea intake with mortality: prospective cohort study of UK Biobank participants. *British Journal of Nutrition* **129** 115–125.

SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94** 1096–1120. https://doi.org/10.1080/01621459.1999.10473862

SHEA, J. and TORGOVITSKY, A. (2023). ivmte: An R package for extrapolating instrumental variable estimates away from compliers. *Observational Studies* **9** 1–42.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P. and DANESH ET AL., J. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* **12** e1001779.

TASNIM, S., TANG, C., MUSINI, V. M. and WRIGHT, J. M. (2020). Effect of alcohol on blood pressure. *Cochrane Database of Systematic Reviews* **7** 1–140.

TOPIWALA, A., EBMEIER, K. P., MAULLIN-SAPEY, T. and NICHOLS, T. E. (2022a). Alcohol consumption and MRI markers of brain structure and function: cohort study of 25,378 UK Biobank participants. *NeuroImage: Clinical* **35** 103066.

TOPIWALA, A., TASCHLER, B., EBMEIER, K. P., SMITH, S., ZHOU, H., LEVEY, D. F., CODD, V., SAMANI, N. J., GELERNTER, J., NICHOLS, T. E. and BURGESS, S. (2022b). Alcohol consumption and telomere length: Mendelian randomization clarifies alcohol's effects. *Molecular Psychiatry* **27** 4001–4008.

VERBANCK, M., CHEN, C. Y., NEALE, B. and DO, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* **50** 693–698.

WILDING, A., MUNFORD, L. and SUTTON, M. (2023). Estimating the heterogeneous health and well-being returns to social participation. *Health Economics* **32** 1921–1940.

WU, Z. and WANG, J. (2024). Interpretation of two-sample Mendelian randomization for binary exposures and outcome. *bioRxiv* 1–16.

Supplementary Material

Ashish Patel, Francis J. DiTraglia & Stephen Burgess

Notation and Preliminaries

For any vectors of observed variables A and B, let $var_{\mathcal{P}}(A) = E_{\mathcal{P}}[(A - E_{\mathcal{P}}(A))(A - E_{\mathcal{P}}(A))']$, and let $cov_{\mathcal{P}}(A, B) = E_{\mathcal{P}}[(A - E_{\mathcal{P}}(A))(B - E_{\mathcal{P}}(B))']$. We also note their sample averages as $var_{\widehat{\mathcal{P}}}(A)$ and $cov_{\widehat{\mathcal{P}}}(A, B)$. Then, given the linear model $m(x, p) = r(x, p)'\gamma$ for the observable mean $m(x, p) = E_{\mathcal{P}}[Y|X = x, \pi_{\mathcal{P}}(X, Z) = p]$, we can define a target estimand $\gamma = \gamma(\mathcal{P})$ as

$$\gamma(\mathcal{P}) = var_{\mathcal{P}}(r(X, \pi_{\mathcal{P}}(X, Z)))^{-1} cov_{\mathcal{P}}(r(X, \pi_{\mathcal{P}}(X, Z)), Y).$$

To calculate the efficient influence function of $\gamma(\mathcal{P})$, we follow the *point mass contamination* strategy recommended by Ichimura and Newey (2022) and Hines et al. (2022). Let $\widetilde{\mathcal{P}}$ denotes a fixed distribution that has a point mass at a single observation $\widetilde{o} = (\tilde{a}, \tilde{x}, \tilde{y}, \tilde{z})$. We wish to perturb the estimand $\gamma(\mathcal{P})$ in the direction $\widetilde{\mathcal{P}}$ by using a mixture model $\mathcal{P}_t = t\widetilde{\mathcal{P}} + (1 - t)\mathcal{P}$ indexed by $t \in [0, 1]$. Then, the efficient influence function at observation $\widetilde{o}$ is given by the Gateaux derivative

$$\phi(\widetilde{o}, \mathcal{P}) = \frac{\partial \gamma(\mathcal{P}_t)}{\partial t}\Big|_{t=0}.$$

The efficient influence function $\phi(\widetilde{o}, \mathcal{P})$ characterises how sensitive the estimand $\gamma(\mathcal{P})$ is to changes in the data-generating distribution $\mathcal{P}$ at observation $\widetilde{o}$, and it plays an important part in the first-order asymptotic properties of an estimator of $\gamma(\mathcal{P})$. The term

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(O_i, \widehat{\mathcal{P}})$$

is called the *drift term* of a von Mises expansion of the nonparametric estimator of $\gamma(\mathcal{P})$. One way to construct an efficient estimator of $\gamma(\mathcal{P})$ is to choose the estimator $\gamma(\widehat{\mathcal{P}})$ that sets this drift term to zero. This is the approach we now take.

Setting the drift term to zero, we find the estimator $\widetilde{\gamma} \equiv \gamma(\widehat{\mathcal{P}})$ as the solution to the estimating equation

$$\begin{aligned} 0 \quad = \quad & -var_{\widehat{\mathcal{P}}}(r(X, \pi_{\widehat{\mathcal{P}}}(X, Z)))^{-1} cov_{\widehat{\mathcal{P}}}\Big(\frac{\partial r(X, \pi_{\widehat{\mathcal{P}}}(X, Z))}{\partial \pi}(A - \pi_{\widehat{\mathcal{P}}}(X, Z)), r(X, \pi_{\widehat{\mathcal{P}}}(X, Z))\Big)\gamma(\widehat{\mathcal{P}}) \\ & -\gamma(\widehat{\mathcal{P}}) + var_{\widehat{\mathcal{P}}}(r(X, \pi_{\widehat{\mathcal{P}}}(X, Z)))^{-1} cov_{\widehat{\mathcal{P}}}(r(X, \pi_{\widehat{\mathcal{P}}}(X, Z)), Y). \end{aligned}$$

Note that this is a different estimator to the simple conventional (OLS) estimator

$$\widehat{\gamma} = var_{\widehat{\mathcal{P}}}(r(\pi_{\widehat{\mathcal{P}}}(X, Z), X))^{-1} cov_{\widehat{\mathcal{P}}}(r(\pi_{\widehat{\mathcal{P}}}(X, Z), X), Y)$$

because it adjusts for the first stage propensity score estimation.

A More General MTE Model

Nonlinear Model with Interaction Effects

We derive results under a more general model for marginal treatment effects $\theta_{MTE}(X, V)$ that permits interactions between covariates X and unobserved heterogeneity V, and also nonlinear effects of V of polynomial order S. The model considered in the main text holds under $S = 1$ (linear effects) and by setting the coefficients of the interactions between X and V to zero.

We assume a linear model for these potential outcomes in terms of the observed covariates X,

$$Y_a = \alpha_a + \beta_a' X + U_a$$

for $a = 0, 1$, where α_a and β_a are unknown parameters, and U_a are unobserved errors which, without loss of generality, have mean zero conditional on X.

Following previous work (Brinch et al., 2017), we consider the following assumption on the conditional means of unobserved confounders,

$$E[U_a|X = x, V = v] = \sum_{s=1}^{S} \zeta_{as}\Big(v^s - \frac{1}{s+1}\Big) + \xi_{as}' x\Big(v^s - \frac{1}{s+1}\Big), \quad \text{for } a = 0, 1, \tag{1}$$

where $(\zeta_{01} \dots \zeta_{0S}, \zeta_{11} \dots \zeta_{1S})$ and $(\xi_{01} \dots \xi_{0S}, \xi_{11} \dots \xi_{1S})$ are unknown parameters, and the constant terms are restricted to satisfy $E_{\mathcal{P}}[U_0|X] = 0$ and $E_{\mathcal{P}}[U_1|X] = 0$.

Within Equation (1), a popular assumption sets $\xi_{01} = \dots = \xi_{0S} = 0$ and $\xi_{11} = \dots = \xi_{1S} = 0$; this is called *additive separability* (Carneiro and Lee, 2009; Brinch et al., 2017), because it will mean that the marginal treatment effect is separable in the observed covariates X and unobserved heterogeneity V. It is implied by (X, Z) being independent of (U_0, U_1, V) which, as Mogstad and Torgovitsky (2018) describe, would "nearly elevate X to the status of an instrument, albeit one that does not need to obey the usual exclusion restriction". In our application, we will take the observed covariates X to be sex, which may satisfy the additive separability assumption.

The advantage of imposing additive separability is that it allows identification of marginal treatment effects using the collective variation in $\pi_{\mathcal{P}}(X, Z)$ and X, rather than the variation in $\pi_{\mathcal{P}}(X, Z)$ conditional on X. In contrast, the more general model where all parameters may be non-zero allows *interaction effects*, so that the marginal treatment effect may vary in V in a different way across covariates X. In practice, choosing a model allowing $X \times V$ interaction effects on the MTE has implications for covariate selection. In an MTE model of interaction effects, including a large number of covariates can be demanding on the instrument-induced variation in the propensity score required for identification.

Another important choice is the order of polynomial S. Our analysis in the main text considers the $S = 1$ case because observational studies suggest a linear association between alcohol consumption and systolic blood pressure (Di Federico et al., 2023), and the individual parameters of the MTE model have an easier interpretation. As we let S increase, the underlying specifications for the conditional potential outcome means $E[Y_0|X, V]$ and $E[Y_1|X, V]$ are akin to a semiparametric model. In particular, higher order polynomials are assumed to approximate the conditional means $E[U_0|X, V]$ and $E[U_1|X, V]$, with

the usual caveat that unknown parameters in more complex models may be difficult to estimate precisely.

Estimands

Under this more general model, the conditional mean of the observed outcome $m(x,p) = E_{\mathcal{P}}[Y|X = x, \pi_{\mathcal{P}}(X,Z) = p]$ in Equation (4) of the main text now takes the form $m(x,p) = r(x,p)'\gamma$ where $r(x,p)$ is a known vector, and $\boldsymbol{\gamma}$ is an unknown vector such that

$$r(x,p) = \begin{bmatrix} 1 \\ x \\ p \\ xp \\ p^2 \\ \vdots \\ p^{S+1} \\ xp^2 \\ \vdots \\ xp^{S+1} \end{bmatrix}, \text{ and } \gamma = \begin{bmatrix} \alpha_0 \\ \beta_0 \\ \alpha_1 - \alpha_0 - \sum_{s=1}^{S} \frac{1}{s+1}(\zeta_{1s} - \zeta_{0s}) \\ \beta_1 - \beta_0 - \sum_{s=1}^{S} \frac{1}{s+1}(\xi_{1s} - \xi_{0s}) \\ \frac{1}{2}(\zeta_{11} - \zeta_{01}) \\ \vdots \\ \frac{1}{S+1}(\zeta_{1S} - \zeta_{0S}) \\ \frac{1}{2}(\xi_{11} - \xi_{01}) \\ \vdots \\ \frac{1}{S+1}(\xi_{1S} - \xi_{0S}) \end{bmatrix}.$$

The MTE function is identified as the derivative $\theta_{\text{MTE}}(x,v) = \partial m(x,p)/\partial p|_{p=v}$, and in the linear model ($S = 1$) the MTE simplifies as $\theta_{\text{MTE}}(x,v) = (\partial r(x,v)/\partial v)'\gamma$ where $\partial r(x,v)/\partial v = (0, 0', 1, x', 2p, 2px')'$. This specification of the MTE means that the treatment effect can vary linearly with covariates x and unobserved heterogeneity v, and the intercept of the linear-in-v MTE slope can vary for subgroups of covariates (in our empirical example, the MTE intercept may be different for women than for men).

Inspecting the signs of individual parameters in γ may be interesting in simple models, but is less insightful in more complex models (with a higher order polynomial S). Instead, scalar treatment effect estimands can be more attractive in terms of interpretability. We state the form of the five target estimands considered in the main text under this general model.

First, the *average treatment effect* (ATE) effect is

$$\theta_{\text{ATE}}(\gamma, \mathcal{P}) = E_{\mathcal{P}}[Y_1 - Y_0] = E_{\mathcal{P}}[r_{\text{ATE}}(X)]'\gamma \tag{2}$$

where $r_{\text{ATE}}(X) = (0, 0', 1, X', 1'_{S\times 1}, \underbrace{X', X', \ldots, X'}_{S \text{ times}})'$.

The *average treatment effect on the treated* (ATT) is

$$\theta_{\text{ATT}}(\gamma, \mathcal{P}) = E_{\mathcal{P}}[Y_1 - Y_0|A = 1] = \mathcal{P}(A = 1)^{-1} E_{\mathcal{P}}[r_{\text{ATT}}(X, Z, \mathcal{P})]'\gamma \tag{3}$$

where

$$\begin{aligned} r_{\text{ATT}}(X, Z, \mathcal{P}) \quad = \quad & (0, 0', \pi_{\mathcal{P}}(X,Z), \pi_{\mathcal{P}}(X,Z)X', \pi_{\mathcal{P}}(X,Z)^2, \ldots, \pi_{\mathcal{P}}(X,Z)^{S+1}, \\ & \pi_{\mathcal{P}}(X,Z)^2 X', \ldots, \pi_{\mathcal{P}}(X,Z)^{S+1} X')'. \end{aligned}$$

The *average treatment effect on the untreated* (ATU) is

$$\theta_{ATU}(\gamma, \mathcal{P}) = E_{\mathcal{P}}[Y_1 - Y_0 | A = 0] = \mathcal{P}(A = 0)^{-1} E_{\mathcal{P}}[r_{\text{ATU}}(X, Z, \mathcal{P})]'\gamma \tag{4}$$

where

$$\begin{aligned} r_{ATU}(X, Z, \mathcal{P}) &= (0, 0', 1 - \pi_{\mathcal{P}}(X, Z), (1 - \pi_{\mathcal{P}}(X, Z))X', 1 - \pi_{\mathcal{P}}(X, Z)^2, \ldots, 1 - \pi_{\mathcal{P}}(X, Z)^{S+1}, \\ &\quad (1 - \pi_{\mathcal{P}}(X, Z)^2)X', \ldots, (1 - \pi_{\mathcal{P}}(X, Z)^{S+1})X')'. \end{aligned}$$

The *average selection on gains* (ASG) is ,

$$\begin{aligned} \theta_{\text{ASG}}(\gamma, \mathcal{P}) &= E_{\mathcal{P}}[Y_1 - Y_0 | A = 1] - E_{\mathcal{P}}[Y_1 - Y_0 | A = 0] \\ &= (E_{\mathcal{P}}[r_{\text{ATT}}(X, Z, \mathcal{P})]\mathcal{P}(A = 1)^{-1} - E_{\mathcal{P}}[r_{\text{ATU}}(X, Z, \mathcal{P})]\mathcal{P}(A = 0)^{-1})'\gamma. \end{aligned} \tag{5}$$

The generalized LATE is

$$\theta_{\text{GLATE}}(\gamma, \mathcal{P}, \underline{\text{v}}, \bar{\text{v}}) = E_{\mathcal{P}}[Y_1 - Y_0 | \, \underline{\text{v}} < V \leq \bar{\text{v}}] = E_{\mathcal{P}}[r_{\text{GLATE}}(X, \underline{\text{v}}, \bar{\text{v}})]'\gamma \tag{6}$$

where

$$\begin{aligned} r_{\text{GLATE}}(X, \underline{\text{v}}, \bar{\text{v}}) &= (\bar{\text{v}} - \underline{\text{v}})^{-1}(0, 0', \bar{\text{v}} - \underline{\text{v}}, (\bar{\text{v}} - \underline{\text{v}})X', \bar{\text{v}}^2 - \underline{\text{v}}^2, \ldots, \bar{\text{v}}^{S+1} - \underline{\text{v}}^{S+1}, \\ &\quad (\bar{\text{v}}^2 - \underline{\text{v}}^2)X', \ldots (\bar{\text{v}}^{S+1} - \underline{\text{v}}^{S+1})X')' \end{aligned}$$

for any choices $0 \leq \underline{\text{v}} < \bar{\text{v}} < 1$.

Heckman and Vytlacil (2005) discussed how some observed estimands can also be expressed in terms of the vector of MTE parameters γ. For example, the IV estimand can be written as

$$\theta_{\text{IV}}(\gamma, \mathcal{P}) = \text{cov}_{\mathcal{P}}(A, Z)^{-1}\text{cov}_{\mathcal{P}}(Y, Z) = \text{cov}_{\mathcal{P}}(A, Z)^{-1} E_{\mathcal{P}}[r_{\text{IV}}(X, Z, \mathcal{P})]'\gamma \tag{7}$$

where $r_{\text{IV}}(X, Z, \mathcal{P}) = r(X, \pi_{\mathcal{P}}(X, Z))(Z - E_{\mathcal{P}}[Z])$.

Useful Lemmas

For the following results, $f(o)$ denotes the probability density function of O, and define $f_t(o) = t\mathcal{I}_{\widetilde{o}}(o) + (1-t)f(o)$, where $\mathcal{I}_{\widetilde{o}}(o)$ denotes the Dirac delta function with respect to $\widetilde{o}$; i.e. a function that equals zero except at $\widetilde{o}$, and that integrates to 1.

Lemma 1. *Under the mixture model $\mathcal{P}_t = t\widetilde{\mathcal{P}} + (1-t)\mathcal{P}$ where $\widetilde{\mathcal{P}}$ denotes a fixed distribution that has a point mass at a single observation $\widetilde{o} = (\tilde{a}, \tilde{x}, \tilde{z})$,*

$$\frac{\partial}{\partial t}\pi_{\mathcal{P}_t}(x,z)\Big|_{t=0} = \frac{1}{f(x,z)}\big(\mathcal{I}_{(\tilde{a},\tilde{x},\tilde{z})}(1,x,z) - \mathcal{I}_{(\tilde{x},\tilde{z})}(x,z)\pi_{\mathcal{P}}(x,z)\big).$$

Proof. For $\pi_{\mathcal{P}}(x,z) = P(A=1|X,Z) = f(1|x,z) = f(1,x,z)/f(x,z)$, we can write

$$\pi_{\mathcal{P}_t}(x,z) = \frac{f_t(1,x,z)}{f_t(x,z)}.$$

Since $f_t(1,x,z) = t\mathcal{I}_{(\tilde{a},\tilde{x},\tilde{z})}(1,x,z) + (1-t)f(1,x,z)$ and $f_t(x,z) = t\mathcal{I}_{(\tilde{x},\tilde{z})}(x,z) + (1-t)f(x,z)$, we have $\frac{\partial}{\partial t}f_t(1,x,z) = \mathcal{I}_{(\tilde{a},\tilde{x},\tilde{z})}(1,x,z) - f(1,x,z)$ and $\frac{\partial}{\partial t}f_t(x,z) = \mathcal{I}_{(\tilde{x},\tilde{z})}(x,z) - f(x,z)$. Therefore,

$$\begin{aligned}
\frac{\partial}{\partial t}\pi_{\mathcal{P}_t}(x,z)\Big|_{t=0} &= \frac{1}{f(x,z)}\cdot\frac{\partial}{\partial t}f_t(1,x,z) - \frac{f(1,x,z)}{f(x,z)^2}\cdot\frac{\partial}{\partial t}f_t(x,z) \\
&= \frac{1}{f(x,z)}\mathcal{I}_{(\tilde{a},\tilde{x},\tilde{z})}(1,x,z) - f(1|x,z) - \frac{f(1,x,z)}{f(x,z)^2}\mathcal{I}_{(\tilde{x},\tilde{z})}(x,z) + f(1|x,z) \\
&= \frac{1}{f(x,z)}\big(\mathcal{I}_{(\tilde{a},\tilde{x},\tilde{z})}(1,x,z) - \mathcal{I}_{(\tilde{x},\tilde{z})}(x,z)f(1|x,z)\big)
\end{aligned}$$

as required.

□

Lemma 2. *Under the mixture model $\mathcal{P}_t = t\widetilde{\mathcal{P}} + (1-t)\mathcal{P}$ where $\widetilde{\mathcal{P}}$ denotes a fixed distribution that has a point mass at a single observation $\widetilde{o} = (\tilde{a}, \tilde{x}, \tilde{z})$,*

$$\frac{\partial}{\partial t}r(x,\pi_{\mathcal{P}_t}(x,z))\Big|_{t=0} = \frac{\partial}{\partial \pi}r(x,\pi_{\mathcal{P}}(x,z))\cdot\frac{1}{f(x,z)}\big(\mathcal{I}_{(\tilde{a},\tilde{x},\tilde{z})}(1,x,z) - \mathcal{I}_{(\tilde{x},\tilde{z})}(x,z)\pi_{\mathcal{P}}(x,z)\big).$$

Proof. Follows immediately from the chain rule and Lemma 1.

□

Lemma 3. *Under the mixture model $\mathcal{P}_t = t\widetilde{\mathcal{P}} + (1-t)\mathcal{P}$ where $\widetilde{\mathcal{P}}$ denotes a fixed distribution that has a point mass at a single observation $\widetilde{o} = (\tilde{a}, \tilde{x}, \widetilde{y}, \tilde{z})$,*

$$\begin{aligned}
\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r(x,\pi_{\mathcal{P}_t}(x,z))y]\Big|_{t=0} &= \frac{\partial}{\partial \pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\mathcal{I}_{\tilde{a}}(1) - \pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))E_{\mathcal{P}}[y|\widetilde{x},\widetilde{z}] + r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\widetilde{y} \\
&\quad - E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}}(x,z))y].
\end{aligned}$$

Proof. Note that

$$\begin{aligned}
\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r(x,\pi_{\mathcal{P}_t}(x,z))y]\Big|_{t=0} &= \frac{\partial}{\partial t}\int r(x,\pi_{\mathcal{P}_t}(x,z))y\cdot f_t(x,y,z)\Big|_{t=0}\,dxdydz \\
&= \int \frac{\partial}{\partial t}r(x,\pi_{\mathcal{P}_t}(x,z))\Big|_{t=0} y\cdot f(x,y,z)\,dxdydz \\
&\quad + \int r(x,\pi_{\mathcal{P}}(x,z))y\cdot \frac{\partial}{\partial t}f_t(x,y,z)\Big|_{t=0}\,dxdydz \\
&= \int \frac{\partial}{\partial \pi}r(x,\pi_{\mathcal{P}}(x,z))\cdot\frac{1}{f(x,z)}\big(\mathcal{I}_{(\tilde{a},\widetilde{x},\widetilde{z})}(1,x,z)-\mathcal{I}_{(\widetilde{x},\widetilde{z})}(x,z)\pi_{\mathcal{P}}(x,z)\big)y \\
&\quad \times f(x,y,z)\,dxdydz \\
&\quad + \int r(x,\pi_{\mathcal{P}}(x,z))y\cdot\mathcal{I}_{(\widetilde{x},\widetilde{y},\widetilde{z})}(x,y,z)\,dxdydz - E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}}(x,z))y] \\
&= \int \frac{\partial}{\partial \pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\mathcal{I}_{\tilde{a}}(1)-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))y\cdot f(y|\widetilde{x},\widetilde{z})\,dy \\
&\quad + r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\widetilde{y} - E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}}(x,z))y] \\
&= \frac{\partial}{\partial \pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\mathcal{I}_{\tilde{a}}(1)-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))E_{\mathcal{P}}[y|\widetilde{x},\widetilde{z}] + r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\widetilde{y} \\
&\quad - E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}}(x,z))y]
\end{aligned}$$

since $\mathcal{I}_{\tilde{a}}(1)$, $\pi_{\mathcal{P}}(\tilde{x},\tilde{z})$, and $\frac{\partial}{\partial\pi}r(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))$ are constant, and where the third equality follows by Lemma 2.

□

Lemma 4. *Under the mixture model $\mathcal{P}_t = t\widetilde{\mathcal{P}} + (1-t)\mathcal{P}$ where $\widetilde{\mathcal{P}}$ denotes a fixed distribution that has a point mass at a single observation $\widetilde{o} = (\tilde{a},\tilde{x},\tilde{z})$,*

$$\begin{aligned}
\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r(x,\pi_{\mathcal{P}_t}(x,z))r(x,\pi_{\mathcal{P}_t}(x,z))']\Big|_{t=0} &= 2\frac{\partial}{\partial\pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\mathcal{I}_{\tilde{a}}(1)-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))' \\
&\quad + r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))' \\
&\quad - E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}}(x,z))r(x,\pi_{\mathcal{P}}(x,z))'].
\end{aligned}$$

Proof. Note that

$$\begin{aligned}
&\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r(x,\pi_{\mathcal{P}_t}(x,z))r(x,\pi_{\mathcal{P}_t}(x,z))']\Big|_{t=0} \\
&= \frac{\partial}{\partial t}\int r(x,\pi_{\mathcal{P}_t}(x,z))r(x,\pi_{\mathcal{P}_t}(x,z))'\cdot f_t(x,z)\,dxdz \\
&= 2\int \frac{\partial}{\partial\pi}r(x,\pi_{\mathcal{P}}(x,z))\cdot\frac{1}{f(x,z)}(\mathcal{I}_{\tilde{a}}(1)-\mathcal{I}_{(\widetilde{x},\widetilde{z})}(x,z)\pi_{\mathcal{P}}(x,z))r(x,\pi_{\mathcal{P}}(x,z))'\cdot f(x,z)\,dxdz \\
&+ r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))' - E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}}(x,z))r(x,\pi_{\mathcal{P}}(x,z))'] \\
&= 2\Big(\frac{\partial}{\partial\pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\Big)(\mathcal{I}_{\tilde{a}}(1)-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))' \\
&+ r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))' - E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}}(x,z))r(x,\pi_{\mathcal{P}}(x,z))']
\end{aligned}$$

where the second equality follows by Lemma 2.

□

Lemma 5. $E_{\mathcal{P}}[Y|X,Z] = E_{\mathcal{P}}[Y|X,\pi_{\mathcal{P}}(X,Z)]$.

Proof. Note that by the law of iterated expectations,

$$\begin{aligned} E_{\mathcal{P}}[Y|X,Z] &= E[AY_1 + (1-A)Y_0|X,Z] \\ &= E[AY_1|X,Z] + E[(1-A)Y_0|X,Z] \\ &= E_{A|X,Z}[E[Y_1|X,Z,A]\cdot A|X,Z] + E_{A|X,Z}[E[Y_0|X,Z,A]\cdot(1-A)|X,Z] \\ &= E[Y_1|X,Z,A=1]\cdot\mathcal{P}(A=1|X,Z) + E[Y_0|X,Z,A=0]\cdot(1-\mathcal{P}(A=1|X,Z)). \end{aligned}$$

For a mean conditional on the propensity score, we can write

$$\begin{aligned} E_{\mathcal{P}}[A|X,\pi_{\mathcal{P}}(X,Z)] &= E_{Z|X,\pi_{\mathcal{P}}(X,Z)}\big[E[A|X,\pi_{\mathcal{P}}(X,Z),Z]|X,\pi_{\mathcal{P}}(X,Z)\big] \\ &= E_{Z|X,\pi_{\mathcal{P}}(X,Z)}\big[E_{\mathcal{P}}[A|X,Z]|X,\pi_{\mathcal{P}}(X,Z)\big] \\ &= E_{Z|X,\pi_{\mathcal{P}}(X,Z)}\big[\pi_{\mathcal{P}}(X,Z)|X,\pi_{\mathcal{P}}(X,Z)\big] \\ &= \pi_{\mathcal{P}}(X,Z) \\ &= \mathcal{P}(A=1|X,Z) \end{aligned}$$

where the first equality follows by the law of iterated expectations, the second equality follows by $\pi_{\mathcal{P}}(X,Z)$ being known given knowledge of (X,Z), and the third and fourth equalities follow by the definition of the propensity score. Therefore, $\mathcal{P}(A=1|X,Z) = \mathcal{P}(A=1|X,\pi_{\mathcal{P}}(X,Z))$.

Now, by assumption, $Y_1 \perp\!\!\!\perp Z|X$. Therefore, $Y_1 \perp\!\!\!\perp (X,Z)|X$. Then, $Y_1 \perp\!\!\!\perp (X,Z)|(X,\pi_{\mathcal{P}}(X,Z))$ since $(X,\pi_{\mathcal{P}}(X,Z))$ gives no information on Y if $(X,Z)|X$ does not. Finally, $Y_1 \perp\!\!\!\perp (X,Z,A=1)|(X,\pi_{\mathcal{P}}(X,Z),A=1)$. By identical arguments, $Y_0 \perp\!\!\!\perp (X,Z,A=0)|(X,\pi_{\mathcal{P}}(X,Z),A=0)$. Hence, $E[Y_1|X,Z,A=1] = E[Y_1|X,\pi_{\mathcal{P}}(X,Z),A=1]$, and likewise, $E[Y_0|X,Z,A=0] = E[Y_0|X,\pi_{\mathcal{P}}(X,Z),A=0]$.

Overall, we have shown that

$$\begin{aligned} E_{\mathcal{P}}[Y|X,Z] &= E[Y_1|X,Z,A=1]\cdot\mathcal{P}(A=1|X,Z) + E[Y_0|X,Z,A=0]\cdot(1-\mathcal{P}(A=1|X,Z)) \\ &= E[Y_1|X,\pi_{\mathcal{P}}(X,Z),A=1]\cdot\mathcal{P}(A=1|X,\pi_{\mathcal{P}}(X,Z)) \\ &\quad + E[Y_0|X,\pi_{\mathcal{P}}(X,Z),A=0]\cdot\mathcal{P}(A=0|X,\pi_{\mathcal{P}}(X,Z)) \\ &= E_{\mathcal{P}}[Y|X,\pi_{\mathcal{P}}(X,Z)] \end{aligned}$$

as required.

□

Lemma 6. *Under the mixture model $\mathcal{P}_t = t\widetilde{\mathcal{P}} + (1-t)\mathcal{P}$ where $\widetilde{\mathcal{P}}$ denotes a fixed distribution that has a*

point mass at a single observation $\widetilde{o} = (\tilde{a}, \tilde{x}, \widetilde{y}, \tilde{z})$*, the efficient influence function for* $\boldsymbol{\gamma}(\mathcal{P})$ *is*

$$\begin{aligned}\phi(\widetilde{o}, \mathcal{P}) &= -E_{\mathcal{P}}[r(x, \pi_{\mathcal{P}}(x,z))r(x, \pi_{\mathcal{P}}(x,z))']^{-1}\frac{\partial}{\partial \pi} r(\widetilde{x}, \pi_{\mathcal{P}}(\widetilde{x}, \widetilde{z}))(\widetilde{a} - \pi_{\mathcal{P}}(\widetilde{x}, \widetilde{z}))r(\widetilde{x}, \pi_{\mathcal{P}}(\widetilde{x}, \widetilde{z}))'\boldsymbol{\gamma}(\mathcal{P}) \\ &\quad -E_{\mathcal{P}}[r(x, \pi_{\mathcal{P}}(x,z))r(x, \pi_{\mathcal{P}}(x,z))']^{-1} r(\widetilde{x}, \pi_{\mathcal{P}}(\widetilde{x}, \widetilde{z}))r(\widetilde{x}, \pi_{\mathcal{P}}(\widetilde{x}, \widetilde{z}))'\boldsymbol{\gamma}(\mathcal{P}) \\ &\quad +E_{\mathcal{P}}[r(x, \pi_{\mathcal{P}}(x,z))r(x, \pi_{\mathcal{P}}(x,z))']^{-1} r(\widetilde{x}, \pi_{\mathcal{P}}(\widetilde{x}, \widetilde{z}))\widetilde{y}.\end{aligned}$$

Proof. Note that

$$\begin{aligned}\phi(\widetilde{o}, \mathcal{P}) &= \frac{\partial}{\partial t}\gamma(\mathcal{P}_t)\Big|_{t=0} \\ &= -E_{\mathcal{P}}[r(\pi_{\mathcal{P}}(x,z), x)r(\pi_{\mathcal{P}}(x,z), x)']^{-1}\frac{\partial}{\partial t} E_{\mathcal{P}_t}[r(\pi_{\mathcal{P}_t}(x,z), x)r(\pi_{\mathcal{P}_t}(x,z), x)']\Big|_{t=0} \\ &\quad \times E_{\mathcal{P}}[r(\pi_{\mathcal{P}}(x,z), x)r(\pi_{\mathcal{P}}(x,z), x)']^{-1} E_{\mathcal{P}}[r(\pi_{\mathcal{P}}(x,z), x)y] \\ &\quad +E_{\mathcal{P}}[r(\pi_{\mathcal{P}}(x,z), x)r(\pi_{\mathcal{P}}(x,z), x)']^{-1}\frac{\partial}{\partial t} E_{\mathcal{P}_t}[r(\pi_{\mathcal{P}_t}(x,z), x)y]\Big|_{t=0}\end{aligned}$$

which leads to the required result after substituting in the results from Lemmas 3, 4 and 5.

□

Proof of Proposition 1

Using Lemma 6, setting the sample average of the efficient influence function to zero, we have

$$\begin{aligned}0 &= E_{\widehat{\mathcal{P}}}[\phi(o, \widehat{\mathcal{P}})] \\ &= -E_{\widehat{\mathcal{P}}}[r(x, \pi_{\widehat{\mathcal{P}}}(x,z))r(x, \pi_{\widehat{\mathcal{P}}}(x,z))']^{-1} E_{\widehat{\mathcal{P}}}\Big[\frac{\partial}{\partial \pi} r(x, \pi_{\widehat{\mathcal{P}}}(x,z))(a - \pi_{\widehat{\mathcal{P}}}(x,z))r(x, \pi_{\widehat{\mathcal{P}}}(x,z))'\Big]\gamma(\widehat{\mathcal{P}}) \\ &\quad -\gamma(\widehat{\mathcal{P}}) + E_{\widehat{\mathcal{P}}}[r(x, \pi_{\widehat{\mathcal{P}}}(x,z))r(x, \pi_{\widehat{\mathcal{P}}}(x,z))']^{-1} E_{\widehat{\mathcal{P}}}[r(x, \pi_{\widehat{\mathcal{P}}}(x,z))y].\end{aligned}$$

Solving for $\widetilde{\gamma} = \gamma(\widehat{\mathcal{P}})$, the efficient estimator $\widetilde{\gamma}$ has a convenient closed form expression

$$\widetilde{\gamma} = (\Omega_{\widehat{\mathcal{P}}} + \Gamma_{\widehat{\mathcal{P}}})^{-1}\Upsilon_{\widehat{\mathcal{P}}}$$

where $\Omega_{\widehat{\mathcal{P}}}$, $\Gamma_{\widehat{\mathcal{P}}}$, and $\Upsilon_{\widehat{\mathcal{P}}}$ are

$$\begin{aligned}\Omega_{\widehat{\mathcal{P}}} &= E_{\widehat{\mathcal{P}}}[r(x, \pi_{\widehat{\mathcal{P}}}(x,z))r(x, \pi_{\widehat{\mathcal{P}}}(x,z))'] \\ \Gamma_{\widehat{\mathcal{P}}} &= E_{\widehat{\mathcal{P}}}[\frac{\partial}{\partial \pi} r(x, \pi_{\widehat{\mathcal{P}}}(x,z))(a - \pi_{\widehat{\mathcal{P}}}(x,z))r(x, \pi_{\widehat{\mathcal{P}}}(x,z))'] \\ \Upsilon_{\widehat{\mathcal{P}}} &= E_{\widehat{\mathcal{P}}}[r(x, \pi_{\widehat{\mathcal{P}}}(x,z))y].\end{aligned}$$

The asymptotic variance of the efficient estimator $\widetilde{\gamma}$ can be calculated as $E_{\widehat{\mathcal{P}}}[\phi(o, \mathcal{P})\phi(o, \mathcal{P})']$. From Lemma 6, this variance expression takes the form $E_{\widehat{\mathcal{P}}}[\phi(o, \mathcal{P})\phi(o, \mathcal{P})'] = \Omega_{\mathcal{P}}^{-1}\Phi_{\mathcal{P}}\Omega_{\mathcal{P}}^{-1\prime}$ where

$$\Phi_{\mathcal{P}} = E_{\mathcal{P}}[(g(o, \gamma, \mathcal{P}) + \psi(o, \gamma, \mathcal{P}))(g(o, \gamma, \mathcal{P}) + \psi(o, \gamma, \mathcal{P}))']$$

for

$$g(o,\gamma,\mathcal{P}) = r(x,\pi_{\mathcal{P}}(x,z))(y - r(x,\pi_{\mathcal{P}}(x,z))'\gamma)$$

and

$$\psi(o,\gamma,\mathcal{P}) = \frac{\partial}{\partial\pi} r(x,\pi_{\mathcal{P}}(x,z))(a - \pi_{\mathcal{P}}(x,z))r(x,\pi_{\mathcal{P}}(x,z))'\gamma$$

and where $\Omega_{\mathcal{P}}$ and $\Gamma_{\mathcal{P}}$ are defined in the main text.

Now we need to show that the conventional (OLS) estimator $\widehat{\gamma} = \Omega_{\widehat{\mathcal{P}}}^{-1} cov_{\widehat{\mathcal{P}}}(r(\pi_{\widehat{\mathcal{P}}}(x,z),x),y)$ has the same asymptotic variance as $\widetilde{\gamma}$. To derive the asymptotic variance of $\widehat{\gamma}$, we start by deriving the so-called first stage influence function (FSIF; Chernozhukov et al. 2022), of the OLS moment condition $E_{\mathcal{P}}[g(o,\gamma,\mathcal{P})]$. where $g(o,\gamma,\mathcal{P}) = r(x,\pi_{\mathcal{P}}(x,z))(y - r(x,\pi_{\mathcal{P}}(x,z))'\gamma)$.

The FSIF term, which we denote $\text{FSIF}(o,\gamma,\mathcal{P})$, relates to the uncertainty due to first-stage propensity score estimation, and under standard regularity conditions, the first-order asymptotic expansion of the conventional estimator $\widehat{\gamma}$ is $\sqrt{n}(\widehat{\gamma} - \gamma) = \Omega_{\mathcal{P}}^{-1}\sqrt{n}E_{\widehat{\mathcal{P}}}\big[g(O,\gamma,\mathcal{P}) + \text{FSIF}(o,\gamma,\mathcal{P})\big] + o_{\mathcal{P}}(1)$.

Using the same point mass contamination strategy used above, we calculate the FSIF at observation $\widetilde{o} = (\widetilde{a},\widetilde{x},\widetilde{y},\widetilde{z})$ as the Gateaux derivative

$$\text{FSIF}(o,\gamma,\mathcal{P}) = \frac{\partial}{\partial t} E_{\mathcal{P}}[g(\widetilde{o},\gamma,\mathcal{P}_t)]\Big|_{t=0}.$$

Calculating this derivative, we have

$$\begin{aligned}
\text{FSIF}(o,\gamma,\mathcal{P}) &= \frac{\partial}{\partial t}\int r(x,\pi_{\mathcal{P}_t}(x,z))(y - r(x,\pi_{\mathcal{P}_t}(x,z))'\gamma)\cdot f(x,y,z)\,dxdydz\Big|_{t=0} \\
&= \frac{\partial}{\partial t} E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}_t}(x,z))y]\Big|_{t=0} - \Big(\frac{\partial}{\partial t} E_{\mathcal{P}}[r(x,\pi_{\mathcal{P}_t}(x,z))r(x,\pi_{\mathcal{P}_t}(x,z))']\Big|_{t=0}\Big)\gamma \\
&= \frac{\partial}{\partial\pi} r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\widetilde{a} - \pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))E_{\mathcal{P}}[y|\widetilde{x},\widetilde{z}] \\
&\quad -2\Big(\frac{\partial}{\partial\pi} r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\Big)(\tilde{a} - \pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma \\
&= -\Big(\frac{\partial}{\partial\pi} r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\Big)(\tilde{a} - \pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma \\
&= \psi(o,\gamma,\mathcal{P})
\end{aligned}$$

using Lemma 5, and where the third equality uses similar arguments used in Proofs of Lemmas 3 and 4. Thus, given $\sqrt{n}(\widehat{\gamma} - \gamma) = \Omega_{\mathcal{P}}^{-1}\sqrt{n}E_{\widehat{\mathcal{P}}}\big[g(O,\gamma,\mathcal{P}) + \text{FSIF}(o,\gamma,\mathcal{P})\big] + o_{\mathcal{P}}(1)$, we have that the asymptotic variance of $\widehat{\gamma}$ is equal to $\Omega_{\mathcal{P}}^{-1}\Phi_{\mathcal{P}}\Omega_{\mathcal{P}}^{-1\prime}$, which is the same asymptotic variance as $\widetilde{\gamma}$. □

Proof of Proposition 2

(1) IV estimand

By Table II of Mogstad et al. (2018, p.1597), we can express the IV estimand $cov(A,Z)^{-1}cov(Y,Z)$ as a weighted average of mean response functions $E[Y_1|X,V]$ and $E[Y_0|X,V]$. In our model, these mean response functions take the form $E[Y_0|X,V] = \alpha_0 + \beta_0'X + E[U_0|X,V]$ and $E[Y_1|X,V] = \alpha_1 + \beta_1'X + E[U_1|X,V]$. By Equation (1), $E[U_0|X,V] = \sum_{s=1}^{S}\zeta_{0s}\Big(V^s - \frac{1}{s+1}\Big) + \xi_{0s}'X\Big(V^s - \frac{1}{s+1}\Big)$ and $E[U_1|X,V] =$

$\sum_{s=1}^{S}\zeta_{1s}\Big(V^s-\frac{1}{s+1}\Big)+\xi_{1s}'X\Big(V^s-\frac{1}{s+1}\Big)$. Hence, using the formula given in Table II of Mogstad et al. (2018, p.1597), we can write

$$\begin{aligned}
cov(Y,Z) &= E\Big[\int_p^1\Big\{\alpha_0+\beta_0'X+\sum_{s=1}^{S}\zeta_{0s}\Big(v^s-\frac{1}{s+1}\Big)+\xi_{0s}'X\Big(v^s-\frac{1}{s+1}\Big)\Big\}dv\cdot(Z-E[Z])\Big]\\
&\quad +E\Big[\int_0^p\Big\{\alpha_1+\beta_1'X+\sum_{s=1}^{S}\zeta_{1s}\Big(v^s-\frac{1}{s+1}\Big)+\xi_{1s}'X\Big(v^s-\frac{1}{s+1}\Big)\Big\}dv\cdot(Z-E[Z])\Big]\\
&= \alpha_0E[Z-E[Z]]+\beta_0'E[X(Z-E[Z])]+\Big(\alpha_1-\alpha_0-\sum_{s=1}^{S}\frac{1}{s+1}(\zeta_{1s}-\zeta_{0s})\Big)E[p(Z-E[Z])]\\
&\quad +\Big(\beta_1-\beta_0-\sum_{s=1}^{S}\frac{1}{s+1}(\xi_{1s}-\xi_{0s})\Big)'E[Xp(Z-E[Z])]\\
&\quad +\sum_{s=1}^{S}\frac{1}{s+1}(\zeta_{1s}-\zeta_{0s})E[p^{s+1}(Z-E[Z])]+\sum_{s=1}^{S}\frac{1}{s+1}(\xi_{1s}-\xi_{0s})E[Xp^{s+1}(Z-E[Z])].
\end{aligned}$$

Therefore, $\theta_{IV}(\mathcal{P})=cov_{\mathcal{P}}(A,Z)^{-1}cov_{\mathcal{P}}(Y,Z)=cov_{\mathcal{P}}(A,Z)^{-1}E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\gamma(\mathcal{P})$.

The efficient influence function of $\theta_{IV}(\mathcal{P})$ can be calculated by perturbing $\theta_{IV}(\mathcal{P})$ in the direction $\widetilde{\mathcal{P}}$ using the mixture model $\mathcal{P}_t=t\widetilde{\mathcal{P}}+(1-t)\mathcal{P}$, where $\widetilde{\mathcal{P}}$ denotes a fixed distribution that has a point mass at a single observation $\tilde{o}=(\tilde{y},\tilde{a},\tilde{x},\tilde{z})$. The efficient influence function at $\tilde{o}$ is

$$\begin{aligned}
\phi_{IV}(\tilde{o},\mathcal{P}) &= \frac{\partial}{\partial t}\theta_{IV}(\mathcal{P}_t)\Big|_{t=0}\\
&= \Big(\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]\Big|_{t=0}\Big)'\gamma(\mathcal{P})\cdot cov_{\mathcal{P}}(A,Z)^{-1}\\
&\quad +E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\cdot\frac{\partial}{\partial t}\gamma(\mathcal{P}_t)\Big|_{t=0}\cdot cov_{\mathcal{P}}(A,Z)^{-1}\\
&\quad +E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\gamma(\mathcal{P})\cdot\frac{\partial}{\partial t}\big(cov_{\mathcal{P}_t}(A,Z)^{-1}\big)\Big|_{t=0}. \qquad (8)
\end{aligned}$$

We will deal with each of the three terms on the RHS of Equation (8) in turn. First,

$$\begin{aligned}
\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]\Big|_{t=0} &= r(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{z}-E_{\mathcal{P}}[Z])-E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]\\
&\quad +E_{\mathcal{P}}\Big[\frac{\partial}{\partial t}r(X,\pi_{\mathcal{P}_t}(X,Z))(Z-E_{\mathcal{P}_t}[Z])\Big|_{t=0}\Big].
\end{aligned}$$

Now,

$$\begin{aligned}
E_{\mathcal{P}}\Big[\frac{\partial}{\partial t}r(X,\pi_{\mathcal{P}_t}(X,Z))(Z-E_{\mathcal{P}_t}[Z])\Big|_{t=0}\Big] &= \int\frac{\partial}{\partial t}r(x,\pi_{\mathcal{P}_t}(x,z))(z-E_{\mathcal{P}_t}[Z])\Big|_{t=0}\cdot f(x,z)\,dxdz\\
&= \int\frac{\partial}{\partial\pi}r(x,\pi_{\mathcal{P}}(x,z))(\mathcal{I}_{\tilde{a}}(1)-\pi_{\mathcal{P}}(x,z))\mathcal{I}_{(\tilde{x},\tilde{z})}(x,z)(z-E_{\mathcal{P}}[Z])\,dxdz\\
&\quad -\int r(x,\pi_{\mathcal{P}}(x,z))(\tilde{z}-E_{\mathcal{P}}[Z])f(x,z)\,dxdz\\
&= \frac{\partial}{\partial\pi}r(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{a}-\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{z}-E_{\mathcal{P}}[Z])\\
&\quad -E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))](\tilde{z}-E_{\mathcal{P}}[Z]),
\end{aligned}$$

where the second equality follows by using Lemma 2, and noting that $\partial E_{\mathcal{P}_t}[Z]/\partial t = \int z(\mathcal{I}_{\tilde{z}}(z) - f(z))dz = \tilde{z} - E_{\mathcal{P}}[Z]$.

Overall,

$$\begin{aligned}\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]\Big|_{t=0} &= r(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{z}-E_{\mathcal{P}}[Z]) - E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])] \\ &\quad + \frac{\partial}{\partial \pi}r(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{a}-\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{z}-E_{\mathcal{P}}[Z]) \\ &\quad - E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))](\tilde{z}-E_{\mathcal{P}}[Z]).\end{aligned}$$

For the second term on the RHS of Equation (8), Lemma 6 tells us

$$\begin{aligned}\frac{\partial}{\partial t}\gamma(\mathcal{P}_t)\Big|_{t=0} &= -\Omega_{\mathcal{P}}^{-1}\frac{\partial}{\partial \pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\widetilde{a}-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P}) \\ &\quad -\Omega_{\mathcal{P}}^{-1}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P}) + \Omega_{\mathcal{P}}^{-1}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\widetilde{y}.\end{aligned}$$

Therefore, the second term on the RHS of Equation (8) is equal to

$$\begin{aligned}&cov_{\mathcal{P}}(A,Z)^{-1}E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\Omega_{\mathcal{P}}^{-1}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\big(\widetilde{y}-r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P})\big) \\ -&cov_{\mathcal{P}}(A,Z)^{-1}E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\Omega_{\mathcal{P}}^{-1}\frac{\partial}{\partial \pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\widetilde{a}-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P}).\end{aligned}$$

For the third and final term on the RHS of Equation (8), note that

$$\begin{aligned}\frac{\partial}{\partial t}E_{\mathcal{P}_t}[(A-E_{\mathcal{P}_t}[A])(Z-E_{\mathcal{P}_t}[Z])]\Big|_{t=0} &= (\mathcal{I}_{\tilde{a}}(1)-E_{\mathcal{P}}[A])(\tilde{z}-E_{\mathcal{P}}[Z]) - cov_{\mathcal{P}}(A,Z) \\ &\quad + E_{\mathcal{P}}\Big[\frac{\partial}{\partial t}(A-E_{\mathcal{P}_t}[A])(Z-E_{\mathcal{P}_t}[Z])\Big|_{t=0}\Big] \\ &= (\mathcal{I}_{\tilde{a}}(1)-E_{\mathcal{P}}[A])(\tilde{z}-E_{\mathcal{P}}[Z]) - cov_{\mathcal{P}}(A,Z).\end{aligned}$$

Hence, the third term on the RHS of Equation (8) is

$$\begin{aligned}&-E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\gamma(\mathcal{P})\cdot cov_{\mathcal{P}}(A,Z)^{-2}\Big\{(\mathcal{I}_{\tilde{a}}(1)-E_{\mathcal{P}}[A])(\tilde{z}-E_{\mathcal{P}}[Z]) - cov_{\mathcal{P}}(A,Z)\Big\} \\ &= -E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\gamma(\mathcal{P})\cdot cov_{\mathcal{P}}(A,Z)^{-2}(\mathcal{I}_{\tilde{a}}(1)-E_{\mathcal{P}}[A])(\tilde{z}-E_{\mathcal{P}}[Z]) + \theta_{IV}(\mathcal{P})\end{aligned}$$

Combining the above results, we calculate the efficient influence function at $\tilde{o}$ as

$$\begin{aligned}\phi_{IV}(\tilde{o},\mathcal{P}) &= cov_{\mathcal{P}}(A,Z)^{-1}\big(r(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{z}-E_{\mathcal{P}}[Z])\big)'\gamma(\mathcal{P}) - \theta_{IV}(\mathcal{P}) \\ &\quad + cov_{\mathcal{P}}(A,Z)^{-1}\Big(\frac{\partial}{\partial \pi}r(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{a}-\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\tilde{z}-E_{\mathcal{P}}[Z])\Big)'\gamma(\mathcal{P}) \\ &\quad - cov_{\mathcal{P}}(A,Z)^{-1}(\tilde{z}-E_{\mathcal{P}}[Z])E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))]'\gamma(\mathcal{P}) \\ &\quad + cov_{\mathcal{P}}(A,Z)^{-1}E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\Omega_{\mathcal{P}}^{-1}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\big(\widetilde{y}-r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\boldsymbol{\gamma}(\mathcal{P})\big) \\ &\quad - cov_{\mathcal{P}}(A,Z)^{-1}E_{\mathcal{P}}[r(X,\pi_{\mathcal{P}}(X,Z))(Z-E_{\mathcal{P}}[Z])]'\Omega_{\mathcal{P}}^{-1}\frac{\partial}{\partial \pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\widetilde{a}-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\boldsymbol{\gamma}(\mathcal{P}) \\ &\quad - cov_{\mathcal{P}}(A,Z)^{-1}(\mathcal{I}_{\tilde{a}}(1)-E_{\mathcal{P}}[A])(\tilde{z}-E_{\mathcal{P}}[Z])\cdot\theta_{IV}(\mathcal{P}) + \theta_{IV}(\mathcal{P}).\end{aligned}$$

Setting the empirical mean of $\phi_{IV}(O,\mathcal{P})$ to zero, we find an efficient estimator as

$$\begin{aligned}\theta_{IV}(\widehat{\mathcal{P}}) &= cov_{\widehat{\mathcal{P}}}(A,Z)^{-1}E_{\widehat{\mathcal{P}}}[r(X,\pi_{\widehat{\mathcal{P}}}(X,Z))(Z-E_{\widehat{\mathcal{P}}}[Z])]'\widehat{\gamma}\\ &\quad +cov_{\widehat{\mathcal{P}}}(A,Z)^{-1}E_{\widehat{\mathcal{P}}}\Big[\frac{\partial}{\partial\pi}r(X,\pi_{\widehat{\mathcal{P}}}(X,Z))(A-\pi_{\widehat{\mathcal{P}}}(X,Z))(Z-E_{\widehat{\mathcal{P}}}[Z])\Big]'\widehat{\gamma}\\ &\quad -cov_{\widehat{\mathcal{P}}}(A,Z)^{-1}E_{\widehat{\mathcal{P}}}[r(X,\pi_{\widehat{\mathcal{P}}}(X,Z))(Z-E_{\widehat{\mathcal{P}}}[Z])]'\Omega_{\widehat{\mathcal{P}}}^{-1}\Gamma_{\widehat{\mathcal{P}}}\widehat{\gamma}\end{aligned}$$

which is the expression for the efficient IV estimand relating to Equation (7).

(II) ATE ESTIMAND

Analogously to the derivation of the IV estimand above, this time we substitute the MTE expression from our model into ATE expression in Table I of Mogstad et al. (2018, p.1595) to get $\theta_{ATE}(\mathcal{P}) = E_{\mathcal{P}}[r_{ATE}(X)]'\gamma(\mathcal{P})$ as defined in Equation (2).

We find the efficient influence function of $\theta_{ATE}(\mathcal{P})$ as

$$\begin{aligned}\frac{\partial}{\partial t}\theta_{ATE}(\mathcal{P}_t)\Big|_{t=0} &= \Big(\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r_{ATE}(X)]\Big|_{t=0}\Big)'\gamma(\mathcal{P}) + E_{\mathcal{P}}[r_{ATE}(X)]'\frac{\partial}{\partial t}\gamma(\mathcal{P}_t)\Big|_{t=0}\\ &= (r_{ATE}(\tilde{x})-E_{\mathcal{P}}[r_{ATE}(X)])'\gamma(\mathcal{P}) + E_{\mathcal{P}}[r_{ATE}(X)]'\Omega_{\mathcal{P}}^{-1}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\big(\tilde{y}-r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P})\big)\\ &\quad -E_{\mathcal{P}}[r_{ATE}(X)]'\Omega_{\mathcal{P}}^{-1}\frac{\partial}{\partial\pi}r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\widetilde{a}-\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P}).\end{aligned}$$

Setting the empirical mean of the efficient influence function to zero, we find an efficient estimator as

$$\theta_{ATE}(\widehat{\mathcal{P}}) = E_{\widehat{\mathcal{P}}}[r_{ATE}(X)]'\widehat{\gamma} - E_{\widehat{\mathcal{P}}}[r_{ATE}(X)]'\Omega_{\widehat{\mathcal{P}}}^{-1}\Gamma_{\widehat{\mathcal{P}}}\widehat{\gamma}.$$

(III) ATT ESTIMAND

Again we substitute the MTE expression from our model into the ATT expression in Table II of Mogstad et al. (2018, p.1595) to get $\theta_{ATT}(\mathcal{P}) = \mathcal{P}(A=1)^{-1}E_{\mathcal{P}}[r_{ATT}(X,Z,\mathcal{P})]'\gamma(\mathcal{P})$ as defined in Equation (3).

We find the efficient influence function of $\theta_{ATT}(\mathcal{P})$ as

$$\begin{aligned}\frac{\partial}{\partial t}\theta_{ATT}(\mathcal{P}_t)\Big|_{t=0} &= \frac{\partial}{\partial t}\mathcal{P}_t(A=1)^{-1}\Big|_{t=0}\cdot E_{\mathcal{P}}[r_{ATT}(X,Z,\mathcal{P})]'\gamma(\mathcal{P}) + \mathcal{P}(A=1)^{-1}\Big(\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r_{ATT}(X,Z,\mathcal{P}_t)]\Big|_{t=0}\Big)'\gamma(\mathcal{P})\\ &\quad +\mathcal{P}(A=1)^{-1}E_{\mathcal{P}}[r_{ATT}(X,Z,\mathcal{P})]'\frac{\partial}{\partial t}\gamma(\mathcal{P}_t)\Big|_{t=0}.\end{aligned}$$

We deal with each of the three terms on the RHS of this expression in turn. For the first term, note that

$$\frac{\partial}{\partial t}\mathcal{P}_t(A=1)^{-1}\Big|_{t=0} = -P(A=1)^{-2}(\mathcal{I}_{\tilde{a}}(1)-P(A=1))$$

so that

$$\frac{\partial}{\partial t}\mathcal{P}_t(A=1)^{-1}\Big|_{t=0}\cdot E_{\mathcal{P}}[r_{ATT}(X,Z,\mathcal{P})]'\gamma(\mathcal{P}) = \theta_{ATT}(\mathcal{P}) - \theta_{ATT}(\mathcal{P})P(A=1)^{-1}\mathcal{I}_{\tilde{a}}(1).$$

Next, for the second term,

$$\begin{aligned}\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r_{ATT}(X,Z,\mathcal{P}_t)]\Big|_{t=0} &= r_{ATT}(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z})) - E_{\mathcal{P}}[r_{ATT}(X,\pi_{\mathcal{P}}(X,Z))] \\ &\quad + \frac{\partial}{\partial \pi} r_{ATT}(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\mathcal{I}_{\tilde{a}}(1) - \pi_{\mathcal{P}}(\tilde{x},\tilde{z}))\end{aligned}$$

by Lemma 1. Hence,

$$\begin{aligned}\mathcal{P}(A=1)^{-1}\Big(\frac{\partial}{\partial t}E_{\mathcal{P}_t}[r_{ATT}(X,Z,\mathcal{P}_t)]\Big|_{t=0}\Big)'\gamma(\mathcal{P}) &= \mathcal{P}(A=1)^{-1} r_{ATT}(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))'\gamma(\mathcal{P}) - \theta_{ATT}(\mathcal{P}) \\ &\quad + \mathcal{P}(A=1)^{-1}\Big(\frac{\partial}{\partial \pi} r_{ATT}(\tilde{x},\pi_{\mathcal{P}}(\tilde{x},\tilde{z}))(\mathcal{I}_{\tilde{a}}(1) - \pi_{\mathcal{P}}(\tilde{x},\tilde{z}))\Big)'\gamma(\mathcal{P}).\end{aligned}$$

Using Lemma 6, the third term is equal to

$$\begin{aligned}&\mathcal{P}(A=1)^{-1}E_{\mathcal{P}}[r_{ATT}(X,Z,\mathcal{P})]'\Omega_{\mathcal{P}}^{-1} r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))\big(\widetilde{y} - r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P})\big) \\ -\mathcal{P}(A=1)^{-1}&E_{\mathcal{P}}[r_{ATT}(X,Z,\mathcal{P})]'\Omega_{\mathcal{P}}^{-1}\frac{\partial}{\partial \pi} r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))(\widetilde{a} - \pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))r(\widetilde{x},\pi_{\mathcal{P}}(\widetilde{x},\widetilde{z}))'\gamma(\mathcal{P}).\end{aligned}$$

Combining the above results and setting the mean of the efficient influence function to zero, we find an efficient estimator as

$$\begin{aligned}\theta_{ATT}(\widehat{\mathcal{P}}) &= \widehat{\mathcal{P}}(A=1)^{-1}E_{\widehat{\mathcal{P}}}[r_{ATT}(X,\pi_{\widehat{\mathcal{P}}}(X,Z))]'\widehat{\gamma} \\ &\quad + \widehat{\mathcal{P}}(A=1)^{-1}E_{\widehat{\mathcal{P}}}\Big[\frac{\partial}{\partial \pi} r_{ATT}(X,\pi_{\widehat{\mathcal{P}}}(X,Z))(A - \pi_{\widehat{\mathcal{P}}}(X,Z))\Big]'\widehat{\gamma} \\ &\quad - \widehat{\mathcal{P}}(A=1)^{-1}E_{\widehat{\mathcal{P}}}[r_{ATT}(X,Z,\widehat{\mathcal{P}})]'\Omega_{\widehat{\mathcal{P}}}^{-1}\Gamma_{\widehat{\mathcal{P}}}\widehat{\gamma}.\end{aligned}$$

(iv) ATU estimand

The efficient ATU estimator can be derived using very similar arguments as used for the efficient ATT estimator above.

(v) ASG estimand

Follows immediately from the expressions of the ATT and ATU estimands.

(vi) Generalized LATE estimand

The structure of the efficient estimator is very similar to the efficient ATE estimator from Part (ii), and its derivation follows nearly identical arguments. □

Sensitivity Analysis: Characterizing The Bias

As discussed in Section 3.3 of the main text, our potential outcomes model that allows a direct instrument effect is

$$Y_a = \alpha_a + \beta_a' X + \tau_i Z + U_a$$

for $a = 0, 1$, with $E[\tau_i|X, Z] = \tau$ and $E[U_a|X, V] = E[U_a|V] = \zeta_a(V - 0.5)$ under an assumption that the MTE is additively separable in its X and V effects. This is the Section 3.3 model with the V-dependence collected into U_a: writing $U_a = \zeta_a(V - 0.5) + \epsilon_a$ with $E[\epsilon_a|X, V, Z] = 0$ recovers the main-text form $Y_a = (\alpha_a - 0.5\zeta_a) + \beta_a' X + \tau_i Z + \zeta_a V + \epsilon_a$, and $E[Y_a|X, V] = (\alpha_a - 0.5\zeta_a) + \beta_a' X + \zeta_a V$ as in Equation (3). The exogeneity condition now required is that Z is conditionally independent of (U_0, U_1, V) given X, which is the same as Equation (2) but with the individual direct instrument variation $\tau_i Z$ partialled out from the potential outcomes.

Since τ_i is common to both potential outcomes, it cancels in the treatment gain $Y_1 - Y_0$ and enters the observed mean $m(x, p) = E[Y|X = x, \pi(X, Z) = p]$ only through the level term $E[\tau_i|X = x, Z = z(x, p)]\, z(x, p) = \tau z(x, p)$. Individual heterogeneity in τ_i, including any dependence on V, therefore enters solely through its average τ. Using identical arguments to Brinch, Mogstad, and Wiswall (2017, p.991-3), we obtain an analogous expression to their Equation (8) for the observed outcome mean $m(x, p)$,

$$m(x, p) = \alpha_0 + \beta_0 x + \tau z(x, p) + p\big(\alpha_1 - \alpha_0 + (\beta_1 - \beta_0)' x\big) + K(x, p)$$

where $z(x, p)$ solves $\pi(x, z) = p$, and $K(x, p) = \int_0^p E[U_1 - U_0|X = x, V = v]\, \mathrm{d}v$.

Recall that $V|X = x$ is a uniform random variable, and our model is $E[U_0|X = x, V = v] = \zeta_0(v - 0.5)$ and $E[U_1|X = x, V = v] = \zeta_1(v - 0.5)$. Hence, $K(x, p) = 0.5(\zeta_1 - \zeta_0)p^2 - 0.5(\zeta_1 - \zeta_0)p$.

Therefore, our expression of $m(x, p)$ simplifies to

$$\begin{aligned} m(x, p) &= \alpha_0 + \beta_0 x + \tau z(x, p) + \big(\alpha_1 - \alpha_0 - 0.5(\zeta_1 - \zeta_0)\big)p + (\beta_1 - \beta_0)' x p \\ &\quad + 0.5(\zeta_1 - \zeta_0)p^2 \\ &= r(x, p)'\gamma + \tau \cdot z(x, p) \end{aligned}$$

where $r(x, p)$ and γ are defined in Equation (4) of the main text.

Thus, if we continue to estimate the MTE using Heckman and Vytlacil (1999)'s approach, i.e., by differentiating $m(x, p)$ with respect to p, we no longer identify $\mathrm{MTE}(x, p)$ since we have an additional bias term compared with Equation (4),

$$\begin{aligned} \frac{\partial}{\partial p} m(x, p) &= \tau \frac{\partial z(x, p)}{\partial p} + \alpha_1 - \alpha_0 - 0.5(\zeta_1 - \zeta_0) + (\beta_1 - \beta_0)' x + (\zeta_1 - \zeta_0)p \\ &= \tau \frac{\partial z(x, p)}{\partial p} + \mathrm{MTE}(x, p) \\ &= \underbrace{\Big(\partial \pi(\boldsymbol{x}, z)/\partial z\Big|_{z = z(\boldsymbol{x}, p)}\Big)^{-1} \tau}_{\text{bias}} + \mathrm{MTE}(x, p) \end{aligned}$$

where the last equality follows by the implicit function theorem.

If τ is known, since $m(x,p) = r(x,p)'\gamma + \tau \cdot z(x,p)$ we can identify the MTE parameters γ (and hence the target estimands) by subtracting τZ from the observed outcome Y, and regressing $Y - \tau Z$ on the vector of regressors $r(x,p)$. Because we subtract the average τZ rather than the individual $\tau_i Z$, the residual $(\tau_i - \tau)Z$ remains; it has conditional mean zero given (X, Z), and so inflates the sampling variance of the estimator but does not affect its consistency. This forms the basis for our sensitivity analysis.

Additional Simulation Results: Bias Performance and Bandwidth Choice

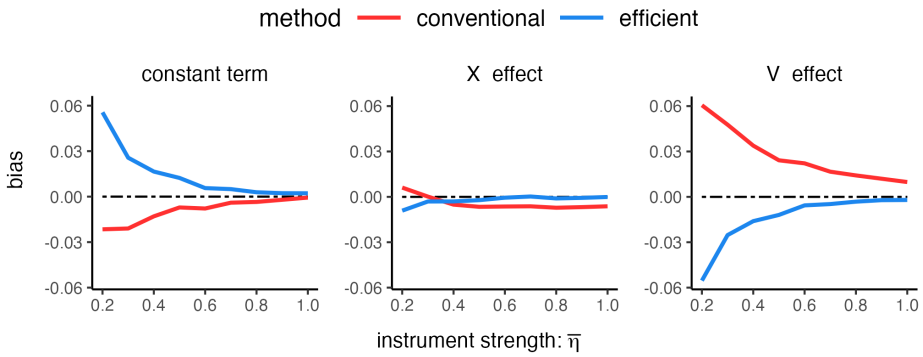

Figure S1. Average bias of MTE parameter estimates over 1000 simulation experiments for varying instrument strength.

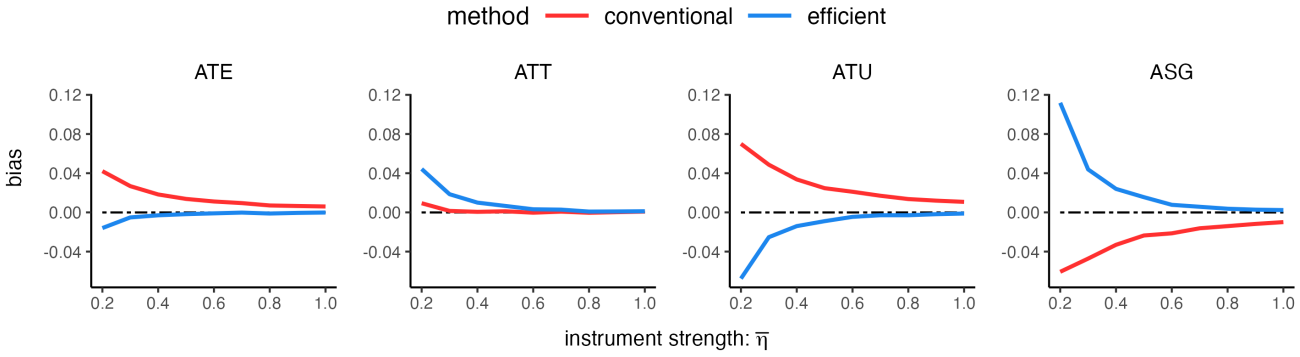

Figure S2. Average bias of target parameter estimates over 1000 simulation experiments for varying instrument strength.

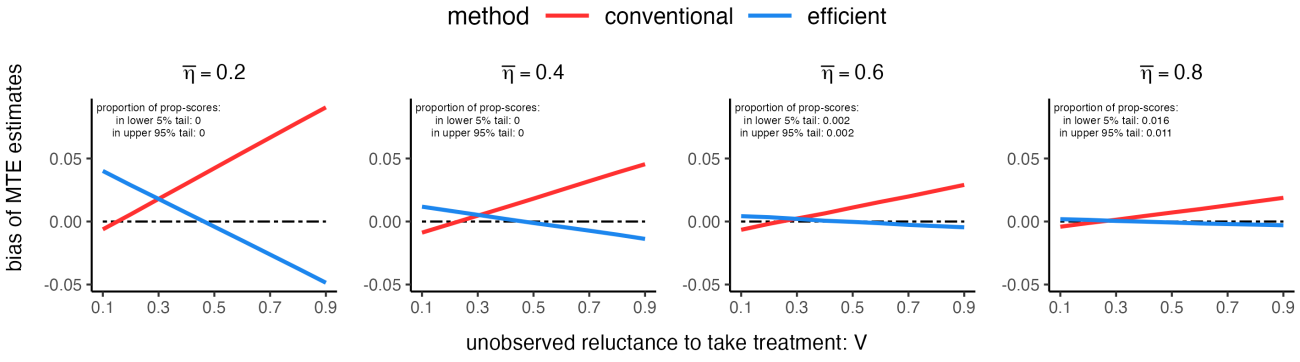

Figure S3. Average bias of the generalized LATE estimand point estimates over 1000 simulation experiments at a moving average of the last 10-percentile points of V, where, for example, at the 90-th percentile, the true generalized LATE was $\theta_{\text{GLATE}}(\underline{v},\bar{v}) = E[Y_1 - Y_0|0.8 < V \leq 0.9]$.

In Figures S4 and S5 below, we considered the sensitivity of estimates to the choice of bandwidth. Let h_0 be the averaged bandwidth using cross-validation on 3 random samples of size $1,000$ as discussed in Section 4.2 of the main text. We present the bias performance of the conventional and efficient estimates as the selected bandwidth $h = \max\{0.005, h_0 + 0.5\kappa\}$ moved away from h_0 by varying the parameter choice κ away from zero.

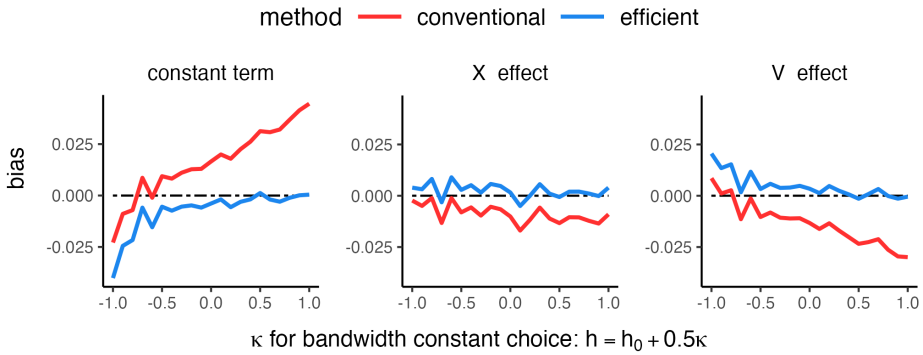

Figure S4. Average bias of MTE parameter estimates over 1000 simulation experiments for bandwidth constant choice h.

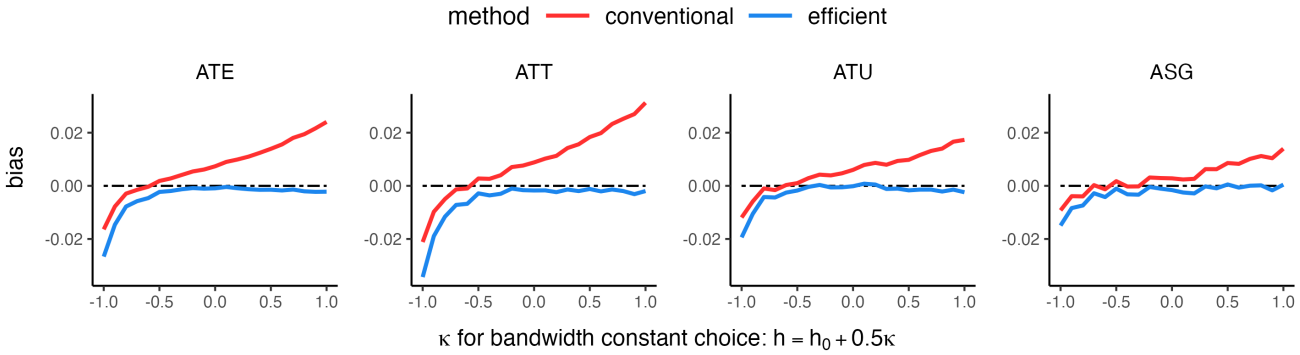

Figure S5. Average bias of target estimand point estimates over 1000 simulation experiments for bandwidth constant choice h.

Additional Empirical Results: MTE Model with Interaction Effects

	Model: MTE with an interaction $X \times V$ effect			
MTE Parameter	constant term	woman (X)	health consciousness (V)	interaction effect ($X \times V$)
Estimate	2.454	-1.797	-3.032	2.001
Standard Error	0.896	1.123	1.329	2.266
95% CI-Lower	0.697	-3.998	-5.638	-2.440
95% CI-Upper	4.211	0.404	-0.427	6.441

Table S1. Efficient estimates of MTE parameters in a model with interaction effects, $\theta_{\text{MTE}}(X, V) \sim \text{constant term} + X_{\text{woman}} + V_{\text{health-consciousness}} + (X_{\text{woman}} \times V_{\text{health-consciousness}})$.

	Model: MTE with an interaction $X \times V$ effect			
Estimand	ATE	ATT	ATU	ASG
Estimate	0.525	1.044	-0.073	1.117
Standard Error	0.202	0.351	0.482	0.730
95% CI-Lower	0.128	0.356	-1.018	-0.314
95% CI-Upper	0.921	1.731	0.871	2.549

Table S2. Efficient estimates of target estimands in a model with interaction effects, $\theta_{\text{MTE}}(X, V) \sim \text{constant term} + X_{\text{woman}} + V_{\text{health-consciousness}} + (X_{\text{woman}} \times V_{\text{health-consciousness}})$.

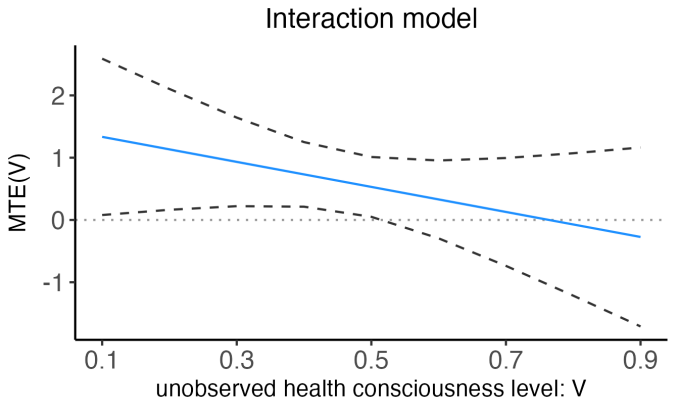

Figure S6. Efficient MTE estimates varying with V and averaged over covariates, $E[Y_1 - Y_0|V]$ (solid blue line) in an MTE model allowing $X \times V$ interactions. The dashed black curves indicate 95% asymptotic confidence intervals.

Additional Empirical Results: Parametric Propensity Score Model

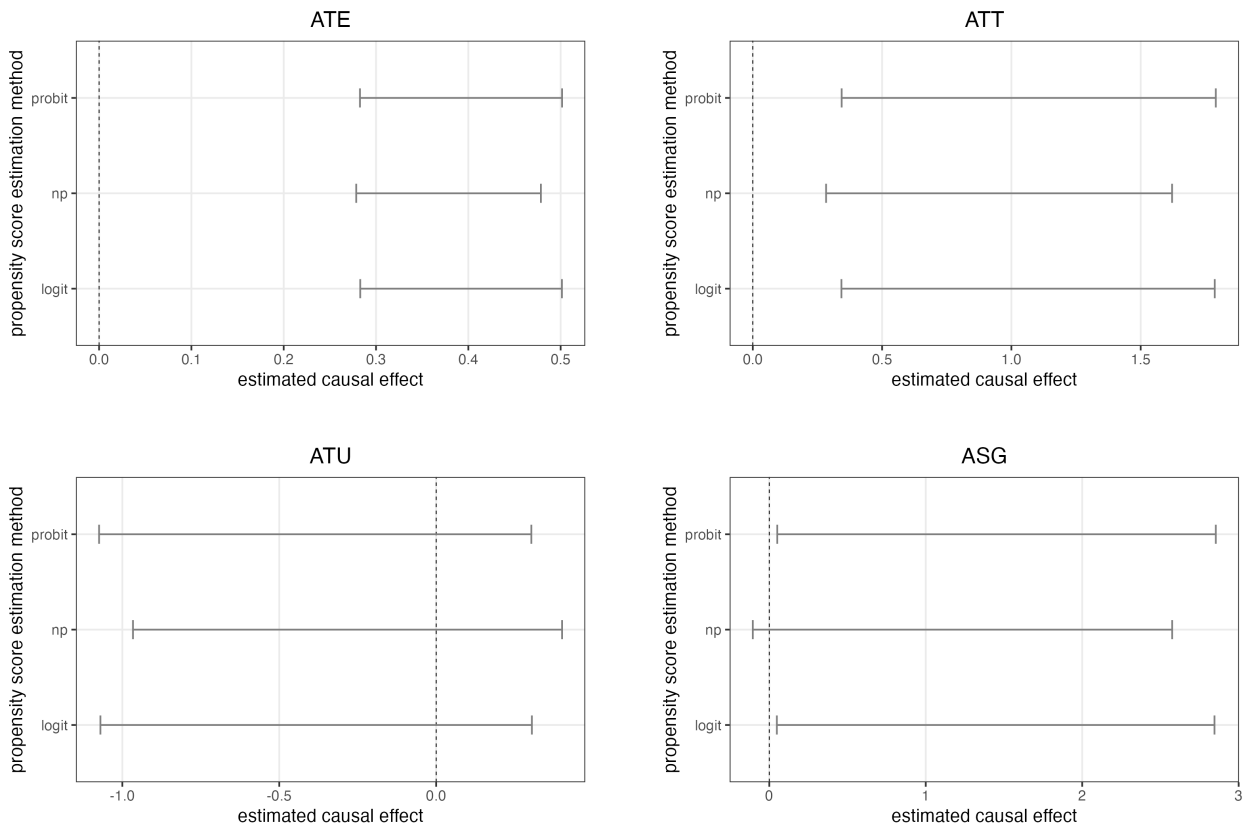

Figure S7. 95% confidence intervals of target estimands estimated under a nonparametric propensity score, and under parametrically (logit and probit specifications) estimated propensity scores. The dashed black curves indicate 95% asymptotic confidence intervals.

References

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond LATE with a discrete instrument. *Journal of Political Economy 125*(4), 985–1039.

Carneiro, P. and S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*(2), 191–208.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2022). Locally robust semiparametric estimation. *Econometrica 90*(4), 1501–1535.

Di Federico, S., T. Filippini, P. K. Whelton, M. Cecchini, I. Iamandii, G. Boriani, and M. Vinceti (2023). Alcohol intake and blood pressure levels: a dose-response meta-analysis of nonexperimental cohort studies. *Hypertension 80*(10), 1961–1969.

Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variables and latent variable models for

identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences of the United States of America 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Hines, O., O. Dukes, K. Diaz-Ordaz, and S. Vansteelandt (2022). Demystifying statistical learning based on efficient influence functions. *American Statistician 76*(3), 292–304.

Ichimura, H. and W. K. Newey (2022). The influence function of semiparametric estimators. *Quantitative Economics 13*(1), 29–61.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica 86*(5), 1589–1619.

Mogstad, M. and A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics 10*, 577–613.