

Supplementary Material

“Mediation Analysis with Mendelian Randomization and Efficient Multiple GWAS Integration”

Rita Qiuran Lyu ^{*} Chong Wu[†] Xinwei Ma[‡] Jingshen Wang[§]

May 15, 2024

Contents

1	Causal Diagram and Instrumental Variable Selection	3
2	Convergence of the Inverse Matrix	6
3	Large-sample Distribution and Statistical Inference	8
4	Asymptotic Efficiency Comparison with MVMR	11
5	Plug-in Estimation and Imperfect IV Selection Bias	12
6	Proofs	15
6.1	Proof of Lemma 1	15
6.2	Proof of Lemma 2	16
6.3	Proof of Lemma 3	16
6.4	Proof of Lemma 4	16
6.5	Proof of Lemma 5	18
6.6	Derivation of Equations (1) and (2)	20
6.7	Proof of Lemma 6	21
6.8	Proof of Lemma 7	22
6.9	Proof of Lemma 8	22
6.10	Derivation of Equation (3)	22
6.11	Proof of Lemma 9	23

Correspondence: Xinwei Ma (x1ma@ucsd.edu), Jingshen Wang (jingshenwang@berkeley.edu).

^{*}Division of Biostatistics, University of California Berkeley.

[†]Department of Biostatistics, The University of Texas MD Anderson Cancer Center.

[‡]Department of Economics, University of California San Diego.

[§]Division of Biostatistics, University of California Berkeley.

6.12	Proof of Theorem 1	25
6.13	Proof of Theorem 2	25
6.14	Proof of Theorem 3	28
7	Simulation	31
7.1	Overlapping Index Set Construction	31
7.2	Index Construction of $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$ under Oracle Setting	31
7.3	Efficiency Gain	32
7.3.1	Simulation Results under $\tau_X = 0$	32
7.3.2	Additional Simulation Results	36
8	Real Data Analysis	37
8.1	Data Harmonization	37
8.2	σ -based LD Clumping Results	38
8.3	LD Clumping Results	41
8.4	LD Pruning Results	44
8.5	Data Source	45

1 Causal Diagram and Instrumental Variable Selection

We start with summary data from three GWAS:

- Exposure dataset: GWAS (I): $\{(\hat{\beta}_{X_j}, \sigma_{X_j}^2)\}_{j=1}^p$,
- Outcome dataset: GWAS (II): $\{(\hat{\beta}_{Y_j}, \sigma_{Y_j}^2)\}_{j=1}^p$,
- Mediator dataset: GWAS (III): $\{(\hat{\beta}_{M_j}, \sigma_{M_j}^2)\}_{j=1}^p$,

where $(\hat{\beta}_{X_j}, \hat{\beta}_{Y_j}, \hat{\beta}_{M_j})$ are estimates, $(\sigma_{X_j}, \sigma_{Y_j}, \sigma_{M_j})$ are corresponding standard errors.

The following assumption links SNP-exposure, SNP-mediator and SNP-outcome associations to the causal and mediation effects in diagram 1.

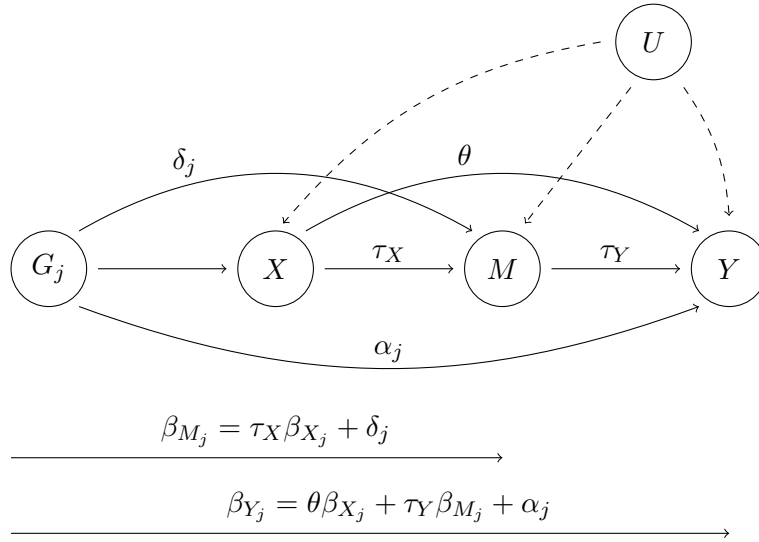


Figure 1: Causal diagram for mediation analysis with multiple GWAS.

Assumption 1 (Causal and mediation effects)

It holds for all j that

$$\beta_{Y_j} = \underbrace{\beta_{X_j} \cdot \theta}_{G_j \rightarrow X \rightarrow Y} + \underbrace{\beta_{M_j} \cdot \tau_Y}_{G_j \rightarrow M \rightarrow Y} + \underbrace{\alpha_j}_{G_j \rightarrow Y}, \quad \beta_{M_j} = \underbrace{\beta_{X_j} \cdot \tau_X}_{G_j \rightarrow X \rightarrow M} + \underbrace{\delta_j}_{G_j \rightarrow M}.$$

■

Throughout this Supplementary Material, we work in the asymptotic regime where $p \rightarrow \infty$. For two sequences of random variables, A and B , write $A \lesssim_p B$ if the ratio A/B is asymptotically bounded in probability. The strict relation, $A \prec_p B$, implies that $A/B \xrightarrow{p} 0$, and $\xrightarrow{p}$ denotes convergence in probability. $A \asymp_p B$ indicates both $A \lesssim_p B$ and $B \lesssim_p A$. We also use the O_p and o_p notation for $\lesssim_p$ and $\prec_p$, depending on which is more convenient. Finally, convergence in distribution is represented by $\xrightarrow{d}$.

Next, we follow the literature and assume that the estimated association effects are independent across j and across the three GWAS. In addition, they are normally distributed. To simplify the presentation, we also assume that the estimation error are of similar magnitude.

Assumption 2 (Estimated association effects)

- (i) For different SNPs $j' \neq j$, $(\hat{\beta}_{X_j}, \hat{\beta}_{M_j}, \hat{\beta}_{Y_j})$ and $(\hat{\beta}_{X_{j'}}, \hat{\beta}_{M_{j'}}, \hat{\beta}_{Y_{j'}})$ are independent;
- (ii) Estimated association effects from three GWAS are independent and normally distributed: The estimated association effects are mutually independent and follow normal distributions:

$$\hat{\beta}_{X_j} \sim \mathcal{N}(\beta_{X_j}, \sigma_{X_j}^2), \quad \hat{\beta}_{M_j} \sim \mathcal{N}(\beta_{M_j}, \sigma_{M_j}^2), \quad \hat{\beta}_{Y_j} \sim \mathcal{N}(\beta_{Y_j}, \sigma_{Y_j}^2).$$

In addition, there exists some $\nu \rightarrow 0$, such that $\{\sigma_{Y_j}/\nu, \sigma_{X_j}/\nu, \sigma_{M_j}/\nu : 1 \leq j \leq p\}$ are bounded and bounded away from zero. ■

Before introducing the other assumptions and additional notation, we first provide a brief discussion on our randomized instrumental variable selection procedure. We also present some results related to the estimated association effects and their standard errors. We recall from the main manuscript that two sets of (possibly) different IVs are selected for the exposure and the mediator using GWAS (I) and (III), respectively. Formally,

$$\mathcal{S}_x = \left\{ j : \left| \frac{\hat{\beta}_{X_j}}{\sigma_{X_j}} + Z_j \right| > \lambda, j = 1, \dots, p \right\} \text{ and } \mathcal{S}_m = \left\{ j : \left| \frac{\hat{\beta}_{M_j}}{\sigma_{M_j}} + Z'_j \right| > \lambda, j = 1, \dots, p \right\},$$

where $Z_1, \dots, Z_p$ and $Z'_1, \dots, Z'_p$ are iid draws from the $\mathcal{N}(0, \eta^2)$ distribution, and η is a pre-specified positive constant. Also recall that the cardinalities of $\mathcal{S}_x$ and $\mathcal{S}_m$ (i.e., number of selected instruments) are denoted by p_x and p_m , respectively.

To remove the winner's curse, we have introduced the following estimated association effects in the main manuscript:

$$\begin{aligned} \hat{\beta}_{X_j, \text{bc}} &= \hat{\beta}_{X_j} - \frac{\sigma_{X_j}}{\eta} \left(\phi(A_{j,+}) - \phi(A_{j,-}) \right) \left(\frac{\mathbf{1}_{(j \in \mathcal{S}_x)}}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} - \frac{\mathbf{1}_{(j \notin \mathcal{S}_x)}}{\Phi(A_{j,+}) - \Phi(A_{j,-})} \right), \\ \hat{\beta}_{M_j, \text{bc}} &= \hat{\beta}_{M_j} - \frac{\sigma_{M_j}}{\eta} \left(\phi(A'_{j,+}) - \phi(A'_{j,-}) \right) \left(\frac{\mathbf{1}_{(j \in \mathcal{S}_m)}}{1 - \Phi(A'_{j,+}) + \Phi(A'_{j,-})} - \frac{\mathbf{1}_{(j \notin \mathcal{S}_m)}}{\Phi(A'_{j,+}) - \Phi(A'_{j,-})} \right), \end{aligned}$$

where

$$A_{j,\pm} = -\frac{\hat{\beta}_{X_j}}{\sigma_{X_j}\eta} \pm \frac{\lambda}{\eta}, \text{ and } A'_{j,\pm} = -\frac{\hat{\beta}_{M_j}}{\sigma_{M_j}\eta} \pm \frac{\lambda}{\eta}.$$

In this Supplementary Material, we use ϕ and Φ to denote the probability density function and the cumulative distribution function of the standard normal distribution.

To understand the formula of, say, $\hat{\beta}_{X_j, \text{bc}}$, we may consider the following auxiliary “initial” estimator:

$$\hat{\beta}_{X_j, \text{ini}} = \hat{\beta}_{X_j} - \frac{\sigma_{X_j}}{\eta^2} Z_j.$$

Standard covariance calculation suggests that $\widehat{\beta}_{X_j, \text{ini}}$ is independent of the selection event, meaning that the initial estimator is unbiased both unconditionally and after conditioning on the selection event: $\mathbb{E}[\widehat{\beta}_{X_j, \text{ini}} | \mathcal{S}_x] = \mathbb{E}[\widehat{\beta}_{X_j, \text{ini}}] = \beta_{X_j}$. The proposed construction takes one step further to reduce conditional variance:

$$\begin{aligned}\widehat{\beta}_{X_j, \text{bc}} &= \mathbf{1}_{(j \in \mathcal{S}_x)} \mathbb{E}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] + \mathbf{1}_{(j \notin \mathcal{S}_x)} \mathbb{E}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \notin \mathcal{S}_x] \\ &= \widehat{\beta}_{X_j} - \frac{\sigma_{X_j}}{\eta^2} \left(\mathbf{1}_{(j \in \mathcal{S}_x)} \mathbb{E}[Z_j | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] + \mathbf{1}_{(j \notin \mathcal{S}_x)} \mathbb{E}[Z_j | \widehat{\beta}_{X_j}, j \notin \mathcal{S}_x] \right).\end{aligned}$$

Lemma 1 *Let Assumptions 1 and 2 hold. Then*

$$\begin{aligned}\mathbb{E}[\widehat{\beta}_{X_j, \text{bc}} | \mathcal{S}_x, \mathcal{S}_m] &= \mathbb{E}[\widehat{\beta}_{X_j, \text{bc}} | \mathcal{S}_x] = \mathbb{E}[\widehat{\beta}_{X_j, \text{bc}} | j \in \mathcal{S}_x] = \mathbb{E}[\widehat{\beta}_{X_j, \text{bc}} | j \notin \mathcal{S}_x] = \beta_{X_j}, \\ \mathbb{E}[\widehat{\beta}_{M_j, \text{bc}} | \mathcal{S}_x, \mathcal{S}_m] &= \mathbb{E}[\widehat{\beta}_{M_j, \text{bc}} | \mathcal{S}_m] = \mathbb{E}[\widehat{\beta}_{M_j, \text{bc}} | j \in \mathcal{S}_m] = \mathbb{E}[\widehat{\beta}_{M_j, \text{bc}} | j \notin \mathcal{S}_m] = \beta_{M_j}.\end{aligned}$$

■

The adjustment made above will also affect the variances of the estimated association effects, and therefore we also propose the following:

$$\begin{aligned}\widehat{\varsigma}_{X_j} &= \sigma_{X_j}^2 \left(1 - \frac{1}{\eta^2} \frac{A_{j,+} \phi(A_{j,+}) - A_{j,-} \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} + \frac{1}{\eta^2} \left(\frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} \right)^2 \right) \mathbf{1}_{(j \in \mathcal{S}_x)} \\ &\quad + \sigma_{X_j}^2 \left(1 - \frac{1}{\eta^2} \frac{-A_{j,+} \phi(A_{j,+}) + A_{j,-} \phi(A_{j,-})}{\Phi(A_{j,+}) - \Phi(A_{j,-})} + \frac{1}{\eta^2} \left(\frac{\phi(A_{j,+}) - \phi(A_{j,-})}{\Phi(A_{j,+}) - \Phi(A_{j,-})} \right)^2 \right) \mathbf{1}_{(j \notin \mathcal{S}_x)}, \\ \widehat{\varsigma}_{M_j} &= \sigma_{M_j}^2 \left(1 - \frac{1}{\eta^2} \frac{A'_{j,+} \phi(A'_{j,+}) - A'_{j,-} \phi(A'_{j,-})}{1 - \Phi(A'_{j,+}) + \Phi(A'_{j,-})} + \frac{1}{\eta^2} \left(\frac{\phi(A'_{j,+}) - \phi(A'_{j,-})}{1 - \Phi(A'_{j,+}) + \Phi(A'_{j,-})} \right)^2 \right) \mathbf{1}_{(j \in \mathcal{S}_m)} \\ &\quad + \sigma_{M_j}^2 \left(1 - \frac{1}{\eta^2} \frac{-A'_{j,+} \phi(A'_{j,+}) + A'_{j,-} \phi(A'_{j,-})}{\Phi(A'_{j,+}) - \Phi(A'_{j,-})} + \frac{1}{\eta^2} \left(\frac{\phi(A'_{j,+}) - \phi(A'_{j,-})}{\Phi(A'_{j,+}) - \Phi(A'_{j,-})} \right)^2 \right) \mathbf{1}_{(j \notin \mathcal{S}_m)}.\end{aligned}$$

To motivate the variance estimator, say of $\widehat{\beta}_{X_j, \text{bc}}$, we also notice that

$$\begin{aligned}\mathbb{V}[\widehat{\beta}_{X_j, \text{ini}}] &= \mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | j \in \mathcal{S}_x] = \mathbb{E}[\mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] | j \in \mathcal{S}_x] + \mathbb{V}[\mathbb{E}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] | j \in \mathcal{S}_x], \\ \mathbb{V}[\widehat{\beta}_{X_j, \text{ini}}] &= \mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | j \notin \mathcal{S}_x] = \mathbb{E}[\mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \notin \mathcal{S}_x] | j \notin \mathcal{S}_x] + \mathbb{V}[\mathbb{E}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \notin \mathcal{S}_x] | j \notin \mathcal{S}_x],\end{aligned}$$

or alternatively, we write

$$\begin{aligned}\varsigma_{X_j} &:= \mathbb{V}[\widehat{\beta}_{X_j, \text{bc}} | \mathcal{S}_x] = \mathbb{V}[\widehat{\beta}_{X_j, \text{ini}}] - \mathbf{1}_{(j \in \mathcal{S}_x)} \mathbb{E}[\mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] | j \in \mathcal{S}_x] \\ &\quad - \mathbf{1}_{(j \notin \mathcal{S}_x)} \mathbb{E}[\mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \notin \mathcal{S}_x] | j \notin \mathcal{S}_x].\end{aligned}$$

The above formula gives the general expression of the variance given the selection event. For the three terms on the right-hand side, the unconditional variance of the auxiliary “initial” estimator is quite easy to obtain. The challenge is that the second and the third term involve the unknown

association effects. To circumvent this technical difficulty, we define

$$\widehat{\varsigma}_{X_j} := \mathbb{V}[\widehat{\beta}_{X_j, \text{ini}}] - \mathbf{1}_{(j \in \mathcal{S}_x)} \mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] - \mathbf{1}_{(j \notin \mathcal{S}_x)} \mathbb{V}[\widehat{\beta}_{X_j, \text{ini}} | \widehat{\beta}_{X_j}, j \notin \mathcal{S}_x].$$

One can further show that ς_{X_j} has the same order as $\sigma_{X_j}^2$ (see Lemma S.9 in [Ma et al. 2023](#)). The following lemma provides a bound on the estimation error of $\widehat{\varsigma}_{X_j}$.

Lemma 2 *Let Assumptions 1 and 2 hold. Then*

$$\left| \sum_{j \in \mathcal{S}_x} (\widehat{\varsigma}_{X_j} - \varsigma_{X_j}) \right| \lesssim_p \nu^2 \sqrt{p_x}, \quad \text{and} \quad \left| \sum_{j \in \mathcal{S}_m} (\widehat{\varsigma}_{M_j} - \varsigma_{M_j}) \right| \lesssim_p \nu^2 \sqrt{p_m}.$$

Another useful result related to the Rao-Blackwellized association effects is their fourth moment calculation, which we discuss in the lemma below. Recall that we define

$$\begin{aligned} u_j &= \widehat{\beta}_{X_j} - \beta_{X_j}, & u_{j, \text{bc}} &= \widehat{\beta}_{X_j, \text{bc}} - \beta_{X_j}, \\ w_j &= \widehat{\beta}_{M_j} - \beta_{M_j}, & w_{j, \text{bc}} &= \widehat{\beta}_{M_j, \text{bc}} - \beta_{M_j}, \\ v_j &= \widehat{\beta}_{Y_j} - \beta_{Y_j}. \end{aligned}$$

Lemma 3 *Let Assumptions 1 and 2 hold. Then*

$$\mathbb{E}[u_{j, \text{bc}}^4 | \mathcal{S}_x, \mathcal{S}_m] \lesssim \nu^4, \quad \text{and} \quad \mathbb{E}[w_{j, \text{bc}}^4 | \mathcal{S}_x, \mathcal{S}_m] \lesssim \nu^4.$$

2 Convergence of the Inverse Matrix

Recall from the main manuscript that the direct and mediation effects are estimated by

$$\begin{bmatrix} \widehat{\theta} \\ \widehat{\tau}_Y \\ \widehat{\tau}_X \end{bmatrix} = \underbrace{\begin{bmatrix} \sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j, \text{bc}}^2 - \widehat{\varsigma}_{X_j}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} & 0 \\ \sum_{j \in \mathcal{S}_m} \frac{\widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_m} \frac{\widehat{\beta}_{M_j, \text{bc}}^2 - \widehat{\varsigma}_{M_j}}{\sigma_{Y_j}^2} & 0 \\ 0 & 0 & \sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j, \text{bc}}^2 - \widehat{\varsigma}_{X_j}}{\sigma_{M_j}^2} \end{bmatrix}^{-1}}_{=: \widehat{\mathbf{M}}^{-1}} \begin{bmatrix} \sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{Y_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_m} \frac{\widehat{\beta}_{M_j, \text{bc}} \widehat{\beta}_{Y_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{M_j, \text{bc}}}{\sigma_{M_j}^2} \end{bmatrix}.$$

In this section, we characterize the asymptotic behavior of the inverse matrix. It is useful to recall the kappas which measure the average instrument strength:

$$\kappa_x = \frac{1}{p_x} \sum_{j \in \mathcal{S}_x} \left(\frac{\beta_{X_j}}{\sigma_{X_j}} \right)^2, \quad \kappa_m = \frac{1}{p_m} \sum_{j \in \mathcal{S}_m} \left(\frac{\beta_{M_j}}{\sigma_{M_j}} \right)^2,$$

$$\kappa_{\mathbf{x}}^{\mathbf{m}} = \frac{1}{p_{\mathbf{x}}} \sum_{j \in \mathcal{S}_{\mathbf{x}}} \left(\frac{\beta_{M_j}}{\sigma_{M_j}} \right)^2, \quad \kappa_{\mathbf{m}}^{\mathbf{x}} = \frac{1}{p_{\mathbf{m}}} \sum_{j \in \mathcal{S}_{\mathbf{m}}} \left(\frac{\beta_{X_j}}{\sigma_{X_j}} \right)^2.$$

It is also helpful to define

$$\kappa_{\mathbf{x}}^{\delta} = \frac{1}{p_{\mathbf{x}}} \sum_{j \in \mathcal{S}_{\mathbf{x}}} \left(\frac{\delta_j}{\sigma_{M_j}} \right)^2, \quad \kappa_{\mathbf{m}}^{\delta} = \frac{1}{p_{\mathbf{m}}} \sum_{j \in \mathcal{S}_{\mathbf{m}}} \left(\frac{\delta_j}{\sigma_{M_j}} \right)^2.$$

While it is possible to allow the kappas to have different asymptotic order, this will necessarily lead to more cumbersome notation and lengthy theoretical discussions. To simplify the presentation, we make the following assumptions on the number of selected instruments and average instrument strength.

Assumption 3 (IV selection and strength)

There exist $\mathbf{p} \rightarrow \infty$ and $\kappa \gtrsim 1$, such that (i) $p_{\mathbf{x}}, p_{\mathbf{m}} \asymp_{\mathbf{p}} \mathbf{p}$, and (ii) $\kappa_{\mathbf{x}}, \kappa_{\mathbf{m}}, \kappa_{\mathbf{m}}^{\delta} \asymp_{\mathbf{p}} \kappa$; $\kappa_{\mathbf{x}}^{\mathbf{m}}, \kappa_{\mathbf{m}}^{\mathbf{x}}, \kappa_{\mathbf{x}}^{\delta} \lesssim_{\mathbf{p}} \kappa$.

We now characterize the asymptotic behavior of the entries of the matrix $\hat{\mathbf{M}}$.

Lemma 4 Let Assumptions 1–3 hold. Then

$$\begin{aligned} \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j}}{\sigma_{Y_j}^2} &= \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} + O_{\mathbf{p}}(\sqrt{\mathbf{p}\kappa}), \\ \sum_{j \in \mathcal{S}_{\mathbf{m}}} \frac{\hat{\beta}_{M_j, \text{bc}}^2 - \hat{\varsigma}_{M_j}}{\sigma_{Y_j}^2} &= \sum_{j \in \mathcal{S}_{\mathbf{m}}} \frac{\beta_{M_j}^2}{\sigma_{Y_j}^2} + O_{\mathbf{p}}(\sqrt{\mathbf{p}\kappa}), \\ \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} &= \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} + O_{\mathbf{p}}(\sqrt{\mathbf{p}\kappa}), \\ \sum_{j \in \mathcal{S}_{\mathbf{m}}} \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} &= \sum_{j \in \mathcal{S}_{\mathbf{m}}} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} + O_{\mathbf{p}}(\sqrt{\mathbf{p}\kappa}), \\ \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j}}{\sigma_{M_j}^2} &= \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\beta_{X_j}^2}{\sigma_{M_j}^2} + O_{\mathbf{p}}(\sqrt{\mathbf{p}\kappa}). \end{aligned}$$

Given the discussion in Lemma 4, it is natural to define

$$\mathbf{M} = \begin{bmatrix} \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} & 0 \\ \sum_{j \in \mathcal{S}_{\mathbf{m}}} \frac{\beta_{M_j} \beta_{X_j}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_{\mathbf{m}}} \frac{\beta_{M_j}^2}{\sigma_{Y_j}^2} & 0 \\ 0 & 0 & \sum_{j \in \mathcal{S}_{\mathbf{x}}} \frac{\beta_{X_j}^2}{\sigma_{M_j}^2} \end{bmatrix}.$$

To understand the properties of the “limiting matrix” $\mathbf{M}$, it is useful to study the off-diagonal terms. For this purpose, we make the following assumption.

Assumption 4 (Balanced pleiotropy)

The pleiotropic effects, $\alpha_1, \dots, \alpha_p$, and $\delta_1, \dots, \delta_p$, are (i) mutually independent, (ii) have a zero mean: $\mathbb{E}[\alpha_j] = \mathbb{E}[\delta_j] = 0$ and (iii) have bounded fourth moments: $\mathbb{E}[|\alpha_j|^4], \mathbb{E}[|\delta_j|^4] \leq C\nu^4\kappa^2$, where C is some constant that does not depend on j or p .

The lemma below shows that the matrix $\mathbf{M}$, after proper scaling, is bounded and asymptotically invertible.

Lemma 5 Let Assumptions 1–4 hold. Then

$$\frac{1}{p\kappa}\mathbf{M} \lesssim_p 1, \quad \text{and} \quad \varsigma_{\min}\left(\frac{1}{p\kappa}\mathbf{M}\right) \gtrsim_p 1,$$

where $\varsigma_{\min}(\cdot)$ denotes the smallest singular value. In addition,

$$\mathbf{M}^{-1}\widehat{\mathbf{M}} \xrightarrow{p} \mathbf{I},$$

and $\mathbf{I}$ is the identity matrix.

3 Large-sample Distribution and Statistical Inference

We start from the following decomposition employed in the main paper (see Section 6.6 for details):

$$\begin{bmatrix} \widehat{\theta} - \theta \\ \widehat{\tau}_Y - \tau_Y \\ \widehat{\tau}_X - \tau_X \end{bmatrix} = \widehat{\mathbf{M}}^{-1} \sum_{j=1}^p U_j,$$

where

$$U_j = \begin{bmatrix} u_{j,\theta} \\ u_{j,\tau_Y} \\ u_{j,\tau_X} \end{bmatrix} := \begin{bmatrix} \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\widehat{\beta}_{X_j, \text{bc}}(v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + \theta \widehat{\varsigma}_{X_j}}{\sigma_{Y_j}^2} \\ \mathbf{1}_{(j \in \mathcal{S}_m)} \cdot \frac{\widehat{\beta}_{M_j, \text{bc}}(v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + \tau_Y \widehat{\varsigma}_{M_j}}{\sigma_{Y_j}^2} \\ \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\widehat{\beta}_{X_j, \text{bc}}(w_{j, \text{bc}} + \delta_j - \tau_X u_{j, \text{bc}}) + \tau_X \widehat{\varsigma}_{X_j}}{\sigma_{M_j}^2} \end{bmatrix}. \quad (1)$$

Alternatively, we may also write

$$\begin{bmatrix} u_{j,\theta} \\ u_{j,\tau_Y} \\ u_{j,\tau_X} \end{bmatrix} := \begin{bmatrix} \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\beta_{X_j}(v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + u_{j, \text{bc}}(v_j + \alpha_j - \tau_Y w_{j, \text{bc}}) + \theta(\widehat{\varsigma}_{X_j} - u_{j, \text{bc}}^2)}{\sigma_{Y_j}^2} \\ \mathbf{1}_{(j \in \mathcal{S}_m)} \cdot \frac{\beta_{M_j}(v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + w_{j, \text{bc}}(v_j + \alpha_j - \theta u_{j, \text{bc}}) + \tau_Y(\widehat{\varsigma}_{M_j} - w_{j, \text{bc}}^2)}{\sigma_{Y_j}^2} \\ \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\beta_{X_j}(w_{j, \text{bc}} + \delta_j - \tau_X u_{j, \text{bc}}) + u_{j, \text{bc}}(w_{j, \text{bc}} + \delta_j) + \tau_X(\widehat{\varsigma}_{X_j} - u_{j, \text{bc}}^2)}{\sigma_{M_j}^2} \end{bmatrix}. \quad (2)$$

In the lemma below, we characterize the asymptotic order of several variance terms.

Lemma 6 *Let Assumptions 1–4 hold. Then*

$$\mathbb{V}\left[\sum_{j=1}^p u_{j,\theta} \middle| \mathcal{S}_x, \mathcal{S}_m\right], \quad \mathbb{V}\left[\sum_{j=1}^p u_{j,\tau_Y} \middle| \mathcal{S}_x, \mathcal{S}_m\right], \quad \mathbb{V}\left[\sum_{j=1}^p u_{j,\tau_X} \middle| \mathcal{S}_x, \mathcal{S}_m\right] \lesssim_p \mathfrak{p}\kappa^2.$$

From Lemmas 5 and 6, we conclude that our estimators converge at the rate $1/\sqrt{\mathfrak{p}}$. To further show that our estimators are asymptotically normally distributed, we verify Lyapunov’s condition with fourth moment calculation. To this end, we impose the following assumption.

Assumption 5 (No dominating instrument)

The instrument strengths satisfy:

$$\frac{\max_{j \in \mathcal{S}_x \cup \mathcal{S}_m} \beta_{X_j}^2}{\sum_{j \in \mathcal{S}_x} \beta_{X_j}^2} \xrightarrow{\mathfrak{p}} 0, \quad \text{and} \quad \frac{\max_{j \in \mathcal{S}_x \cup \mathcal{S}_m} \beta_{M_j}^2}{\sum_{j \in \mathcal{S}_m} \beta_{M_j}^2} \xrightarrow{\mathfrak{p}} 0.$$

The above assumption is very mild, as it only rules out the (unlikely) scenario in which a few instruments are very strong while all the others are weak/irrelevant.

Lemma 7 *Let Assumptions 1–5 hold. Then*

$$\mathbb{E}\left[\sum_{j=1}^p u_{j,\theta}^4 \middle| \mathcal{S}_x, \mathcal{S}_m\right], \quad \mathbb{E}\left[\sum_{j=1}^p u_{j,\tau_Y}^4 \middle| \mathcal{S}_x, \mathcal{S}_m\right], \quad \mathbb{E}\left[\sum_{j=1}^p u_{j,\tau_X}^4 \middle| \mathcal{S}_x, \mathcal{S}_m\right] \prec_p \mathfrak{p}^2 \kappa^4.$$

Our final assumption requires that the “middle matrix” is asymptotically nonsingular, which is critical to the asymptotic normality result. This assumption is again quite mild, as it only requires that the estimated effects are not perfectly correlated. We recall that

$$\mathbf{U} = \sum_{j=1}^p \mathbb{E}[U_j U_j^\top \mid \mathcal{S}_x, \mathcal{S}_m].$$

Assumption 6 (No perfect correlation)

The variance matrix $\mathbf{U}$ satisfies

$$\varsigma_{\min}\left(\frac{1}{\mathfrak{p}\kappa^2}\mathbf{U}\right) \gtrsim_p 1,$$

where as before, $\varsigma_{\min}(\cdot)$ denotes the smallest singular value.

Lemma 8 *Let Assumptions 1–6 hold. Then conditional on the selection events,*

$$\mathbf{U}^{-\frac{1}{2}}\left(\sum_{j=1}^p U_j\right) \xrightarrow{\mathfrak{d}} \mathcal{N}(0, \mathbf{I}),$$

where $\mathbf{I}$ is the identity matrix.

To provide valid standard errors, we make the following observation (See Section 6.10 for details):

$$\begin{aligned}
u_{j,\theta} &= \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\hat{\beta}_{X_j, \text{bc}}(\hat{\beta}_{Y_j} - \tau_Y \hat{\beta}_{M_j, \text{bc}}) + \theta(\hat{\varsigma}_{X_j} - \hat{\beta}_{X_j, \text{bc}}^2)}{\sigma_{Y_j}^2}, \\
u_{j,\tau_Y} &= \mathbf{1}_{(j \in \mathcal{S}_m)} \cdot \frac{\hat{\beta}_{M_j, \text{bc}}(\hat{\beta}_{Y_j} - \theta \hat{\beta}_{X_j, \text{bc}}) + \tau_Y(\hat{\varsigma}_{M_j} - \hat{\beta}_{M_j, \text{bc}}^2)}{\sigma_{Y_j}^2}, \\
u_{j,\tau_X} &= \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}} + \tau_X(\hat{\varsigma}_{X_j} - \hat{\beta}_{X_j, \text{bc}}^2)}{\sigma_{M_j}^2}.
\end{aligned} \tag{3}$$

Therefore, it is natural to define the following feasible sample counterpart:

$$\begin{aligned}
\hat{u}_{j,\theta} &= \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\hat{\beta}_{X_j, \text{bc}}(\hat{\beta}_{Y_j} - \hat{\tau}_Y \hat{\beta}_{M_j, \text{bc}}) + \hat{\theta}(\hat{\varsigma}_{X_j} - \hat{\beta}_{X_j, \text{bc}}^2)}{\sigma_{Y_j}^2}, \\
\hat{u}_{j,\tau_Y} &= \mathbf{1}_{(j \in \mathcal{S}_m)} \cdot \frac{\hat{\beta}_{M_j, \text{bc}}(\hat{\beta}_{Y_j} - \hat{\theta} \hat{\beta}_{X_j, \text{bc}}) + \hat{\tau}_Y(\hat{\varsigma}_{M_j} - \hat{\beta}_{M_j, \text{bc}}^2)}{\sigma_{Y_j}^2}, \\
\hat{u}_{j,\tau_X} &= \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}} + \hat{\tau}_X(\hat{\varsigma}_{X_j} - \hat{\beta}_{X_j, \text{bc}}^2)}{\sigma_{M_j}^2},
\end{aligned}$$

and

$$\hat{U}_j = \begin{bmatrix} \hat{u}_{j,\theta} \\ \hat{u}_{j,\tau_Y} \\ \hat{u}_{j,\tau_X} \end{bmatrix}.$$

Now define

$$\hat{V} = \hat{M}^{-1} \hat{U} (\hat{M}^{-1})^\top, \quad \text{where } \hat{U} = \sum_{j=1}^p \hat{U}_j \hat{U}_j^\top.$$

The lemma below (together with the convergence of $\hat{M}$, established in Lemma 5) shows the validity of the above variance estimator.

Lemma 9 *Let Assumptions 1–6 hold. Then*

$$\mathbf{U}^{-1} \hat{U} \xrightarrow{P} \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix.

Finally, we present the following main theorem, which shows the asymptotic normality of our estimators and the validity of the estimated variance $\widehat{\mathbf{V}}$.

Theorem 1 *Let Assumptions 1–6 hold. Then conditional on the selection events,*

$$\widehat{\mathbf{V}}^{-\frac{1}{2}} \begin{bmatrix} \widehat{\theta} - \theta \\ \widehat{\tau}_Y - \tau_Y \\ \widehat{\tau}_X - \tau_X \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \mathbf{I}),$$

where $\mathbf{I}$ is the identity matrix.

4 Asymptotic Efficiency Comparison with MVMR

In this section, we investigate our proposed method's efficiency gain in oracle settings. To be precise, we assume the sets of relevant instruments are known: $\mathcal{S}_x^* = \{j : \beta_{X_j} \neq 0\}$ and $\mathcal{S}_m^* = \{j : \beta_{M_j} \neq 0\}$. Therefore, no instrument selection is needed. Another assumption we impose is that the estimation errors have proportional variances: $\sigma_{X_j}^2 = c_X \sigma_{Y_j}^2$ and $\sigma_{M_j}^2 = c_M \sigma_{Y_j}^2$ for some constants c_X and c_M . We believe this condition can be relaxed with the cost of having much lengthier derivation.

We compare our proposed method to the measurement error bias corrected version of inverse variance weighted multivariable Mendelian randomization estimator (we denote it as DMVMR). In this oracle setting, they can be written as, respectively,

$$\begin{bmatrix} \widehat{\theta}^* \\ \widehat{\tau}_Y^* \end{bmatrix} = \begin{bmatrix} \sum_{j \in \mathcal{S}_x^*} \frac{\widehat{\beta}_{X_j}^2 - \sigma_{X_j}^2}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_x^*} \frac{\widehat{\beta}_{X_j} \widehat{\beta}_{M_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_m^*} \frac{\widehat{\beta}_{X_j} \widehat{\beta}_{M_j}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_m^*} \frac{\widehat{\beta}_{M_j}^2 - \sigma_{M_j}^2}{\sigma_{Y_j}^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j \in \mathcal{S}_x^*} \frac{\widehat{\beta}_{Y_j} \widehat{\beta}_{X_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_m^*} \frac{\widehat{\beta}_{Y_j} \widehat{\beta}_{M_j}}{\sigma_{Y_j}^2} \end{bmatrix}.$$

and

$$\begin{bmatrix} \widehat{\theta}_{\text{DMVMR}}^* \\ \widehat{\tau}_{Y, \text{DMVMR}}^* \end{bmatrix} = \begin{bmatrix} \sum_{j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*} \frac{\widehat{\beta}_{X_j}^2 - \sigma_{X_j}^2}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*} \frac{\widehat{\beta}_{X_j} \widehat{\beta}_{M_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*} \frac{\widehat{\beta}_{X_j} \widehat{\beta}_{M_j}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*} \frac{\widehat{\beta}_{M_j}^2 - \sigma_{M_j}^2}{\sigma_{Y_j}^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*} \frac{\widehat{\beta}_{Y_j} \widehat{\beta}_{X_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*} \frac{\widehat{\beta}_{Y_j} \widehat{\beta}_{M_j}}{\sigma_{Y_j}^2} \end{bmatrix}.$$

We will ignore the estimation of τ_X in this section, as $\widehat{\tau}_X^*$ (or $\widehat{\tau}_{X, \text{MVMR}}^*$) is obtained through a separate structural equation. With slight abuse of notation, we will use $\mathbf{V}^*$ to denote the asymptotic variance matrix of $\widehat{\theta}^*$ and $\widehat{\tau}_Y^*$, and use $\mathbf{V}_{\text{DMVMR}}^*$ for the measurement error bias corrected version of the MVMR approach.

We demonstrate that our estimator is more efficient asymptotically under the “balanced instruments” assumption below:

Assumption 7 (Balanced instruments)

The cardinalities of $\mathcal{S}_m^* \setminus \mathcal{S}_x^*$ and $\mathcal{S}_x^* \setminus \mathcal{S}_m^*$ satisfy

$$\frac{|\mathcal{S}_m^* \setminus \mathcal{S}_x^*|}{|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|} \leq 6.85, \quad \text{and} \quad \frac{|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|}{|\mathcal{S}_m^* \setminus \mathcal{S}_x^*|} \leq 6.85.$$

We briefly discuss the assumption. As its name suggests, Assumption 7 requires that the two sets of instruments are balanced: there exist IVs only relevant to the exposure but not the mediator ($\mathcal{S}_x^* \setminus \mathcal{S}_m^*$); similarly, there are IVs only relevant to the mediator ($\mathcal{S}_m^* \setminus \mathcal{S}_x^*$). In addition, their numbers do not differ too much. The thresholds, 6.85, is solved from a polynomial equation. See the proof of the theorem below for details.

Theorem 2 Consider the oracle setting, and let Assumption 7 hold. Then $V_{\text{DMVMR}}^* - V^*$ is positive definite.

The previous theorem shows that our proposed MAGIC approach is asymptotically more efficient in the balanced instruments scenario. An equally important scenario, which we discuss below, is when $\mathcal{S}_m^* \supsetneq \mathcal{S}_x^*$. To simplify the discussion, we further assume that $c_X = c_M = 1$. With a little abuse of notation, we use $p_x, p_m, \kappa_x, \kappa_m^\delta$ to denote the average instrument strengths given the oracles $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$. We will show that the estimated causal effects are more efficient (relative to the DMVMR method) under the following assumption.

Assumption 8 (Instrument strengths differential)

(i) $8p_m\kappa_m^\delta > p_x\kappa_x$. (ii) $\tau_X^2(1 + 2\theta^2 + \tau_Y^2) > 2\tau_X\theta\tau_Y$.

To understand this assumption, we notice that both (i) and (ii) are related to the asymptotic behavior of the matrix $\mathbf{M}$. To be even more precise, consider the total instrument strengths for the mediator: $p_m\kappa_m$. By employing the proof technique for Lemma 5, it satisfies

$$p_m\kappa_m = \tau_X^2 p_x\kappa_x + p_m\kappa_m^\delta + \text{negligible terms}.$$

Therefore, what Assumption 8 entails is that the total instrument strength for the mediator is not too small compared to the total instrument strength for the exposure. This is quite reasonable given that we consider the scenario where there are more instruments for the mediator.

To state the next result, we use V_θ^* and $V_{\tau_Y}^*$ to denote the asymptotic variances of $\hat{\theta}^*$ and $\hat{\tau}_Y^*$. Accordingly, $V_{\theta, \text{DMVMR}}^*$ and $V_{\tau_Y, \text{DMVMR}}^*$ are the asymptotic variances using the DMVMR approach.

Theorem 3 Consider the oracle setting. Under Assumption 8(i), $V_{\theta, \text{DMVMR}}^* > V_\theta^*$. Under Assumption 8(ii), $V_{\tau_Y, \text{DMVMR}}^* > V_{\tau_Y}^*$.

5 Plug-in Estimation and Imperfect IV Selection Bias

In this section, we discuss an alternative estimation strategy motivated by the structural equations. While the estimator may seem reasonable, we show that it can be severely biased due to imperfect IV selection.

We start from the two structural equations,

$$\beta_{Y_j} = \beta_{X_j} \cdot \theta + \beta_{M_j} \cdot \tau_Y + \alpha_j, \quad \beta_{M_j} = \beta_{X_j} \cdot \tau_X + \delta_j.$$

A seemingly natural approach is to employ the “sample analogue” or “plug-in” principal:

$$\begin{aligned} \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}} &= \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \left(\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j} \right) \cdot \hat{\tau}_X^{\text{PI}}, \\ \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{Y_j} &= \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \left(\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j} \right) \cdot \hat{\theta}^{\text{PI}} + \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \mathbf{1}_{(j \in \mathcal{S}_m)} \hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}} \cdot \hat{\tau}_Y^{\text{PI}}, \\ \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_m)} \hat{\beta}_{M_j, \text{bc}} \hat{\beta}_{Y_j} &= \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \mathbf{1}_{(j \in \mathcal{S}_m)} \hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}} \cdot \hat{\theta}^{\text{PI}} + \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_m)} \left(\hat{\beta}_{M_j, \text{bc}}^2 - \hat{\varsigma}_{M_j} \right) \cdot \hat{\tau}_Y^{\text{PI}}, \end{aligned}$$

leading to

$$\begin{bmatrix} \hat{\theta}^{\text{PI}} \\ \hat{\tau}_Y^{\text{PI}} \\ \hat{\tau}_X^{\text{PI}} \end{bmatrix} = \underbrace{\begin{bmatrix} \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_x \cap \mathcal{S}_m} \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} & 0 \\ \sum_{j \in \mathcal{S}_x \cap \mathcal{S}_m} \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_m} \frac{\hat{\beta}_{M_j, \text{bc}}^2 - \hat{\varsigma}_{M_j}}{\sigma_{Y_j}^2} & 0 \\ 0 & 0 & \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j}}{\sigma_{M_j}^2} \end{bmatrix}^{-1}}_{=: \hat{\mathbf{M}}^{\text{PI}, -1}} \begin{bmatrix} \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{Y_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_m} \frac{\hat{\beta}_{M_j, \text{bc}} \hat{\beta}_{Y_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}}}{\sigma_{M_j}^2} \end{bmatrix}$$

Compared to our proposed MAGIC approach, this alternative estimator not only replaces the unknown instrument associations with their estimates, but also the unknown sets of relevant IVs, $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$, by their empirical counterparts $\mathcal{S}_x$ and $\mathcal{S}_m$. For this reason, we will call this alternative estimator “plug-in.”

While this estimator may seem appealing, it is actually biased due to imperfect IV selection. To see this new source of bias, we notice that the estimator can be decomposed into

$$\begin{bmatrix} \hat{\theta}^{\text{PI}} \\ \hat{\tau}_Y^{\text{PI}} \\ \hat{\tau}_X^{\text{PI}} \end{bmatrix} = \begin{bmatrix} \theta \\ \tau_Y \\ \tau_X \end{bmatrix} + \hat{\mathbf{M}}^{\text{PI}, -1} \left(\sum_{j=1}^p U_j + \mathbf{R} \right),$$

where U_j remains the same (see, for example, Section 3), and

$$\mathbf{R} = \begin{bmatrix} \tau_Y \sum_{j \in \mathcal{S}_x \setminus \mathcal{S}_m} \frac{\hat{\beta}_{M_j, bc} \hat{\beta}_{X_j, bc}}{\sigma_{Y_j}^2} \\ \theta \sum_{j \in \mathcal{S}_m \setminus \mathcal{S}_x} \frac{\hat{\beta}_{X_j, bc} \hat{\beta}_{M_j, bc}}{\sigma_{Y_j}^2} \\ 0 \end{bmatrix}.$$

Following the same proof technique, one can show results similar to Lemmas 4 and 5 that the inverse matrix converges with proper scaling, which further suggests that

$$\begin{bmatrix} \hat{\theta}^{\text{PI}} \\ \hat{\tau}_Y^{\text{PI}} \\ \hat{\tau}_X^{\text{PI}} \end{bmatrix} = \begin{bmatrix} \theta \\ \tau_Y \\ \tau_X \end{bmatrix} + \tilde{\mathbf{M}}^{-1} \left(\sum_{j=1}^p U_j + \mathbf{R} \right) + \text{asymptotically negligible terms},$$

and

$$\tilde{\mathbf{M}} = \begin{bmatrix} \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_x \cap \mathcal{S}_m} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} & 0 \\ \sum_{j \in \mathcal{S}_x \cap \mathcal{S}_m} \frac{\beta_{M_j} \beta_{X_j}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_m} \frac{\beta_{M_j}^2}{\sigma_{Y_j}^2} & 0 \\ 0 & 0 & \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{M_j}^2} \end{bmatrix}.$$

An important observation is that this alternative plug-in estimator is no longer conditionally unbiased. To be very clear, while one can show

$$\mathbb{E} \left[\tilde{\mathbf{M}}^{-1} \sum_{j=1}^p U_j \middle| \mathcal{S}_x, \mathcal{S}_m \right] = 0,$$

the other remainder term is generally not mean zero:

$$\mathbb{E} \left[\tilde{\mathbf{M}}^{-1} \mathbf{R} \middle| \mathcal{S}_x, \mathcal{S}_m \right] \neq 0 \text{ in general.}$$

As one can anticipate, the rate at which this remainder term $\tilde{\mathbf{M}}^{-1} \mathbf{R}$ vanishes depends crucially on how well/fast the empirically selected instruments, $\mathcal{S}_x$ and $\mathcal{S}_m$, recover their oracle counterparts $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$. Unfortunately, oracle IV selection guarantee may be difficult to establish, which means this alternative plug-in estimator can be severely biased in finite samples. We call this new source of bias as imperfect IV selection bias: it arises due to imperfect IV selection.

Our proposed MAGIC approach does not suffer from this imperfect IV selection bias. To see this reason, we recall from the main manuscript that when constructing the MAGIC estimator, we multiply the two sides of the structural equation by the same selection indicator (either $\mathbf{1}_{(j \in \mathcal{S}_x)}$ or

$\mathbf{1}_{(j \in \mathcal{S}_m)}$). In contrast, the plug-in approach incorporates different selection events ($\mathbf{1}_{(j \in \mathcal{S}_x)}$, $\mathbf{1}_{(j \in \mathcal{S}_m)}$, or the joint event $\mathbf{1}_{(j \in \mathcal{S}_x \cap \mathcal{S}_m)}$) to construct empirical analogues of the structural equations. As a result, our estimator remains asymptotically unbiased even when selection is imperfect. In fact, we do not even require selection to be consistent.

6 Proofs

6.1 Proof of Lemma 1

Proof. We demonstrate the proof with $\widehat{\beta}_{X_j, \text{bc}}$. The analysis for $\widehat{\beta}_{M_j, \text{bc}}$ follows a similar reasoning. First, we notice that $\widehat{\beta}_{X_j, \text{bc}}$ is independent of the selection event $\mathcal{S}_m$ as they are from different GWAS, which means

$$\mathbb{E}[\widehat{\beta}_{X_j, \text{bc}} | \mathcal{S}_x, \mathcal{S}_m] = \mathbb{E}[\widehat{\beta}_{X_j, \text{bc}} | \mathcal{S}_x].$$

We then employ Lemma 10 to find the explicit expression of the Rao-Blackwellized estimators. ■

Lemma 10 *Let $Z \sim \mathcal{N}(\mu, \sigma^2)$. Assume $b > a$. Then*

$$\begin{aligned} \mathbb{E}[Z | a \leq Z \leq b] &= \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ \mathbb{E}[Z^2 | a \leq Z \leq b] &= \mu^2 + \sigma^2 + \sigma^2 \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} + 2\mu\sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}. \end{aligned}$$

In addition,

$$\mathbb{V}[Z | a \leq Z \leq b] = \sigma^2 \left[1 + \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} - \left(\frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \right)^2 \right].$$

Similarly,

$$\begin{aligned} \mathbb{E}[Z | Z \leq a \text{ or } Z \geq b] &= \mu + \sigma \frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} \\ \mathbb{E}[Z^2 | Z \leq a \text{ or } Z \geq b] &= \mu^2 + 2\mu\sigma \frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{\Phi(\frac{a-\mu}{\sigma}) + 1 - \Phi(\frac{b-\mu}{\sigma})} + \sigma^2 + \sigma^2 \frac{-\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) + \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{a-\mu}{\sigma}) + 1 - \Phi(\frac{b-\mu}{\sigma})}, \end{aligned}$$

and

$$\mathbb{V}[Z | Z \leq a \text{ or } Z \geq b] = \sigma^2 \left[1 + \frac{-\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) + \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{a-\mu}{\sigma}) + 1 - \Phi(\frac{b-\mu}{\sigma})} - \left(\frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{\Phi(\frac{a-\mu}{\sigma}) + 1 - \Phi(\frac{b-\mu}{\sigma})} \right)^2 \right].$$

Proof. The first and second moments of the truncated normal distributions can be easily obtained

via the moment generating function. See Lemma S.18 and Lemma S.19 of [Ma et al. \(2023\)](#). $\blacksquare$

6.2 Proof of Lemma 2

Proof. We illustrate with $\widehat{\varsigma}_{X_j}$. The proof for $\widehat{\varsigma}_{M_j}$ is similar. To start, we notice that $\widehat{\varsigma}_{X_j}$ is unbiased for $\sigma_{X_j, \text{bc}}^2$, and hence it suffices to bound the variance:

$$\begin{aligned} \left| \sum_{j \in \mathcal{S}_x} (\widehat{\varsigma}_{X_j} - \varsigma_{X_j}) \right| &= O_p \left(\sqrt{\sum_{j \in \mathcal{S}_x} \mathbb{E} \left[(\widehat{\varsigma}_{X_j} - \varsigma_{X_j})^2 \middle| j \in \mathcal{S}_x \right]} \right) \\ &= \frac{\sigma_{X_j}^2}{\eta^4} O_p \left(\sqrt{\sum_{j \in \mathcal{S}_x} \mathbb{E} \left[\left(\mathbb{V}[Z_j | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] - \mathbb{E}[\mathbb{V}[Z_j | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] | j \in \mathcal{S}_x] \right)^2 \middle| j \in \mathcal{S}_x \right]} \right) \\ &= \frac{\sigma_{X_j}^2}{\eta^4} O_p \left(\sqrt{\sum_{j \in \mathcal{S}_x} \mathbb{E} \left[\left(\mathbb{V}[Z_j | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] \right)^2 \middle| j \in \mathcal{S}_x \right]} \right) \\ &= \frac{\sigma_{X_j}^2}{\eta^4} O_p \left(\sqrt{\sum_{j \in \mathcal{S}_x} \mathbb{E} \left[Z_j^4 | j \in \mathcal{S}_x \right]} \right) = O_p(\nu^2 \sqrt{p_x}). \end{aligned}$$

See Lemma S.12 and its proof in [Ma et al. \(2023\)](#) for additional details. $\blacksquare$

6.3 Proof of Lemma 3

Proof. We illustrate with $u_{j, \text{bc}}$. The proof for the fourth moment of $w_{j, \text{bc}}$ is analogous. To start, we notice that $u_{j, \text{bc}}$ is independent of the selection events in $\mathcal{S}_m$, which means $\mathbb{E}[u_{j, \text{bc}}^4 | \mathcal{S}_x, \mathcal{S}_m] = \mathbb{E}[u_{j, \text{bc}}^4 | \mathcal{S}_x]$. The above further reduces to computing $\mathbb{E}[u_{j, \text{bc}}^4 | j \in \mathcal{S}_x]$ and $\mathbb{E}[u_{j, \text{bc}}^4 | j \notin \mathcal{S}_x]$. We demonstrate the first case.

We write out the explicit form of $\mathbb{E}[u_{j, \text{bc}}^4 | j \in \mathcal{S}_x]$. For simplicity, we let $\eta = 1$ and $\beta_{X_j} = 0$. Then

$$\begin{aligned} \frac{\mathbb{E}[u_{j, \text{bc}}^4 | j \in \mathcal{S}_x]}{\sigma_{X_j}^4} &= \frac{1}{\mathbb{P}[j \in \mathcal{S}_x]} \int_{-\infty}^{\infty} \left(y - \frac{\phi(-\lambda + y) - \phi(-\lambda - y)}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right)^4 \phi(y) \left[\Phi(-\lambda + y) + \Phi(-\lambda - y) \right] dy \\ &\lesssim \int_{-\infty}^{\infty} \left(y - \frac{\phi(-\lambda + y) - \phi(-\lambda - y)}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right)^4 \phi(y) \left[\Phi(-\lambda + y) + \Phi(-\lambda - y) \right] dy. \end{aligned}$$

It is straightforward to see that the above is bounded, and hence $\mathbb{E}[u_{j, \text{bc}}^4 | j \in \mathcal{S}_x] \lesssim \nu^4$. $\blacksquare$

6.4 Proof of Lemma 4

Proof. To start, we write

$$\sum_{j \in \mathcal{S}_x} \widehat{\beta}_{X_j, \text{bc}}^2 = \sum_{j \in \mathcal{S}_x} (\beta_{X_j} + u_{j, \text{bc}})^2 = \sum_{j \in \mathcal{S}_x} (\beta_{X_j}^2 + u_{j, \text{bc}}^2 + 2u_{j, \text{bc}}\beta_{X_j}).$$

Next, we employ Lemma 3, which shows

$$\sum_{j \in \mathcal{S}_x} u_{j,\text{bc}}^2 = \sum_{j \in \mathcal{S}_x} \sigma_{X_j,\text{bc}}^2 + O_p \left(\sqrt{p_x \nu^4} \right).$$

With Lemma 2, we further have that

$$\begin{aligned} \sum_{j \in \mathcal{S}_x} u_{j,\text{bc}}^2 &= \sum_{j \in \mathcal{S}_x} \sigma_{X_j,\text{bc}}^2 + O_p \left(\sqrt{p_x \nu^4} \right) = \sum_{j \in \mathcal{S}_x} \widehat{\varsigma}_{X_j} + \sum_{j \in \mathcal{S}_x} (\sigma_{X_j,\text{bc}}^2 - \widehat{\varsigma}_{X_j}) + O_p \left(\sqrt{p_x \nu^4} \right) \\ &= \sum_{j \in \mathcal{S}_x} \widehat{\varsigma}_{X_j} + O_p \left(\sqrt{p_x \nu^4} \right). \end{aligned}$$

Next, by variance calculation, we have that

$$\sum_{j \in \mathcal{S}_x} u_{j,\text{bc}} \beta_{X_j} = O_p \left(\sqrt{\mathbb{V} \left[\beta_{X_j} u_{j,\text{bc}} \mid j \in \mathcal{S}_x \right]} \right) = O_p(\sqrt{\kappa_x p_x \nu^4}).$$

As a result,

$$\sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j,\text{bc}}^2 - \widehat{\varsigma}_{X_j}}{\sigma_{Y_j}^2} = \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} + O_p \left(\sqrt{p_x (\kappa_x + 1)} \right) = O_p(\sqrt{p \kappa}).$$

By applying similar technique, we can show that

$$\sum_{j \in \mathcal{S}_m} \frac{\widehat{\beta}_{M_j,\text{bc}}^2 - \widehat{\varsigma}_{M_j}}{\sigma_{Y_j}^2} = \sum_{j \in \mathcal{S}_m} \frac{\beta_{M_j}^2}{\sigma_{Y_j}^2} + O_p \left(\sqrt{p_m (\kappa_m + 1)} \right) = O_p(\sqrt{p \kappa}),$$

and

$$\sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j,\text{bc}}^2 - \widehat{\varsigma}_{X_j}}{\sigma_{M_j}^2} = \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{M_j}^2} + O_p \left(\sqrt{p_x (\kappa_x + 1)} \right) = O_p(\sqrt{p \kappa}).$$

Next, we study the off-diagonal terms. For example, we can write

$$\begin{aligned} \sum_{j \in \mathcal{S}_x} \widehat{\beta}_{X_j,\text{bc}} \widehat{\beta}_{M_j,\text{bc}} &= \sum_{j \in \mathcal{S}_x} (\beta_{X_j} + u_{j,\text{bc}})(\beta_{M_j} + w_{j,\text{bc}}) \\ &= \sum_{j \in \mathcal{S}_x} \beta_{X_j} \beta_{M_j} + \sum_{j \in \mathcal{S}_x} \beta_{X_j} w_{j,\text{bc}} + \sum_{j \in \mathcal{S}_x} u_{j,\text{bc}} \beta_{M_j} + \sum_{j \in \mathcal{S}_x} u_{j,\text{bc}} w_{j,\text{bc}}. \end{aligned}$$

With variance calculation, we can show that

$$\sum_{j \in \mathcal{S}_x} \widehat{\beta}_{X_j,\text{bc}} \widehat{\beta}_{M_j,\text{bc}} = \sum_{j \in \mathcal{S}_x} \beta_{X_j} \beta_{M_j} + O_p \left(\sqrt{(\kappa_x + \kappa_m + 1) p_x \nu^4} \right),$$

implying

$$\sum_{j \in \mathcal{S}_x} \frac{\widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} = \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} + O_p(\sqrt{\mathfrak{p}\kappa}).$$

Similarly,

$$\sum_{j \in \mathcal{S}_m} \frac{\widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} = \sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} + O_p(\sqrt{\mathfrak{p}\kappa}),$$

which closes the proof. ■

6.5 Proof of Lemma 5

Proof. To start, we consider the $(1, 2)$ -entry of $\mathbf{M}$, which can be written as

$$\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} = \tau_X \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} + \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j} \delta_j}{\sigma_{Y_j}^2}.$$

Thanks to our Assumption 4, the second term has the bound

$$\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j} \delta_j}{\sigma_{Y_j}^2} = O_p\left(\sqrt{\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2 \mathbb{E}[\delta_j^2]}{\sigma_{Y_j}^4}}\right).$$

Finally, our Assumption 3 suggests that $\mathbb{E}[\delta_j^2] \asymp_p \nu^2 \kappa$. Then,

$$\frac{1}{\mathfrak{p}\kappa} \mathbf{M} = \begin{bmatrix} \frac{1}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} & \frac{\tau_X}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} + O_p\left(\frac{1}{\sqrt{\mathfrak{p}}}\right) & 0 \\ \frac{\tau_X}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} + \frac{1}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j} \delta_j}{\sigma_{Y_j}^2} & \frac{\tau_X^2}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} + \frac{1}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_m} \frac{\delta_j^2}{\sigma_{Y_j}^2} + \frac{2\tau_X}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j} \delta_j}{\sigma_{Y_j}^2} & 0 \\ 0 & 0 & \frac{1}{\mathfrak{p}\kappa} \sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \end{bmatrix}.$$

Obviously the above is bounded. To show that $\frac{1}{\mathfrak{p}\kappa} \mathbf{M}$ is asymptotically invertible, it suffices to focus on the upper left part. The determinant of this submatrix is

$$\begin{aligned} & \frac{\tau_X^2}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) + \frac{1}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\delta_j^2}{\sigma_{Y_j}^2} \right) + \frac{2\tau_X}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j} \delta_j}{\sigma_{Y_j}^2} \right) \\ & - \frac{\tau_X^2}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) - \frac{\tau_X}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\beta_{X_j} \delta_j}{\sigma_{Y_j}^2} \right) + O_p\left(\frac{1}{\sqrt{\mathfrak{p}}}\right) \\ & = \frac{1}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\delta_j^2}{\sigma_{Y_j}^2} \right) + \frac{1}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\tau_X \beta_{X_j} \delta_j}{\sigma_{Y_j}^2} \right) + O_p\left(\frac{1}{\sqrt{\mathfrak{p}}}\right). \end{aligned}$$

If $\tau_X = 0$, then the above is obviously bounded from below since the second term is zero. Otherwise, we have the following

$$\frac{1}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\tau_X \beta_{X_j} \delta_j}{\sigma_{Y_j}^2} \right) = \frac{1}{\mathfrak{p}^2 \kappa^2} \left(\sum_{j \in \mathcal{S}_x} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} \right) \left(\sum_{j \in \mathcal{S}_m} \frac{\tau_X \beta_{X_j} \mathbb{E}[\delta_j | j \in \mathcal{S}_m]}{\sigma_{Y_j}^2} \right) + O_p \left(\frac{1}{\sqrt{\mathfrak{p}}} \right).$$

By Lemma 11 below, one has

$$\tau_X \beta_{X_j} \mathbb{E}[\delta_j | j \in \mathcal{S}_m] \geq 0.$$

Therefore, $\frac{1}{\mathfrak{p}\kappa} \mathbf{M}$ is bounded and asymptotically invertible. To conclude the proof of the lemma, we notice that Lemma 4 implies

$$\frac{1}{\mathfrak{p}\kappa} \widehat{\mathbf{M}} = \frac{1}{\mathfrak{p}\kappa} \mathbf{M} + o_p(1),$$

and therefore

$$\mathbf{M}^{-1} \widehat{\mathbf{M}} = \left(\frac{1}{\mathfrak{p}\kappa} \mathbf{M} \right)^{-1} \left(\frac{1}{\mathfrak{p}\kappa} \mathbf{M} + o_p(1) \right).$$

■

Lemma 11 *Let U , V and W be three independent random variables with finite variances, and (i) V is symmetric around 0; (ii) $W \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 \geq 0$. Let $\lambda > 0$ be a constant. Then*

$$\mathbb{E}[UV \mid |U + V + W| > \lambda] \geq 0.$$

Proof. Case 1: $\sigma > 0$

To begin with, consider the case that $\sigma > 0$. Then it is without loss of generality to assume $\sigma = 1$. We have

$$\mathbb{E}[V \mid |u + V + W| > \lambda] \propto \mathbb{E}[V \mathbf{1}(|u + V + W| > \lambda)] = \int \phi(w - u) g(w) dw,$$

where

$$g(w) = \left(\int_{-\infty}^{-\lambda-w} v dF_V(v) + \int_{\lambda-w}^{\infty} v dF_V(v) \right),$$

and F_V is the distribution of V . It is straightforward to show that $g(0) = 0$, $g(w) \geq 0$ for $w \geq 0$, and g is an odd function. Now consider $u > 0$. We have

$$\int \phi(w - u) g(w) dw = \int_0^{\infty} \left(\phi(w - u) - \phi(w + u) \right) g(w) dw \geq 0.$$

When $u < 0$, the same method can be used to show that

$$\mathbb{E}\left[V \mid |u + V + W| > \lambda\right] \leq 0.$$

As a result, we always have

$$u\mathbb{E}\left[V \mid |u + V + W| > \lambda\right] \geq 0,$$

which implies that

$$\mathbb{E}\left[UV \mid |U + V + W| > \lambda\right] \geq 0.$$

Case 2: $\sigma = 0$

Now assume $\sigma = 0$, which means $W = 0$ with probability 1. We have

$$\mathbb{E}\left[V \mid |u + V| > \lambda\right] \propto \mathbb{E}\left[V \mathbf{1}(|u + V| > \lambda)\right] = \int_{-\infty}^{-\lambda-u} v f_V(v) dv + \int_{\lambda-u}^{\infty} v f_V(v) dv = g(u).$$

By our discussion above, it is easy to see that $ug(u) \geq 0$, which implies

$$u\mathbb{E}\left[V \mid |u + V| > \lambda\right] \geq 0,$$

and therefore

$$\mathbb{E}\left[UV \mid |U + V| > \lambda\right] \geq 0.$$

■

6.6 Derivation of Equations (1) and (2)

Proof. To show the decomposition of our estimators, we start from the following

$$\begin{aligned} \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{Y_j} \hat{\beta}_{X_j, \text{bc}}}{\sigma_{Y_j}^2} &= \sum_{j \in \mathcal{S}_x} \frac{(\theta \beta_{X_j} + \tau_Y \beta_{M_j} + v_j + \alpha_j) \hat{\beta}_{X_j, \text{bc}}}{\sigma_{Y_j}^2} \\ &= \sum_{j \in \mathcal{S}_x} \frac{(\theta \hat{\beta}_{X_j, \text{bc}} + \tau_Y \hat{\beta}_{M_j, \text{bc}} - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}} + v_j + \alpha_j) \hat{\beta}_{X_j, \text{bc}}}{\sigma_{Y_j}^2} \\ &= \theta \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j}}{\sigma_{Y_j}^2} + \tau_Y \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{M_j, \text{bc}} \hat{\beta}_{X_j, \text{bc}}}{\sigma_{Y_j}^2} + \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}} (v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + \theta \hat{\varsigma}_{X_j}}{\sigma_{Y_j}^2}. \end{aligned}$$

Similarly, we have

$$\begin{aligned}
\sum_{j \in \mathcal{S}_m} \frac{\hat{\beta}_{Y_j} \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} &= \sum_{j \in \mathcal{S}_m} \frac{(\theta \beta_{X_j} + \tau_Y \beta_{M_j} + v_j + \alpha_j) \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} \\
&= \sum_{j \in \mathcal{S}_m} \frac{(\theta \hat{\beta}_{X_j, \text{bc}} + \tau_Y \hat{\beta}_{M_j, \text{bc}} - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}} + v_j + \alpha_j) \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} \\
&= \theta \sum_{j \in \mathcal{S}_m} \frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}}}{\sigma_{Y_j}^2} + \tau_Y \sum_{j \in \mathcal{S}_m} \frac{\hat{\beta}_{M_j, \text{bc}}^2 - \hat{\varsigma}_{M_j}}{\sigma_{Y_j}^2} + \sum_{j \in \mathcal{S}_m} \frac{\hat{\beta}_{M_j, \text{bc}} (v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + \tau_Y \hat{\varsigma}_{M_j}}{\sigma_{Y_j}^2},
\end{aligned}$$

and

$$\begin{aligned}
\sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{M_j, \text{bc}} \hat{\beta}_{X_j, \text{bc}}}{\sigma_{M_j}^2} &= \sum_{j \in \mathcal{S}_x} \frac{(\tau_X \beta_{X_j} + w_{j, \text{bc}} + \delta_j) \hat{\beta}_{X_j, \text{bc}}}{\sigma_{M_j}^2} \\
&= \sum_{j \in \mathcal{S}_x} \frac{(\tau_X \hat{\beta}_{X_j, \text{bc}} - \tau_X u_{j, \text{bc}} + w_{j, \text{bc}} + \delta_j) \hat{\beta}_{X_j, \text{bc}}}{\sigma_{M_j}^2} \\
&= \tau_X \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}}^2 - \hat{\varsigma}_{X_j}}{\sigma_{M_j}^2} + \sum_{j \in \mathcal{S}_x} \frac{\hat{\beta}_{X_j, \text{bc}} (w_{j, \text{bc}} + \delta_j - \tau_X u_{j, \text{bc}}) + \tau_X \hat{\varsigma}_{X_j}}{\sigma_{M_j}^2}.
\end{aligned}$$

The above justifies the decomposition of our estimators. ■

6.7 Proof of Lemma 6

Proof. We illustrate the variance computation for $u_{j, \theta}$, and note that the same techniques apply to the two other variances. To start, since instrument selection is carried out in two independent GWAS, we have

$$\begin{aligned}
&\mathbb{V} \left[\sum_{j=1}^p u_{j, \theta} \middle| \mathcal{S}_x, \mathcal{S}_m \right] \\
&= \sum_{j \in \mathcal{S}_x} \mathbb{V} \left[\frac{\beta_{X_j} (v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + u_{j, \text{bc}} (v_j + \alpha_j - \tau_Y w_{j, \text{bc}}) + \theta (\hat{\varsigma}_{X_j} - u_{j, \text{bc}}^2)}{\sigma_{Y_j}^2} \middle| j \in \mathcal{S}_x \right].
\end{aligned}$$

While the above may seem complicated, we further notice that v_j and α_j are from a third GWAS and hence their variances are not affected by the selection event. In addition, as we have discussed above, the variance of $u_{j, \text{bc}}$ and $w_{j, \text{bc}}$ are bounded by the order ν^2 . Finally, the proof of Lemma 2 employs variance calculations of $\hat{\varsigma}_{X_j}$. As a result,

$$\sum_{j \in \mathcal{S}_x} \mathbb{V} \left[\frac{\beta_{X_j} (v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + u_{j, \text{bc}} (v_j + \alpha_j - \tau_Y w_{j, \text{bc}}) + \theta (\hat{\varsigma}_{X_j} - u_{j, \text{bc}}^2)}{\sigma_{Y_j}^2} \middle| j \in \mathcal{S}_x \right]$$

$$\lesssim \frac{1}{\nu^4} \sum_{j \in \mathcal{S}_x} \left(\beta_{X_j}^2 \nu^2 (\kappa + 1) + \nu^4 (\kappa + 1) + \nu^4 \right) \lesssim_p \mathfrak{p} \kappa (\kappa + 1) + \mathfrak{p} (\kappa + 1) + \mathfrak{p} \asymp \mathfrak{p} \kappa^2.$$

■

6.8 Proof of Lemma 7

Proof. The proof is very similar to that of Lemma 6. We again illustrate with $u_{j,\theta}$. To start, we notice that v_j and α_j are independent of the instrument selection events, and hence their moments are not affected. Fourth moments of $u_{j,\text{bc}}$ and $w_{j,\text{bc}}$ are available in Lemma 3. Next, we may bound the fourth moments of $\widehat{\varsigma}_{X_j}$ as

$$\begin{aligned} & \mathbb{E} \left[\left(\widehat{\varsigma}_{X_j} - \varsigma_{X_j} \right)^4 \middle| j \in \mathcal{S}_x \right] \\ & \lesssim \nu^8 \mathbb{E} \left[\left(\mathbb{V}[Z_j | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] - \mathbb{E}[\mathbb{V}[Z_j | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] | j \in \mathcal{S}_x] \right)^4 \middle| j \in \mathcal{S}_x \right] \\ & \lesssim \nu^8 \mathbb{E} \left[\left(\mathbb{V}[Z_j | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] \right)^4 \middle| j \in \mathcal{S}_x \right] \\ & \leq \nu^8 \mathbb{E} \left[\left(\mathbb{E}[Z_j^8 | \widehat{\beta}_{X_j}, j \in \mathcal{S}_x] \right) \middle| j \in \mathcal{S}_x \right] \lesssim \nu^8. \end{aligned}$$

Therefore,

$$\begin{aligned} & \sum_{j \in \mathcal{S}_x} \mathbb{E} \left[\left(\frac{\beta_{X_j} \left(v_j + \alpha_j - \theta u_{j,\text{bc}} - \tau_Y w_{j,\text{bc}} \right) + u_{j,\text{bc}} \left(v_j + \alpha_j - \tau_Y w_{j,\text{bc}} \right) + \theta \left(\widehat{\varsigma}_{X_j} - u_{j,\text{bc}}^2 \right)}{\sigma_{Y_j}^2} \right)^4 \middle| j \in \mathcal{S}_x \right] \\ & \lesssim \frac{1}{\nu^8} \sum_{j \in \mathcal{S}_x} \left(\beta_{X_j}^4 \nu^4 (\kappa^2 + 1) + \nu^8 (\kappa^2 + 1) + \nu^8 \right). \end{aligned}$$

By Assumption 5,

$$\sum_{j \in \mathcal{S}_x} \beta_{X_j}^4 \prec_p (\mathfrak{p} \kappa \nu^2)^2,$$

which closes the proof. ■

6.9 Proof of Lemma 8

Proof. This result follows directly from Lemma 7, Assumption 6, and the multivariate Lindeberg-Feller central limit theorem. ■

6.10 Derivation of Equation (3)

Proof. We start from

$$\widehat{\beta}_{X_j, \text{bc}} \left(v_j + \alpha_j - \theta u_{j,\text{bc}} - \tau_Y w_{j,\text{bc}} \right) + \theta \widehat{\varsigma}_{X_j}$$

$$\begin{aligned}
&= \widehat{\beta}_{X_j, \text{bc}} \left(\widehat{\beta}_{Y_j} - \beta_{Y_j} + \alpha_j - \theta(\widehat{\beta}_{X_j, \text{bc}} - \beta_{X_j}) - \tau_Y(\widehat{\beta}_{M_j, \text{bc}} - \beta_{M_j}) \right) + \theta \widehat{\varsigma}_{X_j} \\
&= \widehat{\beta}_{X_j, \text{bc}} \left(\widehat{\beta}_{Y_j} - \tau_Y \widehat{\beta}_{M_j, \text{bc}} - \underbrace{(\beta_{Y_j} - \alpha_j - \theta \beta_{X_j} - \tau_Y \beta_{M_j})}_{=0} \right) + \theta (\widehat{\varsigma}_{X_j} - \widehat{\beta}_{X_j, \text{bc}}^2) \\
&= \widehat{\beta}_{X_j, \text{bc}} (\widehat{\beta}_{Y_j} - \tau_Y \widehat{\beta}_{M_j, \text{bc}}) + \theta (\widehat{\varsigma}_{X_j} - \widehat{\beta}_{X_j, \text{bc}}^2).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
&\widehat{\beta}_{M_j, \text{bc}} (v_j + \alpha_j - \theta u_{j, \text{bc}} - \tau_Y w_{j, \text{bc}}) + \tau_Y \widehat{\varsigma}_{M_j} \\
&= \widehat{\beta}_{M_j, \text{bc}} (\widehat{\beta}_{Y_j} - \theta \widehat{\beta}_{X_j, \text{bc}}) + \tau_Y (\widehat{\varsigma}_{M_j} - \widehat{\beta}_{M_j, \text{bc}}^2),
\end{aligned}$$

and

$$\begin{aligned}
&\widehat{\beta}_{X_j, \text{bc}} (w_{j, \text{bc}} + \delta_j - \tau_X u_{j, \text{bc}}) + \tau_X \widehat{\varsigma}_{X_j} \\
&= \widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{M_j, \text{bc}} + \tau_X (\widehat{\varsigma}_{X_j} - \widehat{\beta}_{X_j, \text{bc}}^2).
\end{aligned}$$

■

6.11 Proof of Lemma 9

Proof. To illustrate the idea, consider $\widehat{u}_{j, \theta}$, which can be written as

$$\widehat{u}_{j, \theta} = u_{j, \theta} + \underbrace{\mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \frac{\widehat{\beta}_{X_j, \text{bc}} \widehat{\beta}_{M_j, \text{bc}} (\tau_Y - \widehat{\tau}_Y) + (\widehat{\theta} - \theta) (\widehat{\varsigma}_{X_j} - \widehat{\beta}_{X_j, \text{bc}}^2)}{\sigma_{Y_j}^2}}_{=: \check{u}_{j, \theta}}.$$

Using the higher moment calculations in Lemmas 2, 3, the convergence in Lemma 4, and the convergence rate of our estimates, one has

$$\begin{aligned}
\sum_{j=1}^p \check{u}_{j, \theta}^2 &\lesssim_{\mathbf{p}} \frac{1}{\mathbf{p}} \left[\sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \frac{\widehat{\beta}_{X_j, \text{bc}}^2 \widehat{\beta}_{M_j, \text{bc}}^2}{\sigma_{Y_j}^4} + \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \frac{(\widehat{\varsigma}_{X_j} - \widehat{\beta}_{X_j, \text{bc}}^2)^2}{\sigma_{Y_j}^4} \right] \\
&\prec_{\mathbf{p}} \frac{1}{\mathbf{p}} [\mathbf{p}^2 \kappa^2 + \mathbf{p} \kappa + \mathbf{p}] \lesssim \mathbf{p} \kappa^2.
\end{aligned}$$

In other words,

$$\sum_{j=1}^p \check{u}_{j, \theta}^2 = o_{\mathbf{p}} (\mathbf{p} \kappa^2).$$

It is worth mentioning that the leading term stems from the observation

$$\sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x)} \frac{\beta_{X_j}^2 \beta_{M_j}^2}{\sigma_{Y_j}^4} \prec_p \mathbf{p}^2 \kappa^2.$$

Applying similar decomposition to $\hat{u}_{j,\tau_Y}$ and $\hat{u}_{j,\tau_X}$, we define

$$\begin{aligned} \hat{u}_{j,\tau_Y} &= u_{j,\tau_Y} + \mathbf{1}_{(j \in \mathcal{S}_m)} \cdot \underbrace{\frac{\hat{\beta}_{X_j, \text{bc}} \hat{\beta}_{M_j, \text{bc}} (\theta - \hat{\theta}) + (\hat{\tau}_Y - \tau_Y) (\hat{\varsigma}_{M_j} - \hat{\beta}_{M_j, \text{bc}}^2)}{\sigma_{Y_j}^2}}_{=: \check{u}_{j,\tau_Y}}, \\ \hat{u}_{j,\tau_X} &= u_{j,\tau_X} + \mathbf{1}_{(j \in \mathcal{S}_x)} \cdot \underbrace{\frac{(\hat{\tau}_X - \tau_X) (\hat{\varsigma}_{X_j} - \hat{\beta}_{X_j, \text{bc}}^2)}{\sigma_{M_j}^2}}_{=: \check{u}_{j,\tau_X}}, \end{aligned}$$

and we conclude that

$$\begin{aligned} \sum_{j=1}^p \hat{U}_j \hat{U}_j^\top &= \sum_{j=1}^p U_j U_j^\top + \sum_{j=1}^p U_j \check{U}_j^\top + \sum_{j=1}^p \check{U}_j U_j^\top + \underbrace{\sum_{j=1}^p \check{U}_j \check{U}_j^\top}_{= o_p(\mathbf{p} \kappa^2)}. \end{aligned}$$

Here, $\check{U}_j$ is the vector formed by $\check{u}_{j,\theta}$, $\check{u}_{j,\tau_Y}$, and $\check{u}_{j,\tau_X}$. It is straightforward to verify that

$$\mathbf{U}^{-1} \left(\sum_{j=1}^p U_j U_j^\top \right) \xrightarrow{\mathbf{P}} \mathbf{I},$$

and as a byproduct,

$$\sum_{j=1}^p U_j U_j^\top \asymp_p \mathbf{p} \kappa^2.$$

Finally, we can apply Cauchy-Schwartz inequality to bound the interaction terms:

$$\begin{aligned} \sum_{j=1}^p \hat{U}_j \hat{U}_j^\top &= \sum_{j=1}^p U_j U_j^\top + \underbrace{\sum_{j=1}^p U_j \check{U}_j^\top + \sum_{j=1}^p \check{U}_j U_j^\top + \sum_{j=1}^p \check{U}_j \check{U}_j^\top}_{= o_p(\mathbf{p} \kappa^2)}, \end{aligned}$$

which closes the proof. ■

6.12 Proof of Theorem 1

Proof. The theorem follows from the convergence of $\hat{M}$ (Lemma 5), the convergence of $\hat{U}$ (Lemma 9), and the asymptotic normality (Lemma 8). ■

6.13 Proof of Theorem 2

Proof. With a slight abuse of notation, we denote the inverse matrix in our estimator by $\hat{M}^{-1}$, and that in the weak IV bias corrected version of MVMR approach (DMVMR) by $\hat{M}_{\text{DMVMR}}^{-1}$. (We recall that the estimation of τ_X is ignored in this section, and hence they are both 2-by-2 matrices.) Following the proof strategy of Lemmas 4 and 5, one can show that

$$M^{-1}\hat{M} \xrightarrow{P} I, \quad M^{-1}\hat{M}_{\text{DMVMR}} \xrightarrow{P} I,$$

where

$$M = \begin{bmatrix} \sum_{j \in \mathcal{S}_x^*} \frac{\beta_{X_j}^2}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_x^* \cap \mathcal{S}_m^*} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} \\ \sum_{j \in \mathcal{S}_x^* \cap \mathcal{S}_m^*} \frac{\beta_{X_j} \beta_{M_j}}{\sigma_{Y_j}^2} & \sum_{j \in \mathcal{S}_m^*} \frac{\beta_{M_j}^2}{\sigma_{Y_j}^2} \end{bmatrix}.$$

As a result, one only needs to compare the “middle matrix” for asymptotic efficiency comparisons.

In an oracle scenario, from earlier discussions, we have the representation of our estimator as

$$\begin{bmatrix} \hat{\theta}^* - \theta \\ \hat{\tau}_Y^* - \tau_Y \end{bmatrix} = \hat{M}^{-1} \sum_{j=1}^p U_j \approx M^{-1} \sum_{j=1}^p U_j,$$

where, with slight abuse of notation,

$$U_j = \begin{bmatrix} u_{j,\theta} \\ u_{j,\tau_Y} \end{bmatrix} := \begin{bmatrix} \mathbf{1}_{(j \in \mathcal{S}_x^*)} \cdot \frac{\beta_{X_j} (v_j + \alpha_j - \theta u_j - \tau_Y w_j) + u_j (v_j + \alpha_j - \tau_Y w_j) + \theta (\sigma_{X_j}^2 - u_j^2)}{\sigma_{Y_j}^2} \\ \mathbf{1}_{(j \in \mathcal{S}_m^*)} \cdot \frac{\beta_{M_j} (v_j + \alpha_j - \theta u_j - \tau_Y w_j) + w_j (v_j + \alpha_j - \theta u_j) + \tau_Y (\sigma_{M_j}^2 - w_j^2)}{\sigma_{Y_j}^2} \end{bmatrix}.$$

For the DMVMR estimator, notice that

$$\begin{bmatrix} \hat{\theta}_{\text{DMVMR}}^* - \theta \\ \hat{\tau}_{Y,\text{DMVMR}}^* - \tau_Y \end{bmatrix} = \hat{M}_{\text{DMVMR}}^{-1} \sum_{j=1}^p U_{j,\text{DMVMR}} \approx M^{-1} \sum_{j=1}^p U_{j,\text{DMVMR}},$$

where

$$U_{j,\text{DMVMR}} = \begin{bmatrix} u_{j,\theta,\text{DMVMR}} \\ u_{j,\tau_Y,\text{DMVMR}} \end{bmatrix} := \begin{bmatrix} \mathbf{1}_{(j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*)} \cdot \frac{\beta_{X_j} \left(v_j + \alpha_j - \theta u_j - \tau_Y w_j \right) + u_j \left(v_j + \alpha_j - \tau_Y w_j \right) + \theta \left(\sigma_{X_j}^2 - u_j^2 \right)}{\sigma_{Y_j}^2} \\ \mathbf{1}_{(j \in \mathcal{S}_x^* \cup \mathcal{S}_m^*)} \cdot \frac{\beta_{M_j} \left(v_j + \alpha_j - \theta u_j - \tau_Y w_j \right) + w_j \left(v_j + \alpha_j - \theta u_j \right) + \tau_Y \left(\sigma_{M_j}^2 - w_j^2 \right)}{\sigma_{Y_j}^2} \end{bmatrix}.$$

In other words, we can also write

$$U_{j,\text{DMVMR}} = U_j + \tilde{U}_j,$$

and

$$\tilde{U}_j = \begin{bmatrix} \tilde{u}_{j,\theta} \\ \tilde{u}_{j,\tau_Y} \end{bmatrix} := \begin{bmatrix} \mathbf{1}_{(j \in \mathcal{S}_m^* \setminus \mathcal{S}_x^*)} \cdot \frac{u_j \left(v_j + \alpha_j - \tau_Y w_j \right) + \theta \left(\sigma_{X_j}^2 - u_j^2 \right)}{\sigma_{Y_j}^2} \\ \mathbf{1}_{(j \in \mathcal{S}_x^* \setminus \mathcal{S}_m^*)} \cdot \frac{w_j \left(v_j + \alpha_j - \theta u_j \right) + \tau_Y \left(\sigma_{M_j}^2 - w_j^2 \right)}{\sigma_{Y_j}^2} \end{bmatrix}.$$

We make some additional observations:

$$\mathbb{E}[u_{j,\theta} \tilde{u}_{j,\theta}] = 0, \quad \mathbb{E}[u_{j,\tau_Y} \tilde{u}_{j,\tau_Y}] = 0,$$

and

$$\begin{aligned} \mathbb{E}[u_{j,\theta} \tilde{u}_{j,\tau_Y}] &= \mathbf{1}_{(j \in \mathcal{S}_x^* \setminus \mathcal{S}_m^*)} \cdot \mathbb{E} \left[\frac{\beta_{X_j} \left(v_j + \alpha_j - \theta u_j - \tau_Y w_j \right) + u_j \left(v_j + \alpha_j - \tau_Y w_j \right) + \theta \left(\sigma_{X_j}^2 - u_j^2 \right)}{\sigma_{Y_j}^2} \right. \\ &\quad \times \left. \frac{w_j \left(v_j + \alpha_j - \theta u_j \right) + \tau_Y \left(\sigma_{M_j}^2 - w_j^2 \right)}{\sigma_{Y_j}^2} \right] \\ &= \mathbf{1}_{(j \in \mathcal{S}_x^* \setminus \mathcal{S}_m^*)} \theta \tau_Y \frac{\sigma_{X_j}^2 \sigma_{M_j}^2}{\sigma_{Y_j}^4}, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[u_{j,\tau_Y} \tilde{u}_{j,\theta}] &= \mathbf{1}_{(j \in \mathcal{S}_m^* \setminus \mathcal{S}_x^*)} \cdot \mathbb{E} \left[\frac{\beta_{M_j} \left(v_j + \alpha_j - \theta u_j - \tau_Y w_j \right) + w_j \left(v_j + \alpha_j - \theta u_j \right) + \tau_Y \left(\sigma_{M_j}^2 - w_j^2 \right)}{\sigma_{Y_j}^2} \right. \\ &\quad \times \left. \frac{u_j \left(v_j + \alpha_j - \tau_Y w_j \right) + \theta \left(\sigma_{X_j}^2 - u_j^2 \right)}{\sigma_{Y_j}^2} \right] \\ &= \mathbf{1}_{(j \in \mathcal{S}_m^* \setminus \mathcal{S}_x^*)} \theta \tau_Y \frac{\sigma_{X_j}^2 \sigma_{M_j}^2}{\sigma_{Y_j}^4}. \end{aligned}$$

Finally, we compute

$$\begin{aligned}\mathbb{V}[\tilde{u}_{j,\theta}] &= \mathbf{1}_{(j \in \mathcal{S}_m^* \setminus \mathcal{S}_x^*)} \frac{\sigma_{X_j}^2(\sigma_{Y_j}^2 + \mathbb{V}[\alpha_j] + \tau_Y^2 \sigma_{M_j}^2) + 2\theta^2 \sigma_{X_j}^4}{\sigma_{Y_j}^4} \\ \mathbb{V}[\tilde{u}_{j,\tau_Y}] &= \mathbf{1}_{(j \in \mathcal{S}_x^* \setminus \mathcal{S}_m^*)} \frac{\sigma_{M_j}^2(\sigma_{Y_j}^2 + \mathbb{V}[\alpha_j] + \theta^2 \sigma_{X_j}^2) + 2\tau_Y^2 \sigma_{M_j}^4}{\sigma_{Y_j}^4} \\ \mathbb{E}[\tilde{u}_{j,\theta} \tilde{u}_{j,\tau_Y}] &= 0.\end{aligned}$$

Therefore, the variance matrices satisfy

$$\begin{aligned}\mathbb{V}\left[\sum_{j=1}^p U_{j,\text{MVMR}}\right] &= \mathbb{V}\left[\sum_{j=1}^p U_j\right] + \sum_{j=1}^p \left(\mathbb{E}[U_j \tilde{U}_j^\top] + \mathbb{E}[\tilde{U}_j U_j^\top]\right) + \mathbb{V}\left[\sum_{j=1}^p \tilde{U}_j\right] \\ &= \mathbb{V}\left[\sum_{j=1}^p U_j\right] \\ &\quad + \sum_{j=1}^p \begin{bmatrix} \mathbf{1}_{(j \in \mathcal{S}_m^* \setminus \mathcal{S}_x^*)} \frac{\sigma_{X_j}^2(\sigma_{Y_j}^2 + \mathbb{V}[\alpha_j] + \tau_Y^2 \sigma_{M_j}^2) + 2\theta^2 \sigma_{X_j}^4}{\sigma_{Y_j}^4} & \mathbf{1}_{(j \in \mathcal{S}_x^* \Delta \mathcal{S}_m^*)} \theta \tau_Y \frac{\sigma_{X_j}^2 \sigma_{M_j}^2}{\sigma_{Y_j}^4} \\ \mathbf{1}_{(j \in \mathcal{S}_x^* \Delta \mathcal{S}_m^*)} \theta \tau_Y \frac{\sigma_{X_j}^2 \sigma_{M_j}^2}{\sigma_{Y_j}^4} & \mathbf{1}_{(j \in \mathcal{S}_x^* \setminus \mathcal{S}_m^*)} \frac{\sigma_{M_j}^2(\sigma_{Y_j}^2 + \mathbb{V}[\alpha_j] + \theta^2 \sigma_{X_j}^2) + 2\tau_Y^2 \sigma_{M_j}^4}{\sigma_{Y_j}^4} \end{bmatrix}.\end{aligned}$$

Here, Δ denotes the symmetric difference of two sets. Recall that we assumed $\sigma_{X_j}^2 = c_X \sigma_{Y_j}^2$ and $\sigma_{M_j}^2 = c_M \sigma_{Y_j}^2$. Then

$$\begin{aligned}\mathbb{V}\left[\sum_{j=1}^p U_{j,\text{DMVMR}}\right] &= \mathbb{V}\left[\sum_{j=1}^p U_j\right] \\ &\quad + \begin{bmatrix} p_{m \setminus x}(c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M) & p_{m \Delta x} \theta \tau_Y c_X c_M \\ p_{m \Delta x} \theta \tau_Y c_X c_M & p_{x \setminus m}(c_M + \theta^2 c_X c_M + 2\tau_Y^2 c_M^2) \end{bmatrix} \quad (\text{I}) \\ &\quad + \begin{bmatrix} \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_m^* \setminus \mathcal{S}_x^*)} \frac{c_X \mathbb{V}[\alpha_j]}{\sigma_{Y_j}^2} & 0 \\ 0 & \sum_{j=1}^p \mathbf{1}_{(j \in \mathcal{S}_x^* \setminus \mathcal{S}_m^*)} \frac{c_M \mathbb{V}[\alpha_j]}{\sigma_{Y_j}^2} \end{bmatrix}. \quad (\text{II})\end{aligned}$$

In the above, $p_{m \setminus x}$ is the cardinality of $\mathcal{S}_m^* \setminus \mathcal{S}_x^*$, $p_{x \setminus m}$ is the cardinality of $\mathcal{S}_x^* \setminus \mathcal{S}_m^*$, and $p_{m \Delta x}$ is the cardinality of the symmetric difference $\mathcal{S}_m^* \Delta \mathcal{S}_x^*$. Notice that $p_{m \Delta x} = p_{m \setminus x} + p_{x \setminus m}$.

We observe that (II) is positive semi-definite. (I) is positive definite if $\theta \cdot \tau_Y = 0$. In what follows, we assume $\theta \cdot \tau_Y \neq 0$.

To simplify notation, we will set $\tilde{p} = p_{x \setminus m} / p_{m \setminus x}$. Now consider

$$\begin{aligned}&\left(c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M\right) \left(\tilde{p}(c_M + \theta^2 c_X c_M + 2\tau_Y^2 c_M^2)\right) - (1 + \tilde{p})^2 \theta^2 \tau_Y^2 c_X^2 c_M^2 \\ &\geq \tilde{p} \left(2\theta^2 c_X^2 + \tau_Y^2 c_X c_M\right) \left(\theta^2 c_X c_M + 2\tau_Y^2 c_M^2\right) - (1 + \tilde{p})^2 \theta^2 \tau_Y^2 c_X^2 c_M^2 \\ &= 2\tilde{p} c_X^3 c_M \theta^4 + 2\tilde{p} c_X c_M^3 \tau_Y^4 + \left(5\tilde{p} - (1 + \tilde{p})^2\right) c_X^2 c_M^2 \theta^2 \tau_Y^2.\end{aligned}$$

The above will be strictly positive for all θ and τ_Y as long as

$$\left(5\tilde{p} - (1 + \tilde{p})^2\right)^2 - 16\tilde{p}^2 < 0,$$

which holds when $\tilde{p} \in [1/6.85, 6.85]$, which can be found by numerically solving the polynomial equation. ■

6.14 Proof of Theorem 3

Proof. Now consider the scenario: $\mathcal{S}_m \supsetneq \mathcal{S}_x$. Then we can write $\mathbf{M}$ as

$$\mathbf{M} = \begin{bmatrix} p_x \kappa_x & \tau_X p_x \kappa_x \\ \tau_X p_x \kappa_x & \tau_X^2 p_x \kappa_x + p_m \kappa_m^\delta \end{bmatrix} + O_p(\sqrt{p\kappa^2}).$$

Now consider the estimator $\hat{\theta}^*$. Its asymptotic efficiency gain (relative to the DMVMR approach) can be determined by the sign of the following:

$$\begin{aligned} & \begin{bmatrix} \tau_X^2 p_x \kappa_x + p_m \kappa_m^\delta & -\tau_X p_x \kappa_x \end{bmatrix} \begin{bmatrix} c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M & \theta \tau_Y c_X c_M \\ \theta \tau_Y c_X c_M & 0 \end{bmatrix} \begin{bmatrix} \tau_X^2 p_x \kappa_x + p_m \kappa_m^\delta \\ -\tau_X p_x \kappa_x \end{bmatrix} \\ &= \left(\tau_X^2 p_x \kappa_x + p_m \kappa_m^\delta\right)^2 \left(c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M\right) - 2 \left(\tau_X^2 p_x \kappa_x + p_m \kappa_m^\delta\right) \left(\tau_X p_x \kappa_x\right) \left(\theta \tau_Y c_X c_M\right), \end{aligned}$$

or equivalently, the sign of

$$\left(\tau_X^2 p_x \kappa_x + p_m \kappa_m^\delta\right) \left(c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M\right) - 2 \left(\tau_X p_x \kappa_x\right) \left(\theta \tau_Y c_X c_M\right).$$

Assuming $c_X = c_M = 1$, we notice that

$$\frac{2|\theta \tau_Y c_X c_M|}{c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M} \leq \frac{\sqrt{2}}{2}.$$

To see this, we write $\tau_Y = a\theta$ for some $a \geq 0$. Then

$$\frac{2|\theta \tau_Y c_X c_M|}{c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M} = \frac{2a\theta^2}{1 + (2 + a^2)\theta^2} \leq \frac{2a}{2 + a^2} \leq \frac{\sqrt{2}}{2}.$$

Therefore, it suffices to require

$$\tau_X^2 p_x \kappa_x + p_m \kappa_m^\delta \pm \frac{\sqrt{2}}{2} \tau_X p_x \kappa_x > 0$$

for all τ_X . A sufficient condition would be that

$$\frac{p_m \kappa_m^\delta}{p_x \kappa_x} > \frac{1}{8}.$$

Now consider the estimation of τ_Y . Its efficiency gain can be determined from the sign of

$$\begin{aligned} & \begin{bmatrix} -\tau_X p_{\mathbf{x}} \kappa_{\mathbf{x}} & p_{\mathbf{x}} \kappa_{\mathbf{x}} \end{bmatrix} \begin{bmatrix} c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M & \theta \tau_Y c_X c_M \\ \theta \tau_Y c_X c_M & 0 \end{bmatrix} \begin{bmatrix} -\tau_X p_{\mathbf{x}} \kappa_{\mathbf{x}} \\ p_{\mathbf{x}} \kappa_{\mathbf{x}} \end{bmatrix} \\ &= \left(\tau_X p_{\mathbf{x}} \kappa_{\mathbf{x}} \right)^2 \left(c_X + 2\theta^2 c_X^2 + \tau_Y^2 c_X c_M \right) - 2 \left(\tau_X p_{\mathbf{x}} \kappa_{\mathbf{x}} \right) \left(p_{\mathbf{x}} \kappa_{\mathbf{x}} \right) \left(\theta \tau_Y c_X c_M \right), \end{aligned}$$

or equivalently, the sign of

$$\tau_X^2 \left(1 + 2\theta^2 + \tau_Y^2 \right) - 2\tau_X \theta \tau_Y.$$

■

To understand the function $f(\tau_X, \theta, \tau_Y) = \tau_X^2 \left(1 + 2\theta^2 + \tau_Y^2 \right) - 2\tau_X \theta \tau_Y$, we plot the surface of $(\theta, \tau_Y), \theta \in [-1, 1], \tau_Y \in [-1, 1]$ given different τ_X . From Figure 2, we can see that in most combination of (θ, τ_X, τ_Y) , $f(\cdot)$ takes positive values, which means Assumption 8 (ii) is satisfied.

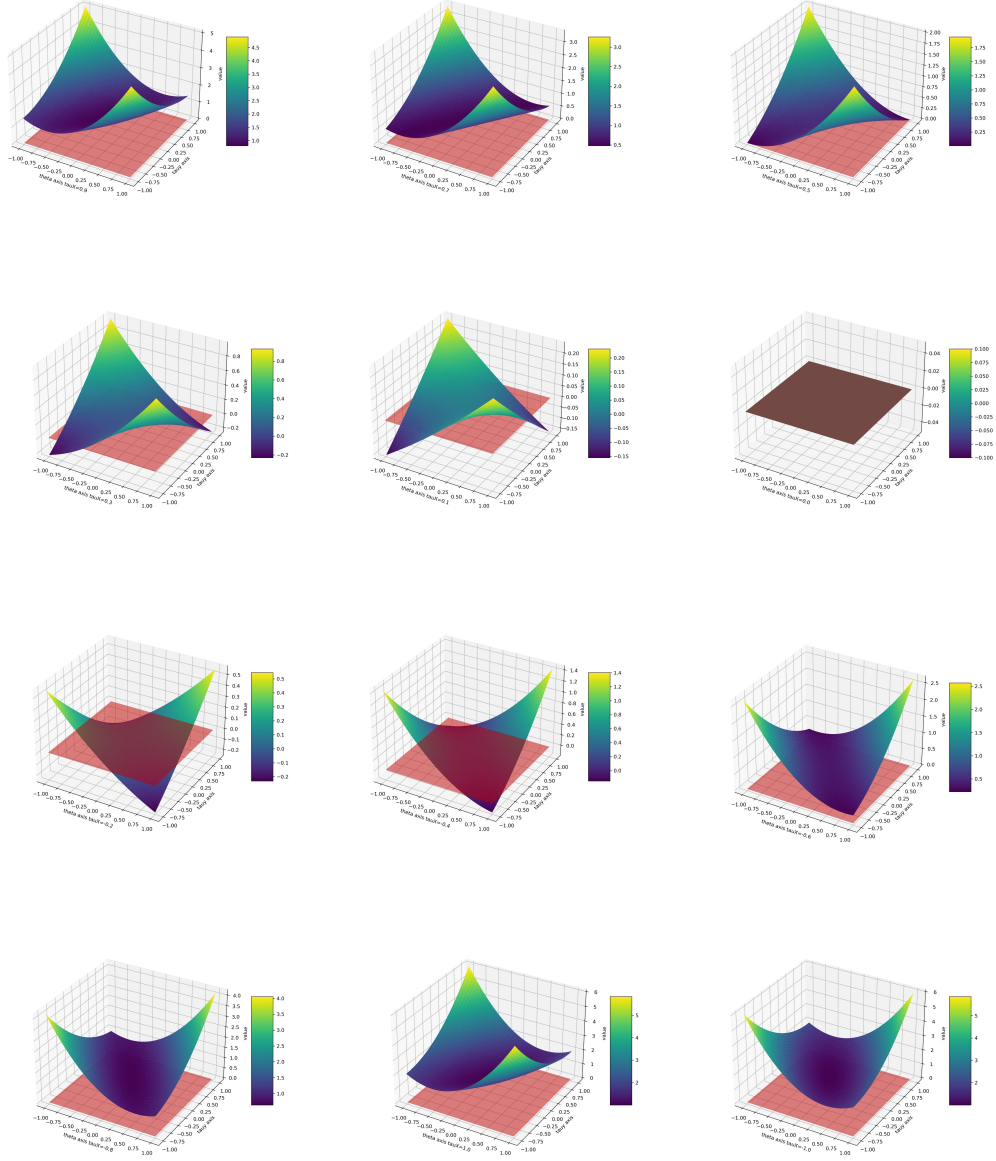


Figure 2: The value of function $f(\tau_X, \theta, \tau_Y) = \tau_X^2(1 + 2\theta^2 + \tau_Y^2) - 2\tau_X\theta\tau_Y$, when $\theta \in [-1, 1]$, $\tau_Y \in [-1, 1]$ given different τ_X . X-axis is the value of θ between -1 and 1. Y-axis is the value of τ_Y between -1 and 1. The red hyperplane is the hyperplane when the Z-axis value is 0.

7 Simulation

7.1 Overlapping Index Set Construction

We illustrate how to construct index sets $\mathcal{S}_x^*$ and $\mathcal{S}_\delta^*$ under DGP 1, 2, 3 here.

Algorithm 1 Construction of index sets $\mathcal{S}_x^*$ and $\mathcal{S}_\delta^*$

Input: π_x, π_δ , the total number of SNPs p , and choice of DGPs (DGP 1 / DGP 2 (i) / DGP 2 (ii) / DGP 3 (i) / DGP 3 (ii)).

Output: $\mathcal{S}_x^*$ and $\mathcal{S}_\delta^*$.

- Define $\mathcal{S}_x^* = \emptyset, \mathcal{S}_\delta^* = \emptyset$.
- 2: Randomly sample $p\pi_x$ integers from 1 to p , and assign these integers to $\mathcal{S}_x^*$.
Define $\mathcal{S}_- = \{1, \dots, p\} \setminus \mathcal{S}_x^*$.
 - 4: **if** DGP 2 (i) or DGP 3 (i) **then**
Randomly sample π_δ integers from $\mathcal{S}_-$ and assign these integers to $\mathcal{S}_\delta^*$.
 - 6: **if** DGP 2 (ii) or DGP 3 (ii) **then**
Randomly sample $\frac{1}{2}p\pi_x$ integers from $\mathcal{S}_x^*$ and assign these integers to $\mathcal{S}_\delta^*$.
 - 8: Randomly sample $p\pi_\delta - \frac{1}{2}p\pi_x$ integers from $\mathcal{S}_-$ and combine $\mathcal{S}_\delta^*$ with these integers.
 - if** DGP 1 **then**
 - 10: Assign $\mathcal{S}_\delta^* = \mathcal{S}_x^*$.
if $p\pi_\delta > p\pi_x$ **then**
 - 12: Randomly sample $p\pi_\delta - p\pi_x$ integers from $\mathcal{S}_-$ and combine $\mathcal{S}_\delta^*$ with these integers.
-

7.2 Index Construction of $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$ under Oracle Setting

We demonstrate the way to generate $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$ under different ratios of $\frac{|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|}{|\mathcal{S}_m^* \setminus \mathcal{S}_x^*|}$ in the following Algorithm.

Algorithm 2 Construction of index sets $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$

Input: $\pi_{x \cap m}, \pi_{m \cap x}, \pi_{x \setminus m}$ the total number of SNPs p , and the ratio $\tilde{p} = \frac{|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|}{|\mathcal{S}_m^* \setminus \mathcal{S}_x^*|}$

Output: $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$.

- Define $\mathcal{S}_x^* = \emptyset, \mathcal{S}_m^* = \emptyset$.
- 2: Randomly sample $p\pi_{x \setminus m}$ integers from 1 to p , and assign these integers to $\mathcal{S}_x^* \setminus \mathcal{S}_m^*$.
Define $\mathcal{S}_- = \{1, \dots, p\} \setminus (\mathcal{S}_x^* \setminus \mathcal{S}_m^*)$.
 - 4: Randomly sample $p\pi_{x \setminus m}$ integers from $\mathcal{S}_-$, and assign these integers to $\mathcal{S}_m^* \setminus \mathcal{S}_x^*$.
Update $\mathcal{S}_- = \mathcal{S}_- \setminus (\mathcal{S}_m^* \setminus \mathcal{S}_x^*)$.
 - 6: Randomly sample $p\pi_{x \cap m}$ integers from $\mathcal{S}_-$, and combine these integers with $\mathcal{S}_x^* \setminus \mathcal{S}_m^*$ to generate $\mathcal{S}_x^*$, combine these integers with $\mathcal{S}_m^* \setminus \mathcal{S}_x^*$ to generate $\mathcal{S}_m^*$.
-

We generate the effect size as $\beta_{X_j} | j \in \mathcal{S}_x^* \sim \mathcal{N}(0, \varepsilon_x^2), \beta_{X_j} | j \notin \mathcal{S}_x^* \sim \delta_0, \beta_{M_j} | j \in \mathcal{S}_m^* \sim \mathcal{N}(\tau_X \beta_{X_j}, \varepsilon_\delta^2)$, and $\beta_{M_j} | j \notin \mathcal{S}_m^* \sim \delta_0$. δ_0 is the Dirac measure centers at 0. $\mathcal{S}_\delta^*$ becomes larger.

Under the total-overlapping indexes setting, the probability of excluding the relevant IVs also reduces.

7.3 Efficiency Gain

In this section, we first provide more simulation results when $\tau_X = 0$ (DGP 2) and then the results under different ratios of $|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|/|\mathcal{S}_m^* \setminus \mathcal{S}_x^*|$. We note that selection events in simulation results for $\tau_X = 0$ from Figure 3- 6 is the empirical estimates of $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$, while Table 1 is the MCSD comparison under the oracle scenario, that is, $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$ are known as priori and we directly use $\mathcal{S}_x^*$ and $\mathcal{S}_m^*$ to estimate causal effects.

7.3.1 Simulation Results under $\tau_X = 0$

We note that when $\tau_X = 0$, $\mathcal{S}_m^* = \mathcal{S}_\delta^*$. By changing π_x and π_δ defined in the manuscript under DGP 2 (i) and DGP 2 (ii), we can have different $|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|/|\mathcal{S}_m^* \setminus \mathcal{S}_x^*|$ ratios. To be more concrete, we change $\pi_x \in \{0.005, 0.01, 0.015\}$ and $\pi_\delta \in \{0.01, 0.015\}$. Then under these two scenarios, we have $|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|/|\mathcal{S}_m^* \setminus \mathcal{S}_x^*| \in \{\frac{1}{5}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}, 3\}$. We present some results here.

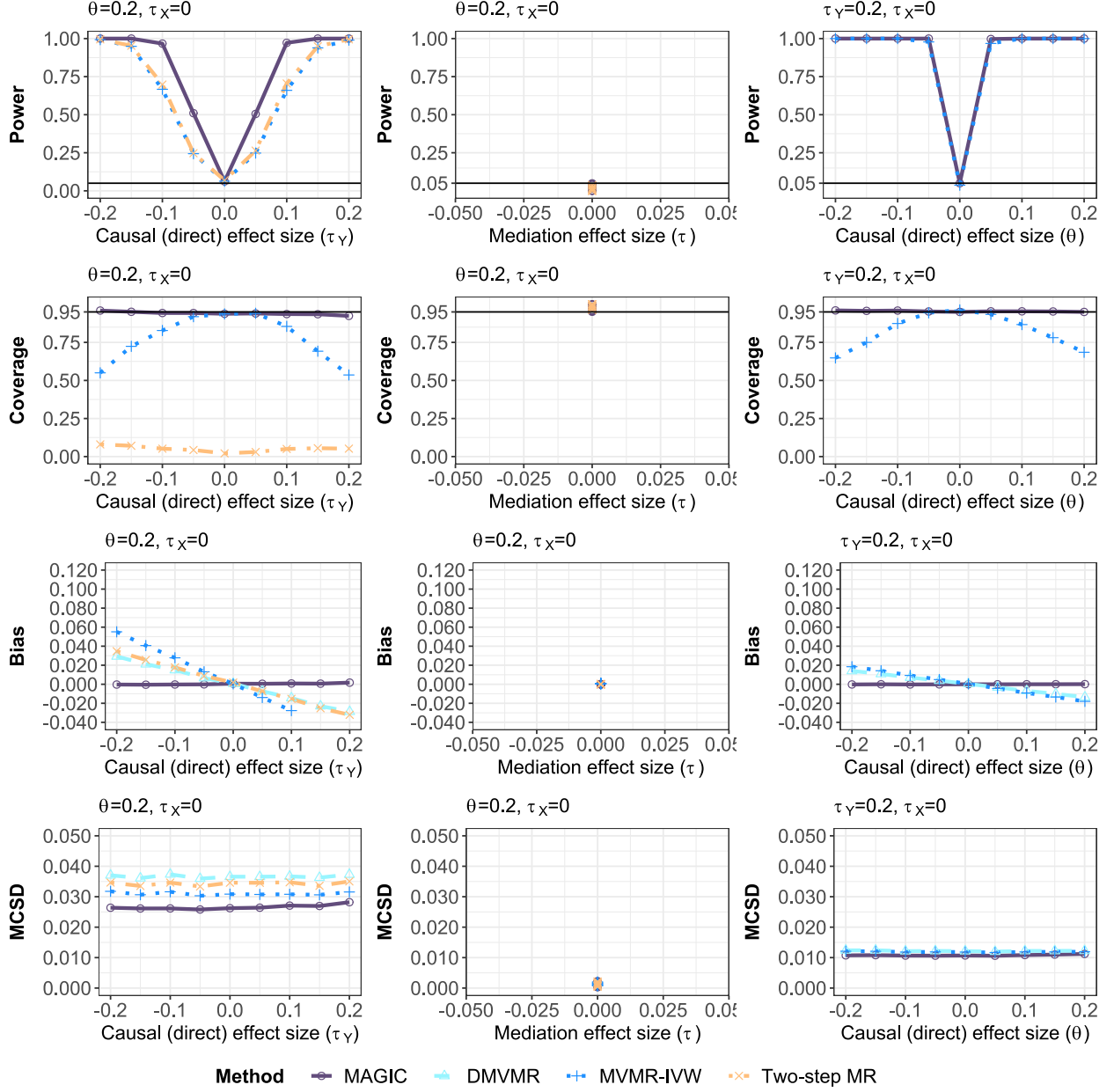


Figure 3: DGP 2 (i) ($\mathcal{S}_x^* \cap \mathcal{S}_\delta^* = \emptyset$): Power, coverage, bias, and MCSD of MAGIC, MVMR-IVW, DMVMR, and Two-step MR computed when $\pi_x = 0.015, \pi_\delta = 0.01$. First and second columns are computed under $\theta = 0.2, \tau_X = 0$ and $\varepsilon_x^2 = 1 \times 10^{-4}, \varepsilon_\delta^2 = 5 \times 10^{-5}$. The third column is computed when $\tau_Y = 0.2$ and $\tau_X = 0$. As two-step MR does not estimate θ , it is not contained in the third row. Now $|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|/|\mathcal{S}_m^* \setminus \mathcal{S}_x^*| = \frac{3}{2}$

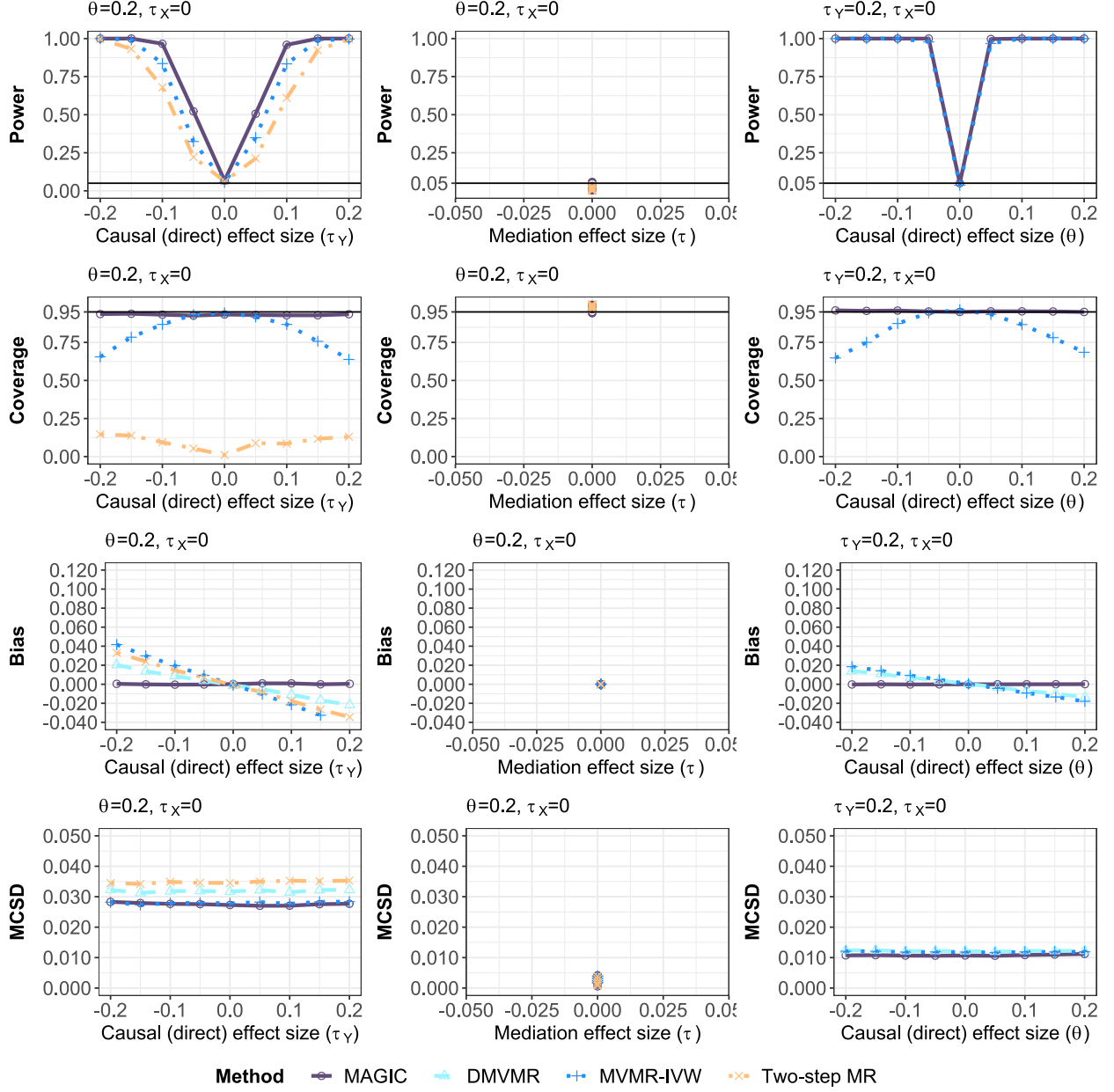


Figure 4: DGP 2(ii) ($|\mathcal{S}_x^* \cap \mathcal{S}_\delta^*| = |\mathcal{S}_x^*|/2$): Power, coverage, bias, and MCSD of MAGIC, MVMR-IVW, DMVMR, and Two-step MR computed when $\pi_x = 0.015, \pi_\delta = 0.01$. First and second columns are computed under $\theta = 0.2, \tau_X = 0$ and $\varepsilon_x^2 = 1 \times 10^{-4}, \varepsilon_\delta^2 = 5 \times 10^{-5}$. The third column is computed when $\tau_Y = 0.2$ and $\tau_X = 0$. As two-step MR does not estimate θ , it is not contained in the third row. Now $|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|/|\mathcal{S}_m^* \setminus \mathcal{S}_x^*| = 3$.

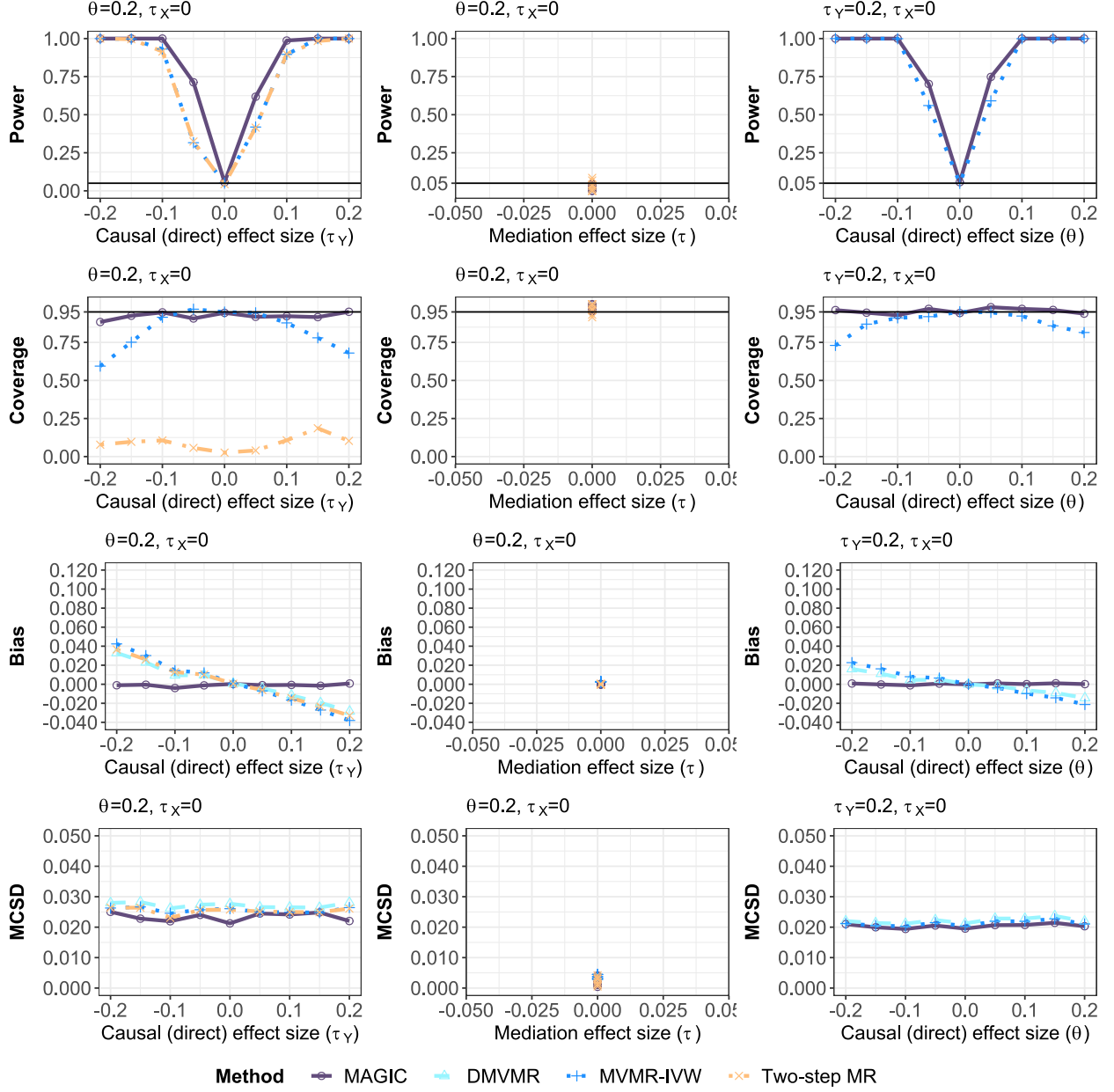


Figure 5: DGP 2(i) ($\mathcal{S}_x^* \cap \mathcal{S}_\delta^* = \emptyset$): Power, coverage, bias, and MCSD of MAGIC, MVMR-IVW, DMVMR, and Two-step MR computed when $\pi_x = 0.005, \pi_\delta = 0.015$. First and second columns are computed under $\theta = 0.2, \tau_X = 0$ and $\varepsilon_x^2 = 1 \times 10^{-4}, \varepsilon_\delta^2 = 5 \times 10^{-5}$. The third column is computed when $\tau_Y = 0.2$ and $\tau_X = 0$. As two-step MR does not estimate θ , it is not contained in the third row. Now $|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|/|\mathcal{S}_m^* \setminus \mathcal{S}_x^*| = \frac{1}{3}$

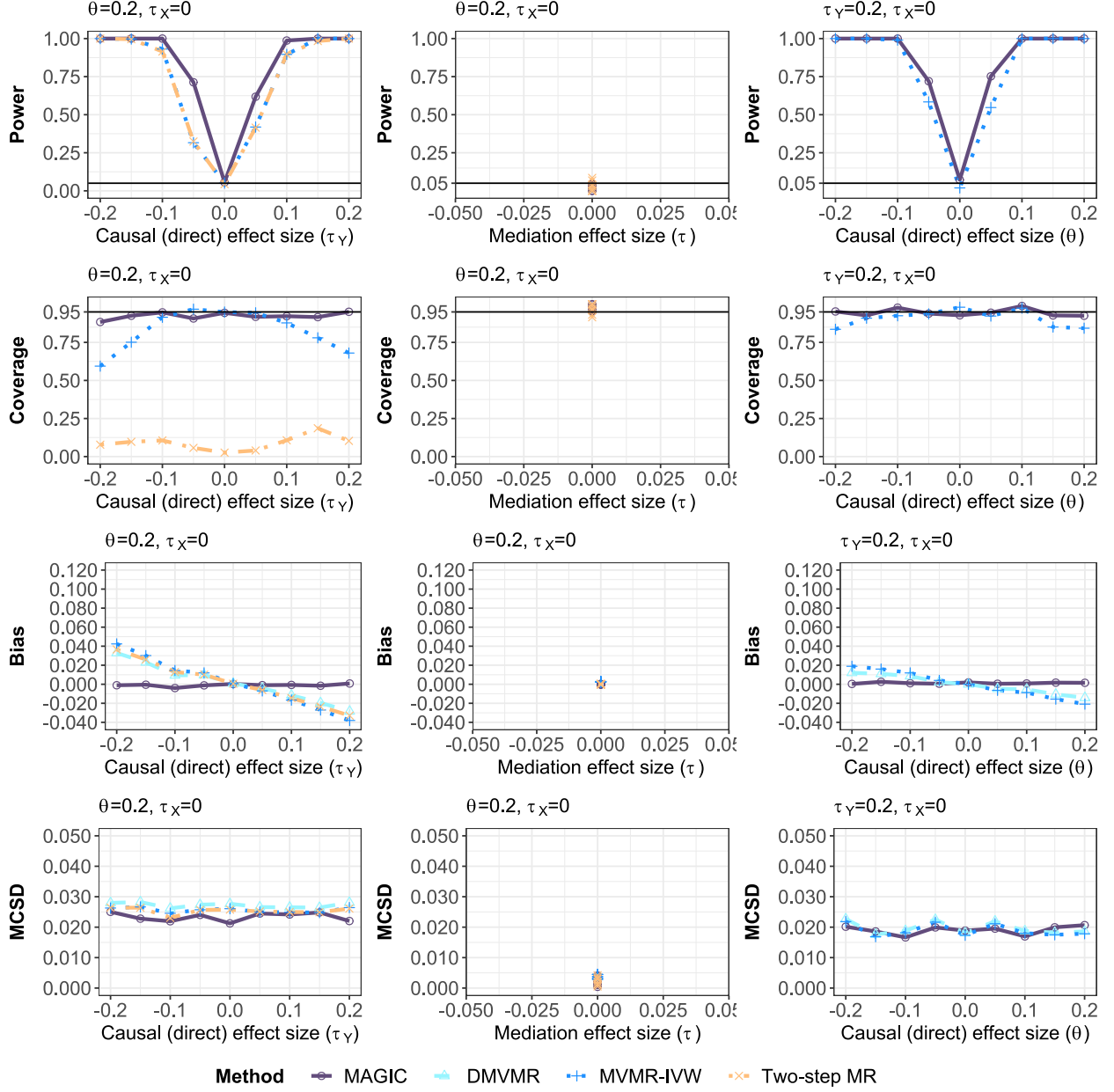


Figure 6: DGP 2(ii) ($|\mathcal{S}_x^* \cap \mathcal{S}_\delta^*| = |\mathcal{S}_x^*|/2$): Power, coverage, bias, and MCSD of MAGIC, MVMR-IVW, DMVMR, and Two-step MR computed when $\pi_x = 0.005, \pi_\delta = 0.015$. First and second columns are computed under $\theta = 0.2, \tau_X = 0$ and $\varepsilon_x^2 = 1 \times 10^{-4}, \varepsilon_\delta^2 = 5 \times 10^{-5}$. The third column is computed when $\tau_Y = 0.2$ and $\tau_X = 0$. As two-step MR does not estimate θ , it is not contained in the third row. Now $|\mathcal{S}_x^* \setminus \mathcal{S}_m^*|/|\mathcal{S}_m^* \setminus \mathcal{S}_x^*| = \frac{1}{5}$.

7.3.2 Additional Simulation Results

We use the index construction way discussed in Section 7.2 and report the MCSD of DMVMR* and MAGIC* under oracle setting.

$ \mathcal{S}_x^* \setminus \mathcal{S}_m^* / \mathcal{S}_m^* \setminus \mathcal{S}_x^* = 0.5$		$ \mathcal{S}_x^* \setminus \mathcal{S}_m^* / \mathcal{S}_m^* \setminus \mathcal{S}_x^* = 6$		$\mathcal{S}_m^* = \mathcal{S}_x^*$	
	DMVMR*	MAGIC*	DMVMR*	MAGIC*	
$\pi_{x \setminus m} = \pi_{m \setminus x} = \pi_{x \cap m} = 0.0005, \varepsilon_x^2 = 5 \times 10^{-5}, \varepsilon_\delta^2 = 5 \times 10^{-5}$					$\pi_{x \cap m} = 0.001$
$\hat{\theta}$	0.060	0.055	0.058	0.057	0.063
$\hat{\tau}_Y$	0.043	0.042	0.068	0.064	0.056
$\pi_{x \setminus m} = \pi_{m \setminus x} = \pi_{x \cap m} = 0.0005, \varepsilon_x^2 = 5 \times 10^{-5}, \varepsilon_\delta^2 = 1 \times 10^{-4}$					$\pi_{x \cap m} = 0.001$
$\hat{\theta}$	0.061	0.056	0.056	0.055	0.058
$\hat{\tau}_Y$	0.030	0.029	0.047	0.045	0.037
$\pi_{x \setminus m} = \pi_{m \setminus x} = \pi_{x \cap m} = 0.0005, \varepsilon_x^2 = 5 \times 10^{-5}, \varepsilon_\delta^2 = 3 \times 10^{-4}$					$\pi_{x \cap m} = 0.001$
$\hat{\theta}$	0.056	0.052	0.053	0.053	0.056
$\hat{\tau}_Y$	0.016	0.016	0.026	0.026	0.021
$\pi_{x \setminus m} = \pi_{m \setminus x} = \pi_{x \cap m} = 0.001, \varepsilon_x^2 = 5 \times 10^{-5}, \varepsilon_\delta^2 = 1 \times 10^{-4}$					$\pi_{x \cap m} = 0.002$
$\hat{\theta}$	0.041	0.038	0.039	0.039	0.039
$\hat{\tau}_Y$	0.020	0.019	0.031	0.030	0.025
$\pi_{x \setminus m} = \pi_{m \setminus x} = \pi_{x \cap m} = 0.001, \varepsilon_x^2 = 5 \times 10^{-5}, \varepsilon_\delta^2 = 3 \times 10^{-4}$					$\pi_{x \cap m} = 0.002$
$\hat{\theta}$	0.040	0.037	0.038	0.038	0.039
$\hat{\tau}_Y$	0.011	0.011	0.018	0.018	0.014
$\pi_{x \setminus m} = \pi_{m \setminus x} = \pi_{x \cap m} = 0.002, \varepsilon_x^2 = 5 \times 10^{-5}, \varepsilon_\delta^2 = 1 \times 10^{-4}$					$\pi_{x \cap m} = 0.004$
$\hat{\theta}$	0.028	0.026	0.028	0.028	0.019
$\hat{\tau}_Y$	0.014	0.014	0.024	0.022	0.017
$\pi_{x \setminus m} = \pi_{m \setminus x} = \pi_{x \cap m} = 0.002, \varepsilon_x^2 = 5 \times 10^{-5}, \varepsilon_\delta^2 = 3 \times 10^{-4}$					$\pi_{x \cap m} = 0.004$
$\hat{\theta}$	0.029	0.026	0.027	0.026	0.020
$\hat{\tau}_Y$	0.008	0.008	0.013	0.013	0.016

Table 1: MCSD under different simulation designs and parameter settings. The true value of $(\theta, \tau_Y, \tau_X) = (0.2, 0.2, 0.6)$. MAGIC* is the MAGIC estimator under the oracle scenario, and DMVMR* is DMVMR under the oracle scenario. The total number of SNPs is 100,000. Here $\pi_{x \setminus m}$ is the proportion of SNPs that are only associated with X , $\pi_{m \setminus x}$ is the proportion of SNPs that are only associated with M , $\pi_{x \cap m}$ is the proportion of SNPs that are associated with both X and M .

8 Real Data Analysis

8.1 Data Harmonization

We harmonize the data through the following steps. First, only genetic variants that exist in GWAS (I), GWAS (II), and GWAS (III) are kept. Second, we select independent genetic variants that have no linkage disequilibrium (R-Squared < 0.001 with an extension of 10,000 Kb in the genome), which has been widely adopted in applied MR studies ([Hemani *et al.*, 2018](#)), with other genetic variants selected through the LD Clumping or the σ -based Clumping by PLINK ([Purcell *et al.*, 2007](#)).

The LD Clumping procedure is based on the p -value of different SNPs. As we have both the exposure dataset and the mediator dataset, we first update the SNPs' p -value using the minimum of its p -value in both the exposure dataset and the mediator dataset and then perform LD clumping. However, LD Clumping will introduce a selection bias that is significantly different from what the existing literature has documented (Hu *et al.*, 2022). In MAGIC, we adopt σ -based Clumping, which is a modified version of LD clumping for MR with mediation analysis. Both clumping algorithms are provided below. In this supplementary material, we also illustrate the performance of σ -based LD Clumping, LD Clumping, and LD Pruning for mediation analysis as follows in Tables 3 -10.

Algorithm 3 LD Clumping

Input: GWAS(I) $\{\widehat{\beta}_{Xj}, \sigma_{Xj}\}_{j=1}^p$ and GWAS(III) $\{\widehat{\beta}_{Mj}, \sigma_{Mj}\}_{j=1}^p$.

Output: Independent SNP set $\mathcal{P}$.

Calculate p -values $\{p_{Xj}\}_{j=1}^p$ and $\{p_{Mj}\}_{j=1}^p$ for GWAS(I) and GWAS(III), respectively.

2: **for** $j = 1, \dots, p$ **do**

Assign $p_{j,\min} = \min\{p_{Xj}, p_{Mj}\}$.

4: Use PLINK to do LD Clumping based on $\{p_{j,\min}\}_{j=1}^p$.

Generate $\mathcal{P}$.

Algorithm 4 σ -based LD Clumping

Input: GWAS(I) $\{\widehat{\beta}_{Xj}, \sigma_{Xj}\}_{j=1}^p$ and GWAS(III) $\{\widehat{\beta}_{Mj}, \sigma_{Mj}\}_{j=1}^p$.

Output: Independent SNP set $\mathcal{P}$.

Generate $\mathbf{rank}_{Xj} = \frac{\sigma_{Xj}^2}{\frac{1}{p} \sum_{i=1}^p \sigma_{Xi}^2}$.

2: Generate $\mathbf{rank}_{Mj} = \frac{\sigma_{Mj}^2}{\frac{1}{p} \sum_{i=1}^p \sigma_{Mi}^2}$.

for $j = 1, \dots, p$ **do**

4: Assign $\mathbf{rank}_{j,\min} = \min\{\mathbf{rank}_{Xj}, \mathbf{rank}_{Mj}\}$.

Use PLINK to do LD Clumping based on $\{\mathbf{rank}_{j,\min}\}_{j=1}^p$.

6: Generate $\mathcal{P}$.

Note that the above-mentioned methods are based on the pooled GWAS(I) and GWAS(III) datasets. Then, by leveraging allele frequency information, we infer the strand direction of ambiguous SNPs and the harmonized exposure-mediator, mediator-outcome, and exposure-outcome datasets by using the TwoSampleMR package (Burgess *et al.*, 2013). We use the default setting with $\lambda = 4.06$ and $\eta = 0.5$ for our proposed MAGIC, and set $\lambda = 5.45$ for both Two-step MR and MVMR-IVW estimators.

8.2 σ -based LD Clumping Results

In this section and next two sections, we provide both adjusted p -values of causal pathways we investigated in the main manuscript and the estimated effects conducted by two-step MR / MAGIC/

MVMR-IVW among different causal pathways using LD Clumping / σ -based LD Clumping / LD pruning. First, we report the adjusted p -values with negative controls (the hair color before graying).

Metabolic Mediators							Negative Controls	
Mediators	Pure hypercholes- -terolemia	LDL	HDL	SBP	DBP	Fasting Glu	Hair Color (blonde)	Hair Color (red)
BMI $\rightarrow$ Mediator $\rightarrow$ CAD								
MAGIC	0.000	0.075	0.000	0.000	0.000	0.783	0.884	0.884
Two-step MR	0.008	0.056	0.000	0.004	0.064	0.079	0.701	0.909
WHR $\rightarrow$ Mediator $\rightarrow$ CAD								
MAGIC	0.050	0.494	0.050	0.532	0.050	0.532	0.946	0.946
Two-step MR	0.024	0.999	0.106	0.198	0.064	0.214	0.999	0.999

Table 2: This table contains BH adjusted p -values for the mediation effects obtained from MAGIC and the two-step MR.

Then, we provide results under different causal pathways (only mediators whose mediation effect is identified as significant after BH procedure).

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{S_x}	# IV _{S_m}
BMI $\rightarrow$ SBP $\rightarrow$ Ischemic stroke										
MAGIC	0.101	0.028	-0.008	0.075	0.749	0.142	0.134	0.027	707	247
Two-step MR	0.069	0.015	-	-	0.681	0.083	0.101	0.018	452	92
MVMR-IVW	0.071	-	0.077	0.050	0.603	0.073	-	-	500	544
BMI $\rightarrow$ DBP $\rightarrow$ Ischemic stroke										
MAGIC	0.140	0.034	0.062	0.075	0.585	0.121	0.240	0.026	719	268
Two-step MR	0.083	0.022	-	-	0.419	0.105	0.198	0.017	448	107
MVMR-IVW	0.068	-	0.068	0.053	0.443	0.076	-	-	492	555
WHR $\rightarrow$ DBP $\rightarrow$ Ischemic stroke										
MAGIC	0.161	0.048	0.120	0.131	0.608	0.085	0.265	0.073	115	555
Two-step MR	0.080	0.029	-	-	0.477	0.063	0.167	0.057	23	224
MVMR-IVW	0.097	-	0.104	0.097	0.487	0.0628	-	-	29	247

Table 3: Causal Pathway: “BMI $\rightarrow$ Mediator $\rightarrow$ Ischemic stroke,” and “WHR $\rightarrow$ Mediator $\rightarrow$ Stroke.”

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{S_x}	# IV _{S_m}
BMI → Pure hypercholesterolemia → CAD										
MAGIC	0.355	0.097	-0.007	0.123	12.332	3.023	0.029	0.003	816	39
Two-step MR	0.377	0.126	-	-	15.092	4.873	0.025	0.002	487	10
MVMR-IVW	0.224	-	0.002	0.073	9.445	1.125	-	-	494	497
BMI → HDL → CAD										
MAGIC	0.178	0.041	0.092	0.107	-0.535	0.127	-0.332	0.028	671	328
Two-step MR	0.124	0.030	-	-	-0.385	0.090	-0.323	0.020	434	151
MVMR-IVW	0.113	-	0.103	0.073	-0.433	0.073	-	-	484	585
BMI → SBP → CAD										
MAGIC	0.130	0.036	0.214	0.093	0.967	0.188	0.135	0.026	712	244
Two-step MR	0.069	0.020	-	-	0.656	0.152	0.104	0.018	452	91
MVMR-IVW	0.070	-	0.147	0.071	0.750	0.104	-	-	500	543
BMI → DBP → CAD										
MAGIC	0.200	0.054	0.155	0.098	0.808	0.195	0.248	0.026	694	250
Two-step MR	0.079	0.039	-	-	0.396	0.189	0.200	0.017	445	107
MVMR-IVW	0.125	-	0.084	0.079	0.529	0.112	-	-	492	552
WHR → DBP → CAD										
MAGIC	0.112	0.048	0.387	0.181	0.700	0.142	0.160	0.061	123	580
Two-step MR	0.085	0.035	-	-	0.504	0.122	0.169	0.057	23	225
MVMR-IVW	0.012	-	0.274	0.189	0.492	0.119	-	-	29	248
WHR → HDL → CAD										
MAGIC	0.217	0.081	-0.169	0.261	-0.667	0.114	-0.324	0.111	89	623
Two-step MR	0.087	0.043	-	-	-0.332	0.072	-0.263	0.114	18	285
MVMR-IVW	-0.021	-	0.215	0.144	-0.305	0.074	-	-	25	303
BMI → SBP → Stroke										
MAGIC	0.099	0.026	0.062	0.064	0.808	0.127	0.122	0.026	720	242
Two-step MR	0.070	0.015	-	-	0.667	0.079	0.105	0.018	453	91
MVMR-IVW	0.087	-	0.103	0.043	0.593	0.064	-	-	499	544
BMI → DBP → Stroke										
MAGIC	0.141	0.034	0.098	0.068	0.618	0.118	0.229	0.026	706	262
Two-step MR	0.105	0.021	-	-	0.522	0.094	0.201	0.017	451	105
MVMR-IVW	0.093	-	0.085	0.047	0.506	0.067	-	-	492	556

Table 4: Causal Pathway: “BMI → Mediator → CAD” and “BMI → Mediator → Stroke” based on σ -based LD Clumping. $\hat{\tau}$, $\hat{\theta}$, $\hat{\tau}_Y$, $\hat{\tau}_X$, SE, # IV_{S_x}, and # IV_{S_m} stand for the estimated effect size, corresponding standard error, the number of IVs in $\mathcal{S}_x$ (MAGIC) / the first step MR (Two-Step MR) / total effects corresponding IVs (MVMR-IVW), and # IV_{S_m}, which represents the number of IVs in $\mathcal{S}_m$ (MAGIC) / the second step MR (Two-Step MR) / the number of IVs in MVMR-IVW, respectively. MAGIC is conducted with $\lambda = 4.06, \eta = 0.5$. The threshold for Two-step MR is 5.45. As θ is not the parameter of interest for Two-step MR and there is no exact way to calculate $\hat{\tau}$ ’s SE for MVMR-IVW, we leave them blank. The τ for MVMR-IVW is conducted by the difference between the total effect and $\hat{\theta}$ from MVMR-IVW.

8.3 LD Clumping Results

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{S_x}	# IV _{S_m}
BMI → Pure hypercholesterolemia → CAD										
MAGIC	0.428	0.083	-0.074	0.091	16.928	2.646	0.025	0.002	821	35
Two-step MR	0.406	0.100	-	-	17.207	3.958	0.024	0.002	466	8
MVMR-IVW	0.267	-	0.137	0.061	10.969	1.005	-	-	472	474
BMI → HDL → CAD										
MAGIC	0.085	0.024	0.290	0.064	-0.259	0.073	-0.326	0.026	686	346
Two-step MR	0.096	0.022	-	-	-0.311	0.066	-0.308	0.023	385	162
MVMR-IVW	0.100	-	0.291	0.065	-0.333	0.053	-	-	465	547
BMI → SBP → CAD										
MAGIC	0.116	0.021	0.280	0.058	0.873	0.123	0.132	0.017	729	213
Two-step MR	0.080	0.017	-	-	0.786	0.121	0.102	0.015	428	86
MVMR-IVW	0.120	-	0.302	0.056	0.827	0.090	-	-	469	514
BMI → DBP → CAD										
MAGIC	0.132	0.029	0.262	0.062	0.647	0.129	0.205	0.021	719	244
Two-step MR	0.125	0.032	-	-	0.662	0.158	0.189	0.017	434	100
MVMR-IVW	0.114	-	0.268	0.063	0.656	0.095	-	-	465	534

Table 5: Causal Pathway: “BMI → Mediator → CAD” based on LD Clumping. $\hat{\tau}$, $\hat{\theta}$, $\hat{\tau}_Y$, $\hat{\tau}_X$, SE, # IV_{S_x}, and # IV_{S_m} stand for the estimated effect size, corresponding standard error, the number of IVs in $\mathcal{S}_x$ (MAGIC) / the first step MR (Two-Step MR) / total effects corresponding IVs (MVMR-IVW), and # IV_{S_m}, which represents the number of IVs in $\mathcal{S}_m$ (MAGIC) / the second step MR (Two-Step MR) / the number of IVs in MVMR-IVW, respectively. MAGIC is conducted with $\lambda = 4.06, \eta = 0.5$. The threshold for Two-step MR is 5.45. As θ is not the parameter of interest for Two-step MR and there is no exact way to calculate $\hat{\tau}$ ’s SE for MVMR-IVW, we leave them blank. The τ for MVMR-IVW is conducted by the difference between the total effect and $\hat{\theta}$ from MVMR-IVW.

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{s_x}	# IV _{s_m}
BMI → Pure hypercholesterolemia → Stroke										
MAGIC	0.095	0.034	0.115	0.047	3.515	1.165	0.027	0.002	822	35
Two-step MR	0.087	0.032	-	-	3.668	1.292	0.024	0.002	465	8
MVMR-IVW	0.065	-	0.138	0.037	2.650	0.580	-	-	471	473
BMI → HDL → Stroke										
MAGIC	0.022	0.009	0.185	0.035	-0.068	0.028	-0.321	0.027	674	349
Two-step MR	0.027	0.009	-	-	-0.086	0.029	-0.309	0.024	389	163
MVMR-IVW	0.057	-	0.147	0.036	-0.102	0.028	-	-	464	552
BMI → SBP → Stroke										
MAGIC	0.074	0.014	0.139	0.034	0.652	0.072	0.114	0.017	715	236
Two-step MR	0.064	0.013	-	-	0.633	0.080	0.101	0.015	425	88
MVMR-IVW	0.067	-	0.126	0.037	0.615	0.058	-	-	469	513
BMI → DBP → Stroke										
MAGIC	0.098	0.018	0.137	0.038	0.470	0.076	0.209	0.020	710	250
Two-step MR	0.078	0.015	-	-	0.413	0.069	0.189	0.017	429	101
MVMR-IVW	0.086	-	0.112	0.037	0.433	0.056	-	-	466	530

Table 6: Causal Pathway: “BMI → Mediator → Stroke.” $\hat{\tau}$, $\hat{\theta}$, $\hat{\tau}_Y$, $\hat{\tau}_X$, SE, # IV_{s_x}, and # IV_{s_m}.

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{s_x}	# IV _{s_m}
BMI → Pure hypercholesterolemia → Ischemic stroke										
MAGIC	0.111	0.037	0.046	0.052	4.092	1.231	0.027	0.002	827	30
Two-step MR	0.104	0.038	-	-	4.387	1.547	0.024	0.002	465	8
MVMR-IVW	0.080	-	0.085	0.043	3.333	0.661	-	-	471	473
BMI → HDL → Ischemic stroke										
MAGIC	0.035	0.012	0.199	0.043	-0.102	0.035	-0.343	0.027	682	356
Two-step MR	0.029	0.011	-	-	-0.093	0.034	-0.310	0.024	389	162
MVMR-IVW	0.052	-	0.128	0.043	-0.113	0.033	-	-	464	551
BMI → SBP → Ischemic stroke										
MAGIC	0.088	0.016	0.110	0.039	0.665	0.085	0.132	0.016	715	216
Two-step MR	0.061	0.013	-	-	0.603	0.090	0.101	0.015	426	87
MVMR-IVW	0.066	-	0.090	0.042	0.617	0.06758	-	-	469	513
BMI → DBP → Ischemic stroke										
MAGIC	0.101	0.022	0.077	0.045	0.457	0.096	0.220	0.019	709	248
Two-step MR	0.067	0.016	-	-	0.356	0.080	0.188	0.017	429	103
MVMR-IVW	0.070	-	0.077	0.043	0.405	0.063	-	-	466	532

Table 7: Causal Pathway: “BMI → Mediator → Ischemic stroke.”

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{S_x}	# IV _{S_m}
Obesity → Pure hypercholesterolemia → CAD										
MAGIC	0.132	0.034	0.039	0.043	20.059	2.772	0.007	0.001	177	241
Two-step MR	0.075	0.024	-	-	15.384	2.479	0.005	0.001	18	17
MVMR-IVW	0.066	-	0.061	0.062	15.180	1.846	-	-	35	35
Obesity → HDL → CAD										
MAGIC	0.029	0.010	0.130	0.041	-0.352	0.072	-0.082	0.024	52	644
Two-step MR	0.023	0.015	-	-	-0.380	0.056	-0.061	0.038	8	284
MVMR-IVW	0.089	-	0.032	0.047	-0.380	0.056	-	-	97	292
Obesity → SBP → CAD										
MAGIC	0.025	0.010	0.129	0.037	0.779	0.092	0.031	0.012	71	580
Two-step MR	0.020	0.008	-	-	0.782	0.099	0.026	0.010	16	191
MVMR-IVW	0.039	-	0.070	0.044	0.782	0.097	-	-	93	207
Obesity → DBP → CAD										
MAGIC	0.019	0.009	0.099	0.040	0.603	0.103	0.032	0.014	67	602
Two-step MR	0.019	0.010	-	-	0.661	0.112	0.029	0.014	15	222
MVMR-IVW	0.024	-	0.088	0.052	0.652	0.109	-	-	100	237
Obesity → Pure hypercholesterolemia → Stroke										
MAGIC	0.030	0.010	0.053	0.022	4.639	1.194	0.006	0.001	180	235
Two-step MR	0.012	0.006	-	-	2.418	0.942	0.005	0.001	18	17
MVMR-IVW	0.018	-	0.060	0.031	2.464	0.820	-	-	35	35
Obesity → HDL → Stroke										
MAGIC	0.012	0.005	0.064	0.029	-0.120	0.0230	-0.104	0.031	53	658
Two-step MR	0.008	0.005	-	-	-0.123	0.028	-0.061	0.038	8	287
MVMR-IVW	0.035	-	0.028	0.026	-0.119	0.028	-	-	125	295
Obesity → SBP → Stroke										
MAGIC	0.020	0.008	0.074	0.025	0.596	0.056	0.033	0.012	70	587
Two-step MR	0.018	0.006	-	-	0.608	0.058	0.030	0.009	15	194
MVMR-IVW	0.013	-	0.053	0.027	0.608	0.057	-	-	92	209
Obesity → DBP → Stroke										
MAGIC	0.026	0.007	0.070	0.021	0.461	0.060	0.056	0.014	60	605
Two-step MR	0.014	0.007	-	-	0.473	0.055	0.029	0.014	15	221
MVMR-IVW	0.023	-	0.045	0.027	0.468	0.055	-	-	125	236

Table 8: Causal Pathway: “Obesity → Mediator → CAD” and “Obesity → Mediator → Stroke” based on LD Clumping. $\hat{\tau}$, $\hat{\theta}$, $\hat{\tau}_Y$, $\hat{\tau}_X$, SE, # IV_{S_x}, and # IV_{S_m} stand for the estimated effect size based on Figure ??, corresponding standard error, the number of IVs in $\mathcal{S}_x$ (proposed) / first step MR (Two-Step MR) / total effects corresponding IVs (MVMR-IVW), and the number of IVs in $\mathcal{S}_m$ (MAGIC) / second step MR (Two-Step MR) / the number of IVs in MVMR-IVW, respectively. Our method uses $\lambda = 4.06, \eta = 0.5$. The threshold for Two-step MR is 5.45. As $\hat{\theta}$ will not be reported by Two-step MR and no exact way to calculate $\hat{\tau}$ ’s SE for MVMR-IVW, we leave them blank. The τ for MVMR-IVW is conducted by the difference between the total effect and MVMR-IVW $\hat{\theta}$.

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{S_x}	# IV _{S_m}
WHR $\rightarrow$ Pure hypercholesterolemia $\rightarrow$ CAD										
MAGIC	0.428	0.153	-0.162	0.147	19.813	2.404	0.022	0.007	180	233
Two-step MR	0.260	0.083	-	-	15.016	2.550	0.017	0.005	28	16
MVMR-IVW	0.346	-	-0.030	0.198	14.709	1.736	-	-	29	44
WHR $\rightarrow$ HDL $\rightarrow$ CAD										
MAGIC	0.158	0.051	0.290	0.141	-0.331	0.072	-0.479	0.117	58	620
Two-step MR	0.171	0.060	-	-	-0.382	0.056	-0.447	0.141	12	280
MVMR-IVW	0.049	-	0.234	0.147	-0.363	0.057	-	-	24	292
WHR $\rightarrow$ SBP $\rightarrow$ CAD										
MAGIC	0.142	0.044	0.534	0.122	0.899	0.094	0.158	0.046	76	582
Two-step MR	0.081	0.038	-	-	0.789	0.098	0.102	0.047	18	197
MVMR-IVW	0.044	-	0.257	0.158	0.791	0.096	-	-	30	215
WHR $\rightarrow$ DBP $\rightarrow$ CAD										
MAGIC	0.080	0.037	0.298	0.145	0.641	0.105	0.125	0.054	75	596
Two-step MR	0.078	0.042	-	-	0.658	0.111	0.060	0.048	17	223
MVMR-IVW	0.062	-	0.227	0.187	0.640	0.109	-	-	28	240
WHR $\rightarrow$ DBP $\rightarrow$ CAD										
MAGIC	0.080	0.037	0.298	0.145	0.641	0.105	0.125	0.054	75	596
Two-step MR	0.078	0.042	-	-	0.658	0.111	0.060	0.048	17	223
MVMR-IVW	0.062	-	0.227	0.187	0.640	0.109	-	-	28	240
WHR $\rightarrow$ HDL $\rightarrow$ Stroke										
MAGIC	0.059	0.019	0.181	0.140	-0.110	-0.534	0.110	0.054	52	665
Two-step MR	0.051	0.021	-	-	-0.115	0.028	-0.447	0.141	12	285
MVMR-IVW	-0.061	-	0.182	0.079	-0.106	0.029	-	-	24	297
WHR $\rightarrow$ DBP $\rightarrow$ Stroke										
MAGIC	0.092	0.033	0.065	0.104	0.486	0.059	0.190	0.066	73	599
Two-step MR	0.055	0.028	-	-	0.459	0.055	0.119	0.060	17	223
MVMR-IVW	0.102	-	-0.027	0.088	0.472	0.054	-	-	28	240

Table 9: Causal Pathway: “WHR $\rightarrow$ Mediator $\rightarrow$ CAD” and “WHR $\rightarrow$ Mediator $\rightarrow$ Stroke.”

8.4 LD Pruning Results

Based on the above-described algorithm, we discover that with $R^2 < 0.001$, we can not always get enough valid variants, that is, the number of valid variants of $\mathcal{S}_m$ is less than 3 for some mediators, because of the inherent randomness in LD Pruning. One possible solution is to relax the R^2 threshold (Sadler *et al.*, 2022). However, the correlated SNPs will affect the downstream analysis. Here, we just report the analysis results with $R^2 < 0.001$. We can see the large SE, reducing the power of estimators, reporting no discovery.

	$\hat{\tau}$	SE	$\hat{\theta}$	SE	$\hat{\tau}_Y$	SE	$\hat{\tau}_X$	SE	# IV _{S_x}	# IV _{S_m}
BMI → HDL → CAD										
MAGIC	-0.080	0.179	-0.339	0.420	0.195	0.432	-0.410	0.073	43	24
Two-step MR	0.184	0.133	-	-	-0.615	0.436	-0.298	0.047	47	13
MVMR-IVW	0.152	-	0.097	0.220	-0.482	0.302	-	-	61	60
BMI → SBP → CAD										
MAGIC	0.078	0.084	-0.344	0.306	0.705	0.727	0.111	0.097	51	15
Two-step MR	0.125	0.076	-	-	0.632	0.351	0.198	0.050	51	9
MVMR-IVW	0.078	-	0.315	0.175	0.215	0.279	-	-	60	60
BMI → DBP → CAD										
MAGIC	-0.025	0.219	-0.187	0.359	-0.098	0.872	0.250	0.086	48	15
Two-step MR	0.090	0.087	-	-	0.451	0.417	0.200	0.058	48	11
MVMR-IVW	0.143	-	0.307	0.199	0.272	0.272	-	-	54	59

Table 10: Causal Pathway: “BMI → Mediator → CAD” based on LD Pruning. $\hat{\tau}$, $\hat{\theta}$, $\hat{\tau}_Y$, $\hat{\tau}_X$, SE, # IV_{S_x}, and # IV_{S_m} stand for the estimated effect size, corresponding standard error, the number of IVs in $\mathcal{S}_x$ (MAGIC) / the first step MR (Two-Step MR) / total effects corresponding IVs (MVMR-IVW), and # IV_{S_m}, which represents the number of IVs in $\mathcal{S}_m$ (MAGIC) / the second step MR (Two-Step MR) / the number of IVs in MVMR-IVW, respectively. MAGIC is conducted with $\lambda = 4.06, \eta = 0.5$. The threshold for Two-step MR is 5.45. As θ is not the parameter of interest for Two-step MR and there is no exact way to calculate $\hat{\tau}$ ’s SE for MVMR-IVW, we leave them blank. The τ for MVMR-IVW is conducted by the difference between the total effect and $\hat{\theta}$ from MVMR-IVW.

8.5 Data Source

We provide the information for all used datasets with corresponding trait name, GWAS ID, sample size, population, the consortium, and their source, to make sure that (i) the datasets come from the real world, (ii) we conduct mediation analysis using three totally independent GWAS, (iii) we focus on the European population. These information are summarized in Table 11.

The MEGASTROKE project received funding from sources specified at

<http://www.megastroke.org/acknowledgments.html>. MEGASTROKE project main authors are in <http://megastroke.org/authors.html>.

Table 11: GWAS data and source

Trait	Consortium/Database	GWAS ID	Sample size	Population	Source
Exposure					
Body Mass Index	GIANT	ieu-b-40	681,275	European	Yengo et al. (2018)
Obesity class 1	GIANT	ieu-a-90	98,697	European	Berndt et al. (2013)
Waist-to-hip ratio	GIANT	ieu-a-73	212,244	European	Shungin et al. (2015)
Mediator					
Fasting glucose	MAGIC	ieu-b-113	58,074	European	Manning et al. (2012)
LDL cholesterol	UK Biobank	ieu-b-109	403,943	European	Richardson et al. (2020)
HDL cholesterol	UK Biobank	ieu-b-110	440,546	European	Richardson et al. (2020)
Systolic blood pressure	UK Biobank	ukb-b-20175	436,419	European	Hemani et al. (2018)
Diastolic blood pressure	UK Biobank	ukb-b-7992	436,424	European	Hemani et al. (2018)
Hair colour (natural, before greying): Blonde	UK Biobank	ukb-d-1747.1	360,270	European	Howrigan et al. (2023)
Hair colour (natural, before greying): Red	UK Biobank	ukb-d-1747.2	360,270	European	Howrigan et al. (2023)
Pure hypercholesterolaemia	UK Biobank	ukb-b-12651	463,010	European	Hemani et al. (2018)
Outcome					
Coronary heart disease	CARDIoGRAMplusC4D	ieu-a-8	86,995	European	Schunkert et al. (2011)
Any ischemic stroke	MEGASTROKE	-	446,696	European	Malik et al. (2018)

References

- Berndt, S. I., Gustafsson, S., Mägi, R., Ganna, A., Wheeler, E., Feitosa, M. F., Justice, A. E., Monda, K. L., Croteau-Chonka, D. C., Day, F. R. *et al.* (2013). “Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture,” *Nature Genetics*, 45(5), 501–512.
- Burgess, S., Butterworth, A., and Thompson, S. G. (2013). “Mendelian randomization analysis with multiple genetic variants using summarized data,” *Genetic Epidemiology*, 37(7), 658–665.
- Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D. *et al.* (2018). “The MR-base platform supports systematic causal inference across the human phenome,” *elife*, 7, e34408.
- Howrigan, D. P., Abbott, L., rkwalter, Palmer, D., Francioli, L., and Hammerbacher, J. (2023). “Nealelab/UK_Biobank_GWAS: v2,” Zenodo, URL: <https://doi.org/10.5281/zenodo.8011558>.
- Hu, X., Zhao, J., Lin, Z., Wang, Y., Peng, H., Zhao, H., Wan, X., and Yang, C. (2022). “Mendelian randomization for causal inference accounting for pleiotropy and sample structure using genome-wide summary statistics,” *Proceedings of the National Academy of Sciences*, 119(28), e2106858119.
- Ma, X., Wang, J., and Wu, C. (2023). “Supplement to “Breaking the winner’s curse in Mendelian randomization: rerandomized inverse variance weighted estimator”.”
- Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A. *et al.* (2018). “Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes,” *Nature Genetics*, 50(4), 524–537.
- Manning, A. K., Hivert, M.-F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H. *et al.* (2012). “A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance,” *Nature Genetics*, 44(6), 659–669.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* (2007). “PLINK: a tool set for whole-genome association and population-based linkage analyses,” *The American Journal of Human Genetics*, 81(3), 559–575.
- Richardson, T. G., Sanderson, E., Palmer, T. M., Ala-Korpela, M., Ference, B. A., Davey Smith, G., and Holmes, M. V. (2020). “Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis,” *PLoS Medicine*, 17(3), e1003062.

- Sadler, M. C., Auwerx, C., Lepik, K., Porcu, E., and Kutalik, Z. (2022). “Quantifying the role of transcript levels in mediating DNA methylation effects on complex traits and diseases,” *Nature Communications*, 13(1), 7559.
- Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H. *et al.* (2011). “Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease,” *Nature Genetics*, 43(4), 333–338.
- Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R. *et al.* (2015). “New genetic loci link adipose and insulin biology to body fat distribution,” *Nature*, 518(7538), 187–196.
- Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N. *et al.* (2018). “Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of European ancestry,” *Human Molecular Genetics*, 27(20), 3641–3649.