

**SUPPLEMENT TO
“BREAKING THE WINNER’S CURSE IN MENDELIAN RANDOMIZATION:
RERANDOMIZED INVERSE VARIANCE WEIGHTED ESTIMATOR”**

BY XINWEI MA^{1,a}, JINGSHEN WANG^{2,b} AND CHONG WU^{3,c}

¹*Department of Economics, University of California San Diego, x1ma@ucsd.edu*

²*Division of Biostatistics, University of California Berkeley, jingshenwang@berkeley.edu*

³*Department of Biostatistics, The University of Texas MD Anderson Cancer Center, cwu18@mdanderson.org*

Contents

I	Setup and Notation	1
II	Inverse Variance Weighting	1
III	De-biased Inverse Variance Weighting	3
IV	Instrumental Variable Selection	4
IV.1	Rao-Blackwellization	6
IV.2	Inverse Variance Weighting with Instrument Selection	8
IV.3	Smoothed RIVW Estimator	9
V	Correlated Instrumental Variables	11
VI	Auxiliary Lemmas	13
VII	Proofs	15
VII.1	Proof of Lemma S.1	15
VII.2	Proof of Lemma S.2	15
VII.3	Proof of Lemma S.3	16
VII.4	Proof of Theorem S.1	16
VII.5	Proof of Lemma S.4	16
VII.6	Proof of Theorem S.2	17
VII.7	Proof of Lemma S.5	17
VII.8	Proof of Lemma S.6	18
VII.9	Proof of Lemma S.7	19
VII.10	Proof of Lemma S.8	19
VII.11	Proof of Lemma S.9	20
VII.12	Proof of Lemma S.10	21
VII.13	Proof of Lemma S.11	22
VII.14	Proof of Lemma S.12	25
VII.15	Proof of Lemma S.13	26
VII.16	Proof of Lemma S.14	27
VII.17	Proof of Theorem S.3	32

*The authors are in alphabetical order.

VII.18 Proof of Theorem S.4	32
VII.19 Omitted Details of Remark S.1	33
VII.20 Proof of Theorem S.5	34
VII.21 Proof of Lemma S.15	34
VII.22 Proof of Lemma S.16	35
VII.23 Proof of Lemma S.17	37
VII.24 Proof of Lemma S.18	38
VII.25 Proof of Lemma S.19	40
VIII Additional Simulation and Case Study Results	41
VIII.1 Simulation Results Corresponding to Lemma S.16	41
VIII.2 Simulation Under Balanced Horizontal Pleiotropy	42
VIII.3 Same Trait Analyses with Standard Clumping and Pruning	46
VIII.4 Winner's Curse Bias and Measurement Error Bias Decomposition	47
VIII.5 Unequal Sample Sizes for the Exposure and Outcome GWAS	48
VIII.6 Supplementary Simulation Evidence	51
References	54

I. Setup and Notation. Recall from the main paper that we start from the estimated SNP-exposure and SNP-outcome associations, $\{(\hat{\gamma}_j, \hat{\Gamma}_j) : 1 \leq j \leq p\}$, as well as their standard errors, $\{(\sigma_{X_j}, \sigma_{Y_j}) : 1 \leq j \leq p\}$. Properties of these estimates are summarized in the following two assumptions.

Assumption S.1 *It holds for all j that $\Gamma_j = \beta\gamma_j$. In addition, not all γ_j are zero.* ■

The above assumption implies that it is possible to identify the parameter of interest from the (population) marginal regression coefficients.

Assumption S.2 (i) *For any $j \neq j'$, the pairs, $(\hat{\Gamma}_j, \hat{\gamma}_j)$ and $(\hat{\Gamma}_{j'}, \hat{\gamma}_{j'})$, are independent.*
(ii) *For each j ,*

$$\begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \Gamma_j \\ \gamma_j \end{bmatrix}, \begin{bmatrix} \sigma_{Y_j}^2 & 0 \\ 0 & \sigma_{X_j}^2 \end{bmatrix} \right),$$

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

This second assumption imposes several restrictions on the estimated associations. First, it requires the estimates to be statistically independent for different SNPs. Second, the estimates are assumed to be normally distributed. Finally, it is required that the standard errors are of the same order.

For future reference, define

$$\kappa = \frac{1}{p} \sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \asymp \frac{1}{p\nu^2} \sum_{j=1}^p \gamma_j^2.$$

II. Inverse Variance Weighting. To estimate β , or more precisely, to aggregate the individual association estimates, it is common to employ a regression of $\hat{\Gamma}_j$ on $\hat{\gamma}_j$, and the slope coefficient obtained thereof will be denoted by $\hat{\beta}_{\text{IVW}}$. This is known as the inverse variance weighting (IVW) approach:

$$\hat{\beta}_{\text{IVW}} = \frac{\sum_{j=1}^p \hat{\Gamma}_j \hat{\gamma}_j / \sigma_{Y_j}^2}{\sum_{j=1}^p \hat{\gamma}_j^2 / \sigma_{Y_j}^2}.$$

Statistical properties of the IVW estimator have been studied in the literature. To facilitate discussion and comparison with our proposed method, we reproduce some of the results in this section.

To begin with, we rewrite the estimates as the sum of the true parameters and the estimation errors, that is,

$$\begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix} = \begin{bmatrix} \Gamma_j \\ \gamma_j \end{bmatrix} + \begin{bmatrix} u_{Y_j} \\ u_{X_j} \end{bmatrix}, \quad \begin{bmatrix} u_{Y_j} \\ u_{X_j} \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \sigma_{Y_j}^2 & 0 \\ 0 & \sigma_{X_j}^2 \end{bmatrix} \right).$$

Then we have $\hat{\Gamma}_j = \Gamma_j + u_{Y_j} = \beta\gamma_j + u_{Y_j} = \beta\hat{\gamma}_j + u_{Y_j} - \beta u_{X_j}$, and hence

$$\hat{\beta}_{\text{IVW}} = \beta + \frac{\sum_{j=1}^p \gamma_j (u_{Y_j} - \beta u_{X_j}) / \sigma_{Y_j}^2}{\sum_{j=1}^p \hat{\gamma}_j^2 / \sigma_{Y_j}^2} + \frac{\sum_{j=1}^p u_{X_j} (u_{Y_j} - \beta u_{X_j}) / \sigma_{Y_j}^2}{\sum_{j=1}^p \hat{\gamma}_j^2 / \sigma_{Y_j}^2}.$$

We will first analyze the denominator, $\sum_{j=1}^p \hat{\gamma}_j^2 / \sigma_{Y_j}^2$, in the above expansion.

Lemma S.1 Assume Assumptions S.1 and S.2 hold, and that $p \rightarrow \infty$. Then

$$\sum_{j=1}^p \frac{\hat{\gamma}_j^2}{\sigma_{Y_j}^2} = \sum_{j=1}^p \frac{\gamma_j^2 + \sigma_{X_j}^2}{\sigma_{Y_j}^2} + O_p\left(\sqrt{p(\kappa+1)}\right) = \left[\sum_{j=1}^p \frac{\gamma_j^2 + \sigma_{X_j}^2}{\sigma_{Y_j}^2} \right] (1 + o_p(1)) \asymp_p p(\kappa+1).$$

■

To summarize, Lemma S.1 demonstrates that as long as $p \rightarrow \infty$, the denominator in the expansion of $\hat{\beta}_{IVW}$ is asymptotically nonrandom, and approximates $\sum_{j=1}^p (\gamma_j^2 + \sigma_{X_j}^2) / \sigma_{Y_j}^2$.

Next we study the term $\sum_{j=1}^p \gamma_j (u_{Y_j} - \beta u_{X_j}) / \sigma_{Y_j}^2$ and show that it is asymptotically normally distributed.

Assumption S.3 The true SNP-exposure associations, γ_j , satisfy

$$\frac{\max_{1 \leq j \leq p} \gamma_j^2}{\sum_{j=1}^p \gamma_j^2} \rightarrow 0.$$

■

Lemma S.2 Assume Assumptions S.1, S.2, and S.3 hold, and that $p \rightarrow \infty$. Then

$$\frac{1}{\sqrt{\mathcal{V}_1}} \sum_{j=1}^p \frac{\gamma_j (u_{Y_j} - \beta u_{X_j})}{\sigma_{Y_j}^2} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{where } \mathcal{V}_1 = \sum_{j=1}^p \frac{\gamma_j^2}{\sigma_{Y_j}^2} \left(1 + \beta^2 \frac{\sigma_{X_j}^2}{\sigma_{Y_j}^2} \right).$$

Therefore,

$$\sum_{j=1}^p \frac{\gamma_j (u_{Y_j} - \beta u_{X_j})}{\sigma_{Y_j}^2} \asymp_p \sqrt{p\kappa}.$$

■

Strictly speaking, Assumption S.3 is not needed for this lemma, as the normality conclusion follows directly from Assumption S.2. In a later section, however, we will study the implications of instrumental variable selection. As the selected SNP-exposure association estimates no longer follow a normal distribution, Assumption S.3 (or more precisely, a modified version thereof) turns out to be crucial.

Finally, we study the last term in the expansion of $\hat{\beta}_{IVW}$.

Lemma S.3 Assume Assumptions S.1 and S.2 hold, and that $p \rightarrow \infty$. Then

$$\frac{1}{\sqrt{\mathcal{V}_2}} \left(\sum_{j=1}^p \frac{u_{X_j} (u_{Y_j} - \beta u_{X_j})}{\sigma_{Y_j}^2} + \beta \sum_{j=1}^p \frac{\sigma_{X_j}^2}{\sigma_{Y_j}^2} \right) \xrightarrow{d} \mathcal{N}(0, 1),$$

$$\text{where } \mathcal{V}_2 = \sum_{j=1}^p \frac{\sigma_{X_j}^2}{\sigma_{Y_j}^2} \left(1 + 2\beta^2 \frac{\sigma_{X_j}^2}{\sigma_{Y_j}^2} \right) \asymp p.$$

■

Now we are ready to present the first main result, which concerns the large-sample properties of the IVW estimator $\hat{\beta}_{IVW}$.

Theorem S.1 Assume Assumptions S.1, S.2, and S.3 hold, and that $p \rightarrow \infty$. Then

$$\mathbf{V}_{\text{IVW}}^{-\frac{1}{2}} \left[\hat{\beta}_{\text{IVW}} - \beta \left(\frac{\sum_{j=1}^p \gamma_j^2 / \sigma_{Y_j}^2}{\sum_{j=1}^p (\gamma_j^2 + \sigma_{X_j}^2) / \sigma_{Y_j}^2} \right) \right] \xrightarrow{d} \mathcal{N}(0, 1),$$

$$\text{where } \mathbf{V}_{\text{IVW}} = \frac{\mathcal{V}_1 + \mathcal{V}_2}{\left(\sum_{j=1}^p (\gamma_j^2 + \sigma_{X_j}^2) / \sigma_{Y_j}^2 \right)^2}.$$

Therefore,

$$\hat{\beta}_{\text{IVW}} - \beta \lesssim_p \frac{1}{\kappa + 1} + \sqrt{\frac{1}{p(\kappa + 1)}}.$$

■

From the above theorem, it should be clear that the IVW estimator is consistent if $p, \kappa \rightarrow \infty$, while asymptotic normality requires that $\kappa/p \rightarrow \infty$. The second requirement is fairly strong and unrealistic. If $\kappa/p \lesssim 1$, then the bias of the IVW estimator might be of first order or even dominate in the asymptotic distribution, rendering the standard normal distributional approximation invalid.

III. De-biased Inverse Variance Weighting. From Theorem S.1, it is clear that the IVW estimator is biased toward zero due to the estimation error in $\hat{\gamma}_j$. [14] propose the following de-biased IVW (dIVW) estimator.

$$\hat{\beta}_{\text{dIVW}} = \frac{\sum_{j=1}^p \hat{\Gamma}_j \hat{\gamma}_j / \sigma_{Y_j}^2}{\sum_{j=1}^p (\hat{\gamma}_j^2 - \sigma_{X_j}^2) / \sigma_{Y_j}^2}.$$

To facilitate discussion, we collect several results on the statistical properties of the de-biased IVW estimator in this section.

To start, consider the following decomposition:

$$\hat{\beta}_{\text{dIVW}} = \beta + \frac{\sum_{j=1}^p \gamma_j (u_{Y_j} - \beta u_{X_j}) / \sigma_{Y_j}^2}{\sum_{j=1}^p (\hat{\gamma}_j^2 - \sigma_{X_j}^2) / \sigma_{Y_j}^2} + \frac{\sum_{j=1}^p (u_{X_j} u_{Y_j} - \beta (u_{X_j}^2 - \sigma_{X_j}^2)) / \sigma_{Y_j}^2}{\sum_{j=1}^p (\hat{\gamma}_j^2 - \sigma_{X_j}^2) / \sigma_{Y_j}^2}.$$

Lemma S.4 Assume Assumptions S.1 and S.2 hold, and that $p \rightarrow \infty$. Then

$$\sum_{j=1}^p \frac{\hat{\gamma}_j^2 - \sigma_{X_j}^2}{\sigma_{Y_j}^2} = \sum_{j=1}^p \frac{\gamma_j^2}{\sigma_{Y_j}^2} + O_p \left(\sqrt{p(\kappa + 1)} \right).$$

If, in addition that $\sqrt{p}\kappa \rightarrow \infty$, then

$$\sum_{j=1}^p \frac{\hat{\gamma}_j^2 - \sigma_{X_j}^2}{\sigma_{Y_j}^2} = \left[\sum_{j=1}^p \frac{\gamma_j^2}{\sigma_{Y_j}^2} \right] (1 + o_p(1)) \asymp_p p\kappa.$$

■

In the lemma, we have ignored the scenario that $\sqrt{p}\kappa$ is bounded. If this is the case, then the denominator will converge to a (possibly noncentral) normal distribution under suitable scaling. In other words, the limiting distribution of the dIVW estimator becomes the ratio of two normal distributions.

Theorem S.2 Assume Assumptions S.1, S.2, and S.3 hold, $p \rightarrow \infty$, and $\sqrt{p}\kappa \rightarrow \infty$. Then

$$V_{\text{dIVW}}^{-\frac{1}{2}} [\hat{\beta}_{\text{dIVW}} - \beta] \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{where } V_{\text{dIVW}} = \frac{\mathcal{V}_1 + \mathcal{V}_2}{\left(\sum_{j=1}^p \gamma_j^2 / \sigma_{Y_j}^2\right)^2}.$$

Therefore,

$$\hat{\beta}_{\text{dIVW}} - \beta \lesssim_p \sqrt{\frac{\kappa + 1}{p\kappa^2}}.$$

■

IV. Instrumental Variable Selection. In this section, we consider the problem of instrument selection, and propose valid post-selection estimation and inference procedures. To start, define the selection event as

$$S_j = \left| \frac{\hat{\gamma}_j}{\sigma_{X_j}} + Z_j \right| - \lambda > 0,$$

where λ is a thresholding sequence, and $Z_j \sim \mathcal{N}(0, \eta_j^2)$ represent the pseudo SNPs that we introduce. The following lemma provides the distribution of $\hat{\gamma}_j$ after selection.

Lemma S.5 Assume Assumption S.2 holds. Then the conditional density of $\hat{\gamma}_j / \sigma_{X_j}$ given the selection event $S_j > 0$ is

$$f_{\frac{\hat{\gamma}_j}{\sigma_{X_j}} | S_j > 0}(y) = \frac{1}{\mathbb{P}[S_j > 0]} \phi\left(y - \frac{\gamma_j}{\sigma_{X_j}}\right) \left[\Phi\left(\frac{-\lambda - y}{\eta_j}\right) + 1 - \Phi\left(\frac{\lambda - y}{\eta_j}\right) \right],$$

where

$$\mathbb{P}[S_j > 0] = \Phi\left(\frac{-\lambda - \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}}\right) + 1 - \Phi\left(\frac{\lambda - \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}}\right).$$

In the above, ϕ and Φ are the standard normal density and cumulative distribution functions, respectively. ■

Despite taking a complicated form, we can draw a few conclusions from the above conditional distribution. First, if we set $\eta_j \rightarrow \infty$, then the selection event will be dominated by the random noise in Z_j . In other words, the event $S_j > 0$ will happen with probability approaching 1. In this case, there will be no selection, which means that the above density will converge to the unconditional density of $\hat{\gamma}_j / \sigma_{X_j} \sim \mathcal{N}(\gamma_j / \sigma_{X_j}, 1)$.

At the other extreme (i.e., when $\eta_j \rightarrow 0$) we see that the above density is nonzero only in the region $(-\infty, -\lambda) \cup (\lambda, \infty)$. This implies that the conditional density of $\hat{\gamma}_j / \sigma_{X_j}$ will have a “hole” in the middle. For any other values of η_j , the post-selection distribution of $\hat{\gamma}_j / \sigma_{X_j}$ will be a smoothed interpolation between the extremes. In other words, incorporating the pseudo SNPs Z_j leads to a soft-thresholding selection rule.

For future reference, we define a few post-selection quantities below. To start, we will denote the collection of selected instruments by $\mathcal{S}_\lambda$, that is,

$$\mathcal{S}_\lambda = \{j : S_j > 0\} = \left\{ j : \left| \frac{\hat{\gamma}_j}{\sigma_{X_j}} + Z_j \right| > \lambda \right\}.$$

We will use the notation, $S_j > 0$ and $j \in \mathcal{S}_\lambda$, interchangeably for the selection event. We may also employ $\mathcal{S}_\lambda$ to denote the selection event if the emphasis is not on a specific SNP j .

The cardinality of $\mathcal{S}_\lambda$ is $p_\lambda = |\mathcal{S}_\lambda|$, which is simply the number of selected instruments. Finally, we define κ_λ as

$$\kappa_\lambda = \frac{1}{p_\lambda} \sum_{j \in \mathcal{S}_\lambda} \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2,$$

whenever $p_\lambda > 0$.

To discuss the limiting behaviors of p_λ and κ_λ , we make the following assumption.

Assumption S.4 (i) *The thresholding sequence diverges: $\lambda \rightarrow \infty$.*

(ii) *Define $\underline{\eta} = \min_{1 \leq j \leq p} \eta_j$ and $\bar{\eta} = \max_{1 \leq j \leq p} \eta_j$, then both $\underline{\eta}$ and $\bar{\eta}$ are bounded and bounded away from zero.* ■

Part (i) suggests that the critical value will diverge, which is quite common for conducting multiple hypotheses testing. The second condition is also intuitive: it requires that the noise we introduce will neither shrink to zero nor explode.

Lemma S.6 *Assume Assumptions S.2 and S.4 hold. Then $p_\lambda \xrightarrow{P} \infty$ if and only if*

$$\sum_{j=1}^p \Phi \left(\frac{-\lambda + \left| \frac{\gamma_j}{\sigma_{X_j}} \right|}{\sqrt{1 + \eta_j^2}} \right) \rightarrow \infty.$$

A necessary condition for $p_\lambda \xrightarrow{P} \infty$ is that, for all $c > 0$,

$$\sum_{j=1}^p \mathbf{1} \left(\left| \frac{\gamma_j}{\sigma_{X_j}} \right| \geq \lambda + \Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right) \right) \rightarrow \infty.$$

A sufficient condition for $p_\lambda \xrightarrow{P} \infty$ is that, for some $c_1 > 0$ and $c_2 \geq 0$:

$$\left(\frac{1}{\log(p)} \right)^{\frac{\underline{\eta}^2}{2(1+\underline{\eta}^2)}} \left(\frac{1}{p} \right)^{\frac{c_2}{1+\underline{\eta}^2}} \sum_{j=1}^p \mathbf{1} \left(\left| \frac{\gamma_j}{\sigma_{X_j}} \right| \geq \lambda + \Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right) \right) \rightarrow \infty.$$

■

Another important quantity is κ_λ , which reflects the overall signal strength of the selected instruments. The following lemma provides sufficient conditions under which κ_λ diverges.

Lemma S.7 *Assume Assumptions S.2 and S.4 hold, and that $p_\lambda \xrightarrow{P} \infty$. Then for any $c > 0$,*

$$\kappa_\lambda \gtrsim_P \left[\left(\lambda + \Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right) \right) \vee 0 \right]^2.$$

If in addition that

$$\lambda \geq \left| \Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right) \right|$$

for some $c_1 > 0$ and $c_2 > 1 + \bar{\eta}^2$, then

$$\kappa_\lambda \gtrsim_P \lambda^2.$$

■

The lemma provides a sufficient condition under which the post-selection average instrument strength, κ_λ , is asymptotically larger than λ^2 . We note that the condition is established in the “worst case scenario” where all of the nonzero SNP-exposure associations are equally weak. In practice, however, it is likely that there will be substantial heterogeneity in γ_j , meaning that κ_λ might be much larger compared to λ^2 .

IV.1. *Rao-Blackwellization.* Using the selected instruments to construct the IVW estimator can pose serious challenges to estimation and inference. In particular, after the selection step, $\hat{\gamma}_j$ is no longer unbiased for γ_j . In what follows, we will propose a bias-corrected version of the post-selection instrument effects.

To introduce our bias correction technique, we make a few observations. Let

$$\hat{b}_{j,\text{ini}} = \frac{\hat{\gamma}_j}{\sigma_{X_j}} - \frac{Z_j}{\eta_j^2},$$

then it follows a normal distribution with mean γ_j/σ_{X_j} . By standard covariance calculation, it is also straightforward to verify that $\hat{b}_{j,\text{ini}}$ is uncorrelated with $\frac{\hat{\gamma}_j}{\sigma_{X_j}} + Z_j$, which means that it is independent of the selection event $S_j > 0$. As a result, the law of iterated expectation implies that

$$\hat{b}_j = \mathbb{E} \left[\hat{b}_{j,\text{ini}} \mid \frac{\hat{\gamma}_j}{\sigma_{X_j}}, S_j > 0 \right]$$

will be unbiased for γ_j/σ_{X_j} , conditioning on the selection event $S_j > 0$. In other words, one can employ the bias-corrected estimator,

$$\hat{\gamma}_{j,\text{RB}} = \sigma_{X_j} \mathbb{E} \left[\hat{b}_{j,\text{ini}} \mid \hat{\gamma}_j, S_j > 0 \right] = \hat{\gamma}_j - \frac{\sigma_{X_j}}{\eta_j^2} \mathbb{E}[Z_j \mid \hat{\gamma}_j, S_j > 0]$$

to construct the IVW estimator.

The exact form of $\hat{\gamma}_{j,\text{RB}}$ is given in the following lemma.

Lemma S.8 *Assume Assumption S.2 holds. Then*

$$\hat{\gamma}_{j,\text{RB}} = \hat{\gamma}_j - \frac{\sigma_{X_j}}{\eta_j} \frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,+})}, \quad \text{where } A_{j,\pm} = -\frac{\hat{\gamma}_j}{\sigma_{X_j}\eta_j} \pm \frac{\lambda}{\eta_j}.$$

■

The next lemma presents the variance of the bias-corrected estimator.

Lemma S.9 *Assume Assumption S.2 holds. Then*

$$\begin{aligned} \sigma_{X_j,\text{RB}}^2 = \mathbb{V}[\hat{\gamma}_{j,\text{RB}} \mid S_j > 0] &= \sigma_{X_j}^2 \left(1 - \frac{1}{\eta_j \mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y \phi(y) \left(\phi(B_{j,+}(y)) - \phi(B_{j,-}(y)) \right) dy \right. \\ &\quad \left. + \frac{1}{\eta_j^2 \mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \phi(y) \frac{\left(\phi(B_{j,+}(y)) - \phi(B_{j,-}(y)) \right)^2}{1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y))} dy \right), \end{aligned}$$

where

$$B_{j,\pm}(y) = -\left(\frac{\gamma_j}{\sigma_{X_j}\eta_j} + \frac{y}{\eta_j} \right) \pm \frac{\lambda}{\eta_j}.$$

■

While the above provides a formula for the variance of the Rao-Blackwellized SNP-exposure association estimates, it is not directly amenable to implementation. This is because the expression involves the true SNP-exposure associations, γ_j . Another complication is that the formula depends on γ_j nonlinearly, rendering a plug-in approach inappropriate. In the following we present an alternative conditional variance formula, and demonstrate how it can be unbiasedly estimated.

Lemma S.10 Assume Assumption S.2 holds. Then

$$\sigma_{X_j, \text{RB}}^2 = \sigma_{X_j}^2 - \frac{\sigma_{X_j}^2}{\eta_j^2} \mathbb{E} \left[\left| \frac{A_{j,+} \phi(A_{j,+}) - A_{j,-} \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} - \left(\frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} \right)^2 \right| S_j > 0 \right],$$

where $A_{j,\pm}$ are defined in Lemma S.8. ■

Although the above variance formula also takes a complicated form, it turns out that one can establish both upper and lower bounds. This is formally stated in the following lemma.

Lemma S.11 Assume Assumptions S.2 and S.4 hold. Then there exist constants $0 < c_1 < c_2 < \infty$, such that

$$c_1 \sigma_{X_j}^2 \leq \sigma_{X_j, \text{RB}}^2 \leq c_2 \sigma_{X_j}^2.$$

In other words, $\{\sigma_{X_j, \text{RB}}^2 / \sigma_{X_j}^2 : 1 \leq j \leq p\}$ are both bounded and bounded away from zero. ■

To estimate the variance of the selected SNP-exposure association estimates, consider the following quantity:

$$\hat{\sigma}_{X_j, \text{RB}}^2 = \sigma_{X_j}^2 \left[1 - \frac{1}{\eta_j^2} \frac{A_{j,+} \phi(A_{j,+}) - A_{j,-} \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} + \left(\frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} \right)^2 \right].$$

It is straightforward to verify that conditional on the selection event, $\hat{\sigma}_{X_j, \text{RB}}^2$ is unbiased for $\sigma_{X_j, \text{RB}}^2$. The next lemma provides an error bound for the estimation of $\sum_{j \in S_\lambda} \sigma_{X_j, \text{RB}}^2$, which is what we need in order to correct measurement error bias.

Lemma S.12 Assume Assumptions S.2 and S.4 hold. Then

$$\begin{aligned} & \left| \sum_{j \in S_\lambda} (\hat{\sigma}_{X_j, \text{RB}}^2 - \sigma_{X_j, \text{RB}}^2) \right| \\ & \lesssim_p \nu^2 \left\{ \sum_{j \in S_\lambda} \mathbb{E} \left[3 - 2m_0(\hat{\gamma}_j)^2 - 3m_0(\hat{\gamma}_j)^4 + 9m_1(\hat{\gamma}_j) + 6m_0(\hat{\gamma}_j)^2 m_1(\hat{\gamma}_j) \right. \right. \\ & \quad \left. \left. - 4m_0(\hat{\gamma}_j) m_2(\hat{\gamma}_j) + m_3(\hat{\gamma}_j) \right] S_j > 0 \right\}^{\frac{1}{2}} \\ & \lesssim_p \nu^2 \lambda \sqrt{p_\lambda}, \end{aligned}$$

where

$$m_\ell(\hat{\gamma}_j) = \frac{A_{j,+}^\ell \phi(A_{j,+}) - A_{j,-}^\ell \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})}.$$
■

The extra λ in our error bound stems from higher-order moments of the Rao-Blackwellized estimator. We suspect that one can establish the sharper bound, $\nu^2 \sqrt{p_\lambda}$, under the same conditions, although the proof eludes us due to the complicated expressions of higher-order moments.

IV.2. *Inverse Variance Weighting with Instrument Selection.* We are now ready to state some results regarding our RIVW estimator using the selected instruments, which is defined as

$$\hat{\beta}_{\text{RIVW}} = \frac{\sum_{j \in \mathcal{S}_\lambda} \hat{\Gamma}_j \hat{\gamma}_{j,\text{RB}} / \sigma_{Y_j}^2}{\sum_{j \in \mathcal{S}_\lambda} (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) / \sigma_{Y_j}^2}.$$

The following lemma provides conditions under which the denominator converges.

Lemma S.13 Assume Assumptions S.1 S.2, and S.4 hold, and that $p_\lambda \xrightarrow{\text{P}} \infty$. Then

$$\sum_{j \in \mathcal{S}_\lambda} (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) = \sum_{j \in \mathcal{S}_\lambda} \gamma_j^2 + O_{\text{P}} \left(\sqrt{p_\lambda \nu^4 (\kappa_\lambda + \lambda^2)} \right).$$

If, in addition that $\sqrt{p_\lambda} \kappa_\lambda / \lambda \xrightarrow{\text{P}} \infty$, then

$$\sum_{j \in \mathcal{S}_\lambda} (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) = \left[\sum_{j \in \mathcal{S}_\lambda} \gamma_j^2 \right] (1 + o_{\text{P}}(1)) \asymp_{\text{P}} p_\lambda \nu^2 \kappa_\lambda.$$

■

To proceed, we further decompose the estimator as

$$\hat{\beta}_{\text{RIVW}} = \beta + \frac{\sum_{j \in \mathcal{S}_\lambda} \left[\gamma_j (u_{Y_j} - \beta u_{X_j,\text{RB}}) + (u_{X_j,\text{RB}} u_{Y_j} - \beta (u_{X_j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2)) \right] / \sigma_{Y_j}^2}{\sum_{j \in \mathcal{S}_\lambda} (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) / \sigma_{Y_j}^2}.$$

For future reference, take

$$u_{j,\text{RIVW}} = \gamma_j (u_{Y_j} - \beta u_{X_j,\text{RB}}) + (u_{X_j,\text{RB}} u_{Y_j} - \beta (u_{X_j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2)),$$

which can be understood as the regression error term.

The following lemma is related to the asymptotic variance of our RIVW estimator.

Lemma S.14 Assume Assumptions S.1 S.2, and S.4 hold, $p_\lambda \xrightarrow{\text{P}} \infty$, and that $\kappa_\lambda / \lambda^2 \xrightarrow{\text{P}} \infty$. Then

$$\begin{aligned} \mathbb{V} \left[\sum_{j \in \mathcal{S}_\lambda} u_{j,\text{RIVW}} \middle| \mathcal{S}_\lambda \right] &= \sum_{j \in \mathcal{S}_\lambda} \left\{ \gamma_j^2 (\sigma_{Y_j}^2 + \beta^2 \sigma_{X_j,\text{RB}}^2) + \sigma_{X_j,\text{RB}}^2 \sigma_{Y_j}^2 \right. \\ &\quad + \beta^2 \left(\mathbb{E}[u_{X_j,\text{RB}}^4 | S_j > 0] + \mathbb{E}[\hat{\sigma}_{X_j,\text{RB}}^4 | S_j > 0] - 2\mathbb{E}[u_{X_j,\text{RB}}^2 \hat{\sigma}_{X_j,\text{RB}}^2 | S_j > 0] \right) \\ &\quad \left. + 2\beta^2 \gamma_j \mathbb{E}[u_{X_j,\text{RB}}^3 - u_{X_j,\text{RB}} \hat{\sigma}_{X_j,\text{RB}}^2 | S_j > 0] \right\} \\ &\asymp_{\text{P}} p_\lambda \nu^4 \kappa_\lambda. \end{aligned}$$

■

It is not surprising that the asymptotic variance of our RIVW estimator takes a complicated form, as the variance of $u_{j,\text{RIVW}}$ involves higher-order moments of $\hat{\gamma}_{j,\text{RB}}$. We employ the condition $\kappa_\lambda / \lambda^2 \xrightarrow{\text{P}} 0$ to simplify the analysis. In particular, this assumption implies that the leading variance stems from the first term in the expression of $u_{j,\text{RIVW}}$. We suspect that it is possible to establish the asymptotic normality of our RIVW estimator and the consistency of

our variance estimator (Theorems S.3 and S.4 below) without this assumption. However, a theoretical justification eludes us due to the complicated higher-moments calculations.

For the asymptotic normality result, we impose the following assumption.

Assumption S.5

$$\frac{\max_{j \in \mathcal{S}_\lambda} \gamma_j^2}{\sum_{j \in \mathcal{S}_\lambda} \gamma_j^2} \xrightarrow{P} 0.$$

■

Theorem S.3 Assume Assumptions S.1, S.2, S.4, and S.5 hold, $p_\lambda \xrightarrow{P} \infty$, and that $\kappa_\lambda/\lambda^2 \xrightarrow{P} \infty$. Then conditional on the selection event $\mathcal{S}_\lambda$, the centered and scaled RIVW estimator is asymptotically normally distributed, that is,

$$V_{\text{RIVW}}^{-\frac{1}{2}} [\hat{\beta}_{\text{RIVW}} - \beta] \xrightarrow{d} \mathcal{N}(0, 1),$$

where

$$V_{\text{RIVW}} = \frac{\mathbb{V} \left[\sum_{j \in \mathcal{S}_\lambda} u_{j, \text{RIVW}} / \sigma_{Y_j}^2 \middle| \mathcal{S}_\lambda \right]}{\left(\sum_{j \in \mathcal{S}_\lambda} \gamma_j^2 / \sigma_{Y_j}^2 \right)^2} \asymp_P \frac{1}{p_\lambda \kappa_\lambda}.$$

■

Finally, we show that our standard error construction is valid. Recall from the main paper that

$$\hat{V}_{\text{RIVW}} = \frac{\sum_{j \in \mathcal{S}_\lambda} \left[\hat{\Gamma}_j \hat{\gamma}_{j, \text{RB}} - \hat{\beta}_{\text{RIVW}} (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \right]^2 / \sigma_{Y_j}^4}{\left[\sum_{j \in \mathcal{S}_\lambda} \left(\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2 \right) / \sigma_{Y_j}^2 \right]^2}$$

Theorem S.4 Assume Assumptions S.1, S.2, and that S.4 hold, $p_\lambda \xrightarrow{P} \infty$, and $\kappa_\lambda/\lambda^2 \xrightarrow{P} \infty$. Then

$$\frac{\hat{V}_{\text{RIVW}}}{V_{\text{RIVW}}} \xrightarrow{P} 1.$$

■

IV.3. Smoothed RIVW Estimator. Our RIVW estimator is constructed based on the selected instruments, that is,

$$\hat{\beta}_{\text{RIVW}} = \frac{\sum_{j \in \mathcal{S}_\lambda} \hat{\Gamma}_j \hat{\gamma}_{j, \text{RB}} / \sigma_{Y_j}^2}{\sum_{j \in \mathcal{S}_\lambda} (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) / \sigma_{Y_j}^2} = \frac{\sum_{j=1}^p \hat{\Gamma}_j \hat{\gamma}_{j, \text{RB}} \mathbf{1}_{j \in \mathcal{S}_\lambda} / \sigma_{Y_j}^2}{\sum_{j=1}^p (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \mathbf{1}_{j \in \mathcal{S}_\lambda} / \sigma_{Y_j}^2}.$$

Due to the pseudo SNPs we introduced into the selection, however, the RIVW estimator may have a larger variance. One possibility is to consider a weighting approach, where for each SNP the weight corresponds to the (conditional) probability of being selected. Formally, we define

$$\hat{\omega}_j = \mathbb{P}[j \in \mathcal{S}_\lambda | \hat{\gamma}_j] = 1 - \Phi(A_{j,+}) + \Phi(A_{j,-}),$$

where we recall (from Lemma S.8) that

$$A_{j,\pm} = -\frac{\hat{\gamma}_j}{\sigma_{X_j} \eta_j} \pm \frac{\lambda}{\eta_j}.$$

The conditional selection probabilities motivate the following smoothed RIVW (sRIVW) estimator

$$\hat{\beta}_{\text{sRIVW}} = \frac{\sum_{j=1}^p \hat{\Gamma}_j \hat{\gamma}_{j,\text{RB}} \hat{\omega}_j / \sigma_{Y_j}^2}{\sum_{j=1}^p (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \hat{\omega}_j / \sigma_{Y_j}^2}.$$

To provide some intuition regarding properties of the sRIVW estimator, we may consider the expectations of the numerator and the denominator. To be specific, we have

$$\mathbb{E}[\hat{\gamma}_{j,\text{RB}} \hat{\omega}_j] = \mathbb{E}[\hat{\gamma}_{j,\text{RB}} \mathbf{1}_{j \in S_\lambda}] = \gamma_j \omega_j,$$

where the second equality can be deduced using the law of iterated expectation, and ω_j is the unconditional probability of an SNP being selected:

$$\omega_j = \mathbb{E}[\hat{\omega}_j] = 1 - \Phi\left(\frac{\lambda - \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}}\right) + \Phi\left(\frac{-\lambda - \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}}\right).$$

As a result, the numerator in our sRIVW estimator satisfies

$$\mathbb{E}\left[\sum_{j=1}^p \frac{\hat{\Gamma}_j \hat{\gamma}_{j,\text{RB}} \hat{\omega}_j}{\sigma_{Y_j}^2}\right] = \beta \sum_{j=1}^p \frac{\omega_j}{\sigma_{Y_j}^2} \gamma_j^2.$$

It is also straightforward to find the expectation of the denominator, as

$$\mathbb{E}\left[\sum_{j=1}^p (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \hat{\omega}_j / \sigma_{Y_j}^2\right] = \mathbb{E}\left[\sum_{j=1}^p (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \mathbf{1}_{j \in S_\lambda} / \sigma_{Y_j}^2\right] = \sum_{j=1}^p \frac{\omega_j}{\sigma_{Y_j}^2} \gamma_j^2.$$

As a result, we expect that the sRIVW approach also delivers accurate estimates of the causal parameter β . To present formal results, we first define a few quantities:

$$p_{\lambda,\mathbf{s}} = \sum_{j=1}^p \omega_j, \quad \kappa_{\lambda,\mathbf{s}} = \frac{1}{p_{\lambda,\mathbf{s}}} \sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}}\right)^2 \omega_j, \quad \tilde{\kappa}_{\lambda,\mathbf{s}} = \frac{1}{p_{\lambda,\mathbf{s}}} \sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}}\right)^2 \omega_j^2.$$

Remark S.1 We have considered bounds on p_λ and κ_λ in Lemmas S.6 and S.7. To provide additional insights, consider the setting that the true SNP-exposure associations are sampled from a spike-and-slab distribution, that is, $\gamma_j / \sigma_{X_j} \sim \pi \mathcal{N}(0, \varepsilon^2) + (1 - \pi) \delta_0$ for some ε . Provided that $\varepsilon \gtrsim \lambda$, it is straightforward to show $p_{\lambda,\mathbf{s}} \gtrsim p\pi\varepsilon^2$. Similarly, we also have the following lower bound

$$\kappa_{\lambda,\mathbf{s}} \gtrsim \varepsilon^2.$$

It will also be interesting to provide a lower bound on $\tilde{\kappa}_{\lambda,\mathbf{s}}$, as we will demonstrate later that this quantity features in the asymptotic variance of the sRIVW estimator. Assuming $\varepsilon/\lambda \rightarrow \infty$, one has that

$$\tilde{\kappa}_{\lambda,\mathbf{s}} \asymp \kappa_{\lambda,\mathbf{s}} \gtrsim \varepsilon^2.$$

This result is quite intuitive: in genetic applications, despite that many instruments are weak, there always exist strong instruments whose signal strength well exceed the selection threshold λ . ■

As before, we will continue assuming that there are no dominating instruments.

Assumption S.6

$$\frac{\max_{1 \leq j \leq p} \gamma_j^2}{\sum_{1 \leq j \leq p} \gamma_j^2 \omega_j} \rightarrow 0.$$

■

The following lemma provides a characterization of the denominator.

Lemma S.15 *Assume Assumptions S.1 S.2, S.4, and S.6 hold, $p_{\lambda, s} \rightarrow \infty$, and $\sqrt{p_{\lambda, s}} \kappa_{\lambda, s} / \lambda \rightarrow \infty$. Then*

$$\sum_{j=1}^p (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \hat{\omega}_j = \left[\sum_{j=1}^p \gamma_j^2 \omega_j \right] (1 + o_p(1)) \asymp_p p_{\lambda, s} \nu^2 \kappa_{\lambda, s}.$$

■

To study the large-sample properties of our sRIVW estimator, we employ the following decomposition:

$$\hat{\beta}_{\text{sRIVW}} = \beta + \frac{\sum_{j=1}^p u_{j, \text{RIVW}} \hat{\omega}_j / \sigma_{Y_j}^2}{\sum_{j=1}^p (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \hat{\omega}_j / \sigma_{Y_j}^2},$$

where $u_{j, \text{RIVW}}$ is defined the same as before. As a result, it is quite straightforward to show that the numerator is mean zero via iterative expectation. In addition, we propose the following variance estimator

$$\hat{V}_{\text{sRIVW}} = \frac{\sum_{j=1}^p \left(\hat{\Gamma}_j \hat{\gamma}_{j, \text{RB}} - \hat{\beta}_{\text{sRIVW}} (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \right)^2 \hat{\omega}_j^2 / \sigma_{Y_j}^4}{\left(\sum_{j=1}^p (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \hat{\omega}_j / \sigma_{Y_j}^2 \right)^2}.$$

Theorem S.5 *Assume Assumptions S.1 S.2, S.4, and S.6 hold, $p_{\lambda, s} \rightarrow \infty$, $\tilde{\kappa}_{\lambda, s} \asymp \kappa_{\lambda, s}$, and $\kappa_{\lambda, s} / \lambda^2 \xrightarrow{P} \infty$. Then the centered and scaled sRIVW estimator is asymptotically normally distributed, that is,*

$$V_{\text{sRIVW}}^{-\frac{1}{2}} \left[\hat{\beta}_{\text{sRIVW}} - \beta \right] \xrightarrow{d} \mathcal{N}(0, 1),$$

where

$$V_{\text{sRIVW}} = \frac{\mathbb{V} \left[\sum_{j=1}^p u_{j, \text{RIVW}} \hat{\omega}_j / \sigma_{Y_j}^2 \right]}{\left(\sum_{j=1}^p \gamma_j^2 \omega_j / \sigma_{Y_j}^2 \right)^2} \asymp \frac{1}{p_{\lambda} \kappa_{\lambda}}.$$

In addition,

$$\frac{\hat{V}_{\text{sRIVW}}}{V_{\text{sRIVW}}} \xrightarrow{P} 1.$$

■

V. Correlated Instrumental Variables. Our earlier assumption requires that all the marginal SNP-exposure association estimates are independent (Assumption S.2). In practice, this is usually achieved by LD clumping. LD clumping uses p-values from GWAS to sort the SNPs by their importance. It then takes the most significant SNP and removes all other SNPs that are correlated with the selected one. The procedure continues with the next most significant SNP bypassing a significance threshold that has not been removed yet. This

p-value based LD clumping procedure, in fact, introduces a selection bias different from the one discussed in the main paper, which we shall demonstrate below. In this section, we consider a setting where SNPs form clusters, and the covariance matrices of the estimated SNP-exposure associations are known a priori. Building on the covariance information, we will be able to lift this new winner's curse as well.

We assume there are $g = 1, 2, \dots, G$ clusters, and within cluster g , SNPs are indexed by $j = 1, 2, \dots, J_g$. Therefore, the true SNP-exposure associations are denoted by $\gamma_g = (\gamma_{g1}, \dots, \gamma_{gJ_g})'$. The estimated SNP-exposure associations are just $\hat{\gamma}_g = (\hat{\gamma}_{g1}, \dots, \hat{\gamma}_{gJ_g})'$. To account for correlated estimates within each cluster, we assume

Assumption S.7 (i) For any $g \neq g'$, $\hat{\gamma}_g$ and $\hat{\gamma}_{g'}$ are independent.
(ii) For each g , $\hat{\gamma}_g \sim \mathcal{N}(\gamma_g, \Sigma_{X_g})$. ■

To avoid cumbersome notation, we focus on one cluster and drop g from the subscript. We denote the entries (variances and covariances) in the matrix, Σ_X , by $\sigma_{X_j}^2$ and $\sigma_{X_j X_{j'}}$, respectively. The corresponding correlation matrix will be represented by Ω_X with entries $\rho_{X_j X_{j'}}$. To lift the winner's curse, we also introduce pseudo SNPs. Let $\mathbf{Z} = (Z_1, \dots, Z_J)' \sim \mathcal{N}(0, \eta^2 \Omega_X)$. We note that η can be cluster specific, that is, a more appropriate notation would be η_g . We drop this dependence for simplicity.

Then in the g th cluster, we select the SNP with the most significant SNP-exposure association provided that its associated p-value is smaller than $2\Phi(-|\lambda|)$. Without loss of generality, suppose SNP j is selected, we characterize the selection event as

$$(1) \quad \mathcal{S}_{\lambda,j} = \left| \frac{\hat{\gamma}_j}{\sigma_{X_j}} + Z_j \right| \geq \max \left\{ \max_{\substack{1 \leq j' \leq J \\ j' \neq j}} \left| \frac{\hat{\gamma}_{j'}}{\sigma_{X_{j'}}} + Z_{j'} \right|, \lambda \right\}.$$

Similar to the strategy we adopted in the main paper, we again proceed with creating an initial estimator that is independent of the select event: $\mathcal{S}_{\lambda,j}$

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

Different from our proposed estimator when the IVs are independent, the challenging step, however, is to form a Rao-Blackwellized estimator. This is because projecting the initial estimator onto the selection event requires us to work out a different conditional selection likelihood function. We first introduce a few auxiliary quantities. Let

$$R_{jj'} = \frac{\eta^2 \rho_{X_j X_{j'}}}{\sigma_{X_j}} \hat{\gamma}_j - \rho_{X_j X_{j'}} Z_j + \frac{\hat{\gamma}_{j'}}{\sigma_{X_{j'}}} + Z_{j'},$$

$$\begin{aligned} \text{for } j' \neq j \quad \tilde{R}_{jj',-} &= \min \left\{ \frac{\sigma_{X_j}}{\eta^2(1 - \rho_{X_j X_{j'}})} (R_{jj} - R_{jj'}), \frac{\sigma_{X_j}}{\eta^2(1 + \rho_{X_j X_{j'}})} (R_{jj} + R_{jj'}) \right\} \\ \tilde{R}_{jj',+} &= \max \left\{ \frac{\sigma_{X_j}}{\eta^2(1 - \rho_{X_j X_{j'}})} (R_{jj} - R_{jj'}), \frac{\sigma_{X_j}}{\eta^2(1 + \rho_{X_j X_{j'}})} (R_{jj} + R_{jj'}) \right\}, \end{aligned}$$

and

$$\tilde{R}_{jj,-} = \frac{\sigma_{X_j}}{\eta^2} (R_{jj} - \lambda), \quad \tilde{R}_{jj,+} = \frac{\sigma_{X_j}}{\eta^2} (R_{jj} + \lambda).$$

Given $\tilde{R}_{jj',\pm}$, we consider the following intersection, which can be written as the union of disjoint intervals:

$$\bigcap_{j'=1}^J (-\infty, \tilde{R}_{jj',-}] \cup [\tilde{R}_{jj',+}, \infty) = \bigcup_{\ell=1}^{L_j} [a_{j\ell}, b_{j\ell}].$$

Finding the above representation (i.e., the end points $a_{j\ell}$ and $b_{j\ell}$) may seem difficult. Therefore we provide Algorithm S.1.

Algorithm S.1

Require: Fix cluster g and the selected SNP j

Require: Set $\mathcal{J} \leftarrow \{1, 2, \dots, J\}$, $\ell \leftarrow 1$, $\text{leftEnd} \leftarrow -\infty$

Require: Compute $\tilde{R}_{jj',-}$ and $\tilde{R}_{jj',+}$ for $j' = 1, 2, \dots, J$

Set $a_{j\ell} \leftarrow \max_{j' \in \mathcal{J}} \tilde{R}_{jj',+}$, $b_{j\ell} \leftarrow \infty$
 Set $\ell \leftarrow 2$

while $\mathcal{J} \neq \emptyset$ **do**

if $\min_{j' \in \mathcal{J}} \tilde{R}_{jj',-} > \text{leftEnd}$ **then**

 Set $\text{index} \leftarrow \text{argmin}_{j' \in \mathcal{J}} \tilde{R}_{jj',-}$

 Set $a_{j\ell} \leftarrow \text{leftEnd}$, $b_{j\ell} \leftarrow \tilde{R}_{j\text{index},-}$

 Set $\text{leftEnd} \leftarrow \tilde{R}_{j\text{index},+}$

 Set $\ell \leftarrow \ell + 1$

 Remove index from $\mathcal{J}$

else

 Set $\text{index} \leftarrow \text{argmin}_{j' \in \mathcal{J}} \tilde{R}_{jj',+}$

 Set $\text{leftEnd} \leftarrow \max\{\text{leftEnd}, \tilde{R}_{j\text{index},+}\}$

 Remove index from $\mathcal{J}$

end if

end while

Now we are ready to define the Rao-Blackwellized estimator after borrowing some techniques from [10], which takes the form

$$\hat{\gamma}_{j,\text{RB}} = \sigma_{X_j} \left(\frac{R_{jj}}{\eta^2 + 1} - \frac{1}{\eta} \frac{\sum_{\ell=1}^{L_j} (\phi(W_{j\ell,-}) - \phi(W_{j\ell,+}))}{\sum_{\ell=1}^{L_j} (\Phi(W_{j\ell,-}) - \Phi(W_{j\ell,+}))} \right),$$

where

$$W_{j\ell,-} = \frac{\eta}{\sigma_{X_j}} b_{j\ell} - \frac{\eta}{1 + \eta^2} R_{jj}, \quad W_{j\ell,+} = \frac{\eta}{\sigma_{X_j}} a_{j\ell} - \frac{\eta}{1 + \eta^2} R_{jj}.$$

We show in the lemma below that the above estimator is unbiased for γ_j conditional on the selection event.

Lemma S.16 Assume Assumption S.7 holds. Then

$$\mathbb{E}[\hat{\gamma}_{j,\text{RB}} | \text{SNP } j \text{ is selected}] = \gamma_j.$$

■

In Section VIII.1, we conduct real data based simulation to showcase the unbiasedness result established in the lemma above.

VI. Auxiliary Lemmas.

Lemma S.17 (i) For all $0 < b \leq \Phi(-1)$,

$$|\Phi^{-1}(b)| \leq \sqrt{2 \log \left(\frac{1}{\sqrt{2\pi}b} \right)}.$$

(ii) For all $0 < b \leq \Phi(-\sqrt{2})$,

$$|\Phi^{-1}(b)| \geq \sqrt{\frac{2}{3} \log \left(\frac{1}{\sqrt{2\pi}b} \right)}.$$

(iii) Let $a \geq 1$. Then for all $0 < b \leq \Phi(-\sqrt{2a})$,

$$\frac{\sqrt{a}}{4} \left(\frac{1}{\log \left(\frac{1}{\sqrt{2\pi}b} \right)} \right)^{\frac{a-1}{2a}} b^{\frac{1}{a}} \leq \Phi \left(\frac{1}{\sqrt{a}} \Phi^{-1}(b) \right) \leq 3\sqrt{a} \left(\frac{1}{\log \left(\frac{1}{\sqrt{2\pi}b} \right)} \right)^{\frac{a-1}{2a}} b^{\frac{1}{a}}.$$

■

Lemma S.18 Let $Z \sim \mathcal{N}(\mu, \sigma^2)$, and $b > a$ be two real numbers. Then

$$\begin{aligned} \mathbb{E}[Z|a \leq Z \leq b] &= \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ \mathbb{E}[Z^2|a \leq Z \leq b] &= \mu^2 + \sigma^2 + \sigma^2 \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} + 2\mu\sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ \mathbb{E}[Z^3|a \leq Z \leq b] &= \mu^3 + 3\mu\sigma^2 + (2\sigma^3 + 3\mu^2\sigma) \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 3\mu\sigma^2 \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} + \sigma^3 \frac{(\frac{a-\mu}{\sigma})^2 \phi(\frac{a-\mu}{\sigma}) - (\frac{b-\mu}{\sigma})^2 \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ \mathbb{E}[Z^4|a \leq Z \leq b] &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 + 4(\mu^3\sigma + 2\mu\sigma^3) \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 3(2\mu^2\sigma^2 + 3\sigma^4) \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} + 4\mu\sigma^3 \frac{(\frac{a-\mu}{\sigma})^2 \phi(\frac{a-\mu}{\sigma}) - (\frac{b-\mu}{\sigma})^2 \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + \sigma^4 \frac{(\frac{a-\mu}{\sigma})^3 \phi(\frac{a-\mu}{\sigma}) - (\frac{b-\mu}{\sigma})^3 \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}. \end{aligned}$$

In addition,

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

■

Lemma S.19 Let $Z \sim \mathcal{N}(\mu, \sigma^2)$, and $b > a$ be two real numbers. Then

$$\begin{aligned} \mathbb{E}[Z|Z \leq a \text{ or } Z \geq b] &= \mu + \sigma \frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} \\ \mathbb{E}[Z^2|Z \leq a \text{ or } Z \geq b] &= \mu^2 + \sigma^2 + 2\mu\sigma \frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} + \sigma^2 \frac{\frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma}) - \frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} \\ \mathbb{E}[Z^3|Z \leq a \text{ or } Z \geq b] &= \mu^3 + 3\mu\sigma^2 + (2\sigma^3 + 3\mu^2\sigma) \frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 3\mu\sigma^2 \frac{\frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma}) - \frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} + \sigma^3 \frac{(\frac{b-\mu}{\sigma})^2 \phi(\frac{b-\mu}{\sigma}) - (\frac{a-\mu}{\sigma})^2 \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} \end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Z^4 | Z \leq a \text{ or } Z \geq b] &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 + 4(\mu^3\sigma + 2\mu\sigma^3) \frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} \\
&\quad + 3(2\mu^2\sigma^2 + 3\sigma^4) \frac{\frac{b-\mu}{\sigma}\phi(\frac{b-\mu}{\sigma}) - \frac{a-\mu}{\sigma}\phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} + 4\mu\sigma^3 \frac{(\frac{b-\mu}{\sigma})^2\phi(\frac{b-\mu}{\sigma}) - (\frac{a-\mu}{\sigma})^2\phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})} \\
&\quad + \sigma^4 \frac{(\frac{b-\mu}{\sigma})^3\phi(\frac{b-\mu}{\sigma}) - (\frac{a-\mu}{\sigma})^3\phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{b-\mu}{\sigma}) + \Phi(\frac{a-\mu}{\sigma})}.
\end{aligned}$$

In addition,

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

■

VII. Proofs.

VII.1. Proof of Lemma S.1.

PROOF. Again we write $\hat{\gamma}_j = \gamma_j + u_{X_j}$, which implies

$$\sum_{j=1}^p \frac{\hat{\gamma}_j^2}{\sigma_{Y_j}^2} = \sum_{j=1}^p \frac{(\gamma_j + u_{X_j})^2}{\sigma_{Y_j}^2} = \sum_{j=1}^p \frac{\gamma_j^2}{\sigma_{Y_j}^2} + \sum_{j=1}^p \frac{u_{X_j}^2}{\sigma_{Y_j}^2} + 2 \sum_{j=1}^p \frac{\gamma_j u_{X_j}}{\sigma_{Y_j}^2}.$$

Asymptotic order of the second term is easily found by its expectation:

$$\sum_{j=1}^p \frac{u_{X_j}^2}{\sigma_{Y_j}^2} = O_p \left(\mathbb{E} \left[\sum_{j=1}^p \frac{u_{X_j}^2}{\sigma_{Y_j}^2} \right] \right) = O_p \left(\sum_{j=1}^p \frac{\sigma_{X_j}^2}{\sigma_{Y_j}^2} \right) = O_p(p).$$

In addition, it is easy to show that

$$\sum_{j=1}^p \frac{u_{X_j}^2}{\sigma_{Y_j}^2} = \sum_{j=1}^p \frac{\sigma_{X_j}^2}{\sigma_{Y_j}^2} + O_p(\sqrt{p}),$$

which implies that $\sum_{j=1}^p u_{X_j}^2 / \sigma_{Y_j}^2$ is concentrated around its mean as long as $p \rightarrow \infty$.

Similarly, asymptotic order of the third term is given by the square root of its variance:

$$2 \sum_{j=1}^p \frac{\gamma_j u_{X_j}}{\sigma_{Y_j}^2} = O_p \left(\sqrt{\mathbb{V} \left[\sum_{j=1}^p \frac{\gamma_j u_{X_j}}{\sigma_{Y_j}^2} \right]} \right) = O_p \left(\sqrt{\sum_{j=1}^p \frac{\gamma_j^2 \sigma_{X_j}^2}{\sigma_{Y_j}^4}} \right) = O_p(\sqrt{p\kappa}),$$

which clearly has a smaller order compared to $p(\kappa + 1)$ by considering the two scenarios separately: $\kappa \rightarrow \infty$ and $\kappa \lesssim 1$. ■

VII.2. Proof of Lemma S.2.

PROOF. Although the asymptotic normality conclusion can be drawn directly from Assumption S.2, we present a proof based on verifying the Lindeberg-Feller condition. This will be helpful when the estimation error is not exactly normal, say, due to instrument selection.

First and foremost, the variance is easily found to be

$$\mathbb{V} \left[\sum_{j=1}^p \gamma_j (u_{Y_j} - \beta u_{X_j}) \right] = \sum_{j=1}^p \gamma_j^2 (\sigma_{Y_j}^2 + \beta^2 \sigma_{X_j}^2) \asymp p\nu^4 \kappa.$$

Next, we consider the Lindeberg-Feller condition:

$$\begin{aligned} & \sum_{j=1}^p \mathbb{E} \left[\frac{\gamma_j^2 (u_{Y_j}^2 + \beta^2 u_{X_j}^2)}{p\nu^4 \kappa} \mathbf{1} \left(\gamma_j^2 (u_{Y_j}^2 + \beta^2 u_{X_j}^2) > \varepsilon p\nu^4 \kappa \right) \right] \\ &= \frac{1}{p\nu^2 \kappa} \sum_{j=1}^p \gamma_j^2 \mathbb{E} \left[\frac{u_{Y_j}^2 + \beta^2 u_{X_j}^2}{\nu^2} \mathbf{1} \left(\frac{u_{Y_j}^2 + \beta^2 u_{X_j}^2}{\nu^2} > \varepsilon \frac{p\nu^2 \kappa}{\gamma_j^2} \right) \right]. \end{aligned}$$

To proceed, define

$$\varrho = \frac{\max_{1 \leq j \leq p} \gamma_j^2}{\sum_{1 \leq j \leq p} \gamma_j^2},$$

then the previous Lindeberg-Feller condition reduces to

$$\frac{1}{p\nu^2 \kappa} \sum_{j=1}^p \gamma_j^2 \mathbb{E} \left[\frac{u_{Y_j}^2 + \beta^2 u_{X_j}^2}{\nu^2} \mathbf{1} \left(\frac{u_{Y_j}^2 + \beta^2 u_{X_j}^2}{\nu^2} > \frac{\varepsilon}{\varrho} \right) \right] \rightarrow 0.$$

■

VII.3. Proof of Lemma S.3.

PROOF. We first calculate the mean of this term.

$$\mathbb{E} \left[\sum_{j=1}^p u_{X_j} (u_{Y_j} - \beta u_{X_j}) \right] = -\beta \sum_{j=1}^p \sigma_{X_j}^2.$$

The variance is also straightforward to obtain:

$$\begin{aligned} \mathbb{V} \left[\sum_{j=1}^p u_{X_j} (u_{Y_j} - \beta u_{X_j}) \right] &= \sum_{j=1}^p \mathbb{V} [u_{X_j} (u_{Y_j} - \beta u_{X_j})] \\ &= \sum_{j=1}^p \sigma_{X_j}^2 [\sigma_{Y_j}^2 + 2\beta^2 \sigma_{X_j}^2] = O(p\nu^4). \end{aligned}$$

Verifying a Lindeberg-Feller condition is also straightforward.

■

VII.4. Proof of Theorem S.1.

PROOF. This theorem follows from Lemmas S.1–S.3. We also note that the covariance between $\sum_{j=1}^p \gamma_j (u_{Y_j} - \beta u_{X_j})$ and $\sum_{j=1}^p u_{X_j} (u_{Y_j} - \beta u_{X_j})$ is

$$\begin{aligned} & \text{Cov} \left[\sum_{j=1}^p \gamma_j (u_{Y_j} - \beta u_{X_j}), \sum_{j=1}^p u_{X_j} (u_{Y_j} - \beta u_{X_j}) \right] \\ &= \sum_{j=1}^p \beta^2 \mathbb{E} [u_{X_j} (u_{X_j}^2 - \sigma_{X_j}^2)] = \sum_{j=1}^p \beta^2 \mathbb{E} [u_{X_j}^3] = 0, \end{aligned}$$

where the last equality is due to Assumption S.2.

■

VII.5. Proof of Lemma S.4.

PROOF. This follows directly from the proof of Lemma S.1.

■

VII.6. *Proof of Theorem S.2.*

PROOF. This follows from the proofs of Theorem S.1 and Lemma S.4. ■

VII.7. *Proof of Lemma S.5.*

PROOF. To start, consider the probability

$$\begin{aligned}
\mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} \leq y, S_j > 0\right] &= \mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} \leq y, \left|\frac{\hat{\gamma}_j}{\sigma_j} + Z_j\right| > \lambda\right] = \int \mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} \leq y, \left|\frac{\hat{\gamma}_j}{\sigma_j} + w\right| > \lambda\right] \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&= \int \left(\mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} \leq y, \frac{\hat{\gamma}_j}{\sigma_j} > \lambda - w\right] + \mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} \leq y, \frac{\hat{\gamma}_j}{\sigma_j} < -\lambda - w\right]\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&= \int \mathbf{1}(y \leq -\lambda - w) \mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} \leq y\right] \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&\quad + \int \mathbf{1}(-\lambda - w < y \leq \lambda - w) \mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} < -\lambda - w\right] \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&\quad + \int \mathbf{1}(y > \lambda - w) \left(\mathbb{P}\left[\lambda - w < \frac{\hat{\gamma}_j}{\sigma_j} \leq y\right] + \mathbb{P}\left[\frac{\hat{\gamma}_j}{\sigma_j} < -\lambda - w\right]\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&= \int \mathbf{1}(y \leq -\lambda - w) \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&\quad + \int \mathbf{1}(-\lambda - w < y \leq \lambda - w) \Phi\left(-\lambda - w - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&\quad + \int \mathbf{1}(y > \lambda - w) \left(\Phi\left(-\lambda - w - \frac{\gamma_j}{\sigma_j}\right) + \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) - \Phi\left(\lambda - w - \frac{\gamma_j}{\sigma_j}\right)\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw,
\end{aligned}$$

which can be further simplified as

$$\begin{aligned}
&= \int_{-\infty}^{-\lambda-y} \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&\quad + \int_{-\lambda-y}^{\lambda-y} \Phi\left(-\lambda - w - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&\quad + \int_{\lambda-y}^{\infty} \left(\Phi\left(-\lambda - w - \frac{\gamma_j}{\sigma_j}\right) + \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) - \Phi\left(\lambda - w - \frac{\gamma_j}{\sigma_j}\right)\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&= \int_{-\infty}^{-\lambda-y} \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw + \int_{\lambda-y}^{\infty} \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw \\
&\quad + \int_{-\lambda-y}^{\infty} \Phi\left(-\lambda - w - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw - \int_{\lambda-y}^{\infty} \Phi\left(\lambda - w - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw
\end{aligned}$$

Next, we compute the derivatives with respect to y , which gives

$$\begin{aligned}
\frac{d}{dy} \int_{-\infty}^{-\lambda-y} \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw &= \phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \Phi\left(\frac{-\lambda-y}{\eta_j}\right) - \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{-\lambda-y}{\eta_j}\right) \\
\frac{d}{dy} \int_{\lambda-y}^{\infty} \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw &= \phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \left(1 - \Phi\left(\frac{\lambda-y}{\eta_j}\right)\right) + \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{\lambda-y}{\eta_j}\right) \\
\frac{d}{dy} \int_{-\lambda-y}^{\infty} \Phi\left(-\lambda - w - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw &= \Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{-\lambda-y}{\eta_j}\right) \\
\frac{d}{dy} - \int_{\lambda-y}^{\infty} \Phi\left(\lambda - w - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{w}{\eta_j}\right) dw &= -\Phi\left(y - \frac{\gamma_j}{\sigma_j}\right) \frac{1}{\eta_j} \phi\left(\frac{\lambda-y}{\eta_j}\right).
\end{aligned}$$

Therefore,

$$\frac{d}{dy} \mathbb{P} \left[\frac{\hat{\gamma}_j}{\sigma_j} \leq y, S_j > 0 \right] = \phi \left(y - \frac{\gamma_j}{\sigma_j} \right) \left[\Phi \left(\frac{-\lambda - y}{\eta_j} \right) + 1 - \Phi \left(\frac{\lambda - y}{\eta_j} \right) \right].$$

■

VII.8. Proof of Lemma S.6.

PROOF. Without loss of generality, we will assume $\gamma_j \geq 0$ for all j , then

$$\Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \leq \mathbb{P}[S_j > 0] \leq 2\Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right).$$

The mean and variance of p_λ are

$$\mathbb{E}[p_\lambda] = \sum_{j=1}^p \mathbb{P}[S_j > 0] \asymp \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right),$$

and

$$\mathbb{V}[p_\lambda] = \sum_{j=1}^p \mathbb{P}[S_j > 0] (1 - \mathbb{P}[S_j > 0]) \asymp \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \left(1 - \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \right).$$

As a result,

$$p_\lambda \asymp_p \sum_{j=1}^p \mathbb{P}[S_j > 0] \asymp \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right).$$

To show the necessary condition, fix some $c > 0$, and consider

$$\begin{aligned} \mathbb{E}[p_\lambda] &= \sum_{j=1}^p \mathbb{P}[S_j > 0] \leq 2 \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \\ &= 2 \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right) \right) + 2 \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} < \lambda + \Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right) \right). \end{aligned}$$

Next, we will show that the second term cannot diverge. For this purpose, we employ lemma S.17, which implies that

$$\begin{aligned} \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} < \lambda + \Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right) \right) &\leq \Phi \left(\frac{\Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right)}{\sqrt{1 + \eta_j^2}} \right) \leq \Phi \left(\frac{\Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right)}{\sqrt{1 + \bar{\eta}^2}} \right) \\ &\leq 3\sqrt{1 + \bar{\eta}^2} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right)^{\frac{1}{1+\bar{\eta}^2}} \asymp p^{-1}. \end{aligned}$$

As a result,

$$\mathbb{E}[p_\lambda] \lesssim \sum_{j=1}^p \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c}{p^{1+\bar{\eta}^2}} \right) \right) + 1.$$

Now consider the other direction.

$$\begin{aligned}
\mathbb{E}[p_\lambda] &= \sum_{j=1}^p \mathbb{P}[S_j > 0] \geq \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \\
&= \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right) \right) + \sum_{j=1}^p \Phi \left(\frac{\frac{\gamma_j}{\sigma_{X_j}} - \lambda}{\sqrt{1 + \eta_j^2}} \right) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} < \lambda + \Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right) \right) \\
&\geq \sum_{j=1}^p \Phi \left(\frac{\Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right)}{\sqrt{1 + \eta_j^2}} \right) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right) \right) \\
&\geq \frac{\sqrt{1 + \eta^2}}{4} \left(\frac{1}{\log \left(\frac{p^{c_2}}{c_1} \right)} \right)^{\frac{1 + \eta^2 - 1}{2(1 + \eta^2)}} \left(\frac{c_1}{p^{c_2}} \right)^{\frac{1}{1 + \eta^2}} \sum_{j=1}^p \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right) \right) \\
&\asymp \left(\frac{1}{\log(p)} \right)^{\frac{\eta^2}{2(1 + \eta^2)}} \left(\frac{1}{p} \right)^{\frac{c_2}{1 + \eta^2}} \sum_{j=1}^p \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c_1}{p^{c_2}} \right) \right).
\end{aligned}$$

■

VII.9. Proof of Lemma S.7.

PROOF. To start, consider

$$\begin{aligned}
p_\lambda \kappa_\lambda &= \sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \mathbf{1}(S_j > 0) \\
&= \sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \mathbf{1}(S_j > 0) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c}{p^{1 + \bar{\eta}^2}} \right) \right) + \sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \mathbf{1}(S_j > 0) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} < \lambda + \Phi^{-1} \left(\frac{c}{p^{1 + \bar{\eta}^2}} \right) \right) \\
&\geq \sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \mathbf{1}(S_j > 0) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c}{p^{1 + \bar{\eta}^2}} \right) \right) \\
&\geq \left[\left(\lambda + \Phi^{-1} \left(\frac{c}{p^{1 + \bar{\eta}^2}} \right) \right) \vee 0 \right]^2 \underbrace{\sum_{j=1}^p \mathbf{1}(S_j > 0) \mathbf{1} \left(\frac{\gamma_j}{\sigma_{X_j}} \geq \lambda + \Phi^{-1} \left(\frac{c}{p^{1 + \bar{\eta}^2}} \right) \right)}_{(I)}.
\end{aligned}$$

Next, the proof of Lemma S.6 suggests that, as long as $p_\lambda \xrightarrow{P} \infty$,

$$\frac{(I)}{p_\lambda} \xrightarrow{P} 1.$$

Therefore, we showed that

$$\kappa_\lambda \gtrsim_p \left[\left(\lambda + \Phi^{-1} \left(\frac{c}{p^{1 + \bar{\eta}^2}} \right) \right) \vee 0 \right]^2.$$

The second claim follows from properties of the normal quantile function. ■

VII.10. Proof of Lemma S.8.

PROOF. Consider the conditional expectation

$$\begin{aligned}
\mathbb{E}[Z_j | \hat{\gamma}_j = r, S_j > 0] &= \mathbb{E} \left[Z_j \middle| \hat{\gamma}_j = r, \left| \frac{r}{\sigma_{X_j}} + Z_j \right| > \lambda \right] = \mathbb{E} \left[Z_j \middle| \left| \frac{r}{\sigma_{X_j}} + Z_j \right| > \lambda \right] \\
&= \mathbb{E} \left[Z_j \middle| Z_j > \lambda - \frac{r}{\sigma_{X_j}} \text{ or } Z_j < -\lambda - \frac{r}{\sigma_{X_j}} \right] \\
&= \frac{\mathbb{E}[Z_j] - \mathbb{E}[Z_j | -\lambda - \frac{r}{\sigma_{X_j}} \leq Z_j \leq \lambda - \frac{r}{\sigma_{X_j}}] \mathbb{P}[-\lambda - \frac{r}{\sigma_{X_j}} \leq Z_j \leq \lambda - \frac{r}{\sigma_{X_j}}]}{\mathbb{P}[Z_j > \lambda - \frac{r}{\sigma_{X_j}} \text{ or } Z_j < -\lambda - \frac{r}{\sigma_{X_j}}]} \\
&= -\mathbb{E} \left[Z_j \middle| -\lambda - \frac{r}{\sigma_{X_j}} \leq Z_j \leq \lambda - \frac{r}{\sigma_{X_j}} \right] \frac{\mathbb{P}[-\lambda - \frac{r}{\sigma_{X_j}} \leq Z_j \leq \lambda - \frac{r}{\sigma_{X_j}}]}{\mathbb{P}[Z_j > \lambda - \frac{r}{\sigma_{X_j}} \text{ or } Z_j < -\lambda - \frac{r}{\sigma_{X_j}}]} \\
&= \eta_j \frac{\phi \left(\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) - \phi \left(-\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right)}{1 - \Phi \left(\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) + \Phi \left(-\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right)}.
\end{aligned}$$

As a result,

$$\hat{\gamma}_{j,\text{RB}} = \hat{\gamma}_j - \frac{\sigma_{X_j}}{\eta_j} \frac{\phi \left(\frac{\lambda}{\eta_j} - \frac{\hat{\gamma}_j}{\sigma_{X_j} \eta_j} \right) - \phi \left(-\frac{\lambda}{\eta_j} - \frac{\hat{\gamma}_j}{\sigma_{X_j} \eta_j} \right)}{1 - \Phi \left(\frac{\lambda}{\eta_j} - \frac{\hat{\gamma}_j}{\sigma_{X_j} \eta_j} \right) + \Phi \left(-\frac{\lambda}{\eta_j} - \frac{\hat{\gamma}_j}{\sigma_{X_j} \eta_j} \right)} = \hat{\gamma}_j - \frac{\sigma_{X_j}}{\eta_j} \frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})}.$$

■

VII.11. Proof of Lemma S.9.

PROOF. To start, we write

$$\begin{aligned}
u_{X_j,\text{RB}} &= \hat{\gamma}_{j,\text{RB}} - \gamma_j = \hat{\gamma}_j - \gamma_j - \frac{\sigma_{X_j}}{\eta_j} \frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} \\
&= u_{X_j} - \frac{\sigma_{X_j}}{\eta_j} \frac{\phi \left(\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j} \right) - \phi \left(-\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j} \right)}{1 - \Phi \left(\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j} \right) + \Phi \left(-\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j} \right)}.
\end{aligned}$$

From Lemma S.5, we also have the following conditional distribution:

$$f_{\frac{u_{X_j}}{\sigma_{X_j}} | S_j > 0}(y) = \frac{1}{\mathbb{P}[S_j > 0]} \phi(y) \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right],$$

where

$$\mathbb{P}[S_j > 0] = 1 - \Phi \left(\frac{\lambda - \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}} \right) + \Phi \left(\frac{-\lambda - \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}} \right).$$

Using integration-by-parts, it is not difficult to show that

$$\mathbb{E}[u_{X_j,\text{RB}} | S_j > 0] = 0,$$

because

$$\mathbb{E}[u_{X_j} | S_j > 0] = \frac{\sigma_{X_j}}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y \phi(y) \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right] dy$$

$$\begin{aligned}
&= \frac{\sigma_{X_j}}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right] d(-\phi(y)) \\
&= \frac{\sigma_{X_j}}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \phi(y) d \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right] \\
&= \frac{\sigma_{X_j}}{\eta_j \mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \phi(y) \left[\phi(B_{j,+}(y)) - \phi(B_{j,-}(y)) \right] dy.
\end{aligned}$$

The above also constitutes a more direct proof that $\hat{\gamma}_{j,\text{RB}}$ is unbiased.

Now we find the variance of $u_{X_j,\text{RB}}$ (which is also the variance of $\hat{\gamma}_{j,\text{RB}}$). To start,

$$\begin{aligned}
\mathbb{E}[u_{X_j}^2 | S_j > 0] &= \frac{\sigma_{X_j}^2}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y^2 \phi(y) \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right] dy \\
&= \frac{\sigma_{X_j}^2}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right] d(-\phi(y)) \\
&= \frac{\sigma_{X_j}^2}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \phi(y) d \left\{ y \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right] \right\} \\
&= \frac{\sigma_{X_j}^2}{\eta_j \mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y \phi(y) \left[\phi(B_{j,+}(y)) - \phi(B_{j,-}(y)) \right] dy \\
&+ \frac{\sigma_{X_j}^2}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \phi(y) \left[1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y)) \right] dy \\
&= \sigma_{X_j}^2 + \frac{\sigma_{X_j}^2}{\eta_j \mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y \phi(y) \left[\phi(B_{j,+}(y)) - \phi(B_{j,-}(y)) \right] dy.
\end{aligned}$$

Therefore, we can also write the variance as

$$\begin{aligned}
\mathbb{E}[u_{X_j,\text{RB}}^2 | S_j > 0] &= \sigma_{X_j}^2 - \frac{\sigma_{X_j}^2}{\eta_j \mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y \phi(y) \left[\phi(B_{j,+}(y)) - \phi(B_{j,-}(y)) \right] dy \\
&+ \frac{\sigma_{X_j}^2}{\eta_j^2 \mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \phi(y) \frac{\left[\phi(B_{j,+}(y)) - \phi(B_{j,-}(y)) \right]^2}{1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y))} dy.
\end{aligned}$$

■

VII.12. Proof of Lemma S.10.

PROOF. Recall that the bias-corrected estimator is the following conditional expectation:

$$\hat{\gamma}_{j,\text{RB}} = \sigma_{X_j} \mathbb{E} \left[\hat{b}_{j,\text{ini}} \mid \hat{\gamma}_j, S_j > 0 \right].$$

We also note that

$$\mathbb{V} \left[\hat{b}_{j,\text{ini}} \mid S_j > 0 \right] = \mathbb{V} \left[\mathbb{E} \left[\hat{b}_{j,\text{ini}} \mid \hat{\gamma}_j, S_j > 0 \right] \mid S_j > 0 \right] + \mathbb{E} \left[\mathbb{V} \left[\hat{b}_{j,\text{ini}} \mid \hat{\gamma}_j, S_j > 0 \right] \mid S_j > 0 \right].$$

As a result,

$$\mathbb{V} \left[\hat{\gamma}_{j,\text{RB}} \mid S_j > 0 \right] = \sigma_{X_j}^2 \left(\mathbb{V} \left[\hat{b}_{j,\text{ini}} \mid S_j > 0 \right] - \mathbb{E} \left[\mathbb{V} \left[\hat{b}_{j,\text{ini}} \mid \hat{\gamma}_j, S_j > 0 \right] \mid S_j > 0 \right] \right).$$

To find the above, we first note that $\hat{b}_{j,\text{ini}}$ is independent of S_j , and hence

$$\mathbb{V} \left[\hat{b}_{j,\text{ini}} \mid S_j > 0 \right] = \mathbb{V} \left[\hat{b}_{j,\text{ini}} \right] = 1 + \frac{1}{\eta_j^2}.$$

Next,

$$\mathbb{V} \left[\widehat{b}_{j,\text{ini}} \mid \widehat{\gamma}_j = r, S_j > 0 \right] = \mathbb{V} \left[\frac{\widehat{\gamma}_j}{\sigma_{X_j}} - \frac{Z_j}{\eta_j^2} \mid \widehat{\gamma}_j = r, S_j > 0 \right] = \frac{1}{\eta_j^4} \mathbb{V} \left[Z_j \mid \widehat{\gamma}_j = r, S_j > 0 \right].$$

The conditional variance above is

$$\begin{aligned} \mathbb{V} \left[Z_j \mid \widehat{\gamma}_j = r, S_j > 0 \right] &= \mathbb{V} \left[Z_j \mid \left| Z_j + \frac{r}{\sigma_{X_j}} \right| > \lambda \right] = \mathbb{V} \left[Z_j \mid Z_j > \lambda - \frac{r}{\sigma_{X_j}} \text{ or } Z_j < -\lambda - \frac{r}{\sigma_{X_j}} \right] \\ &= \eta_j^2 \left[1 + \frac{\left(\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) \phi \left(\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) - \left(-\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) \phi \left(-\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right)}{1 - \Phi \left(\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) + \Phi \left(-\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right)} \right. \\ &\quad \left. - \left(\frac{\phi \left(\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) - \phi \left(-\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right)}{1 - \Phi \left(\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right) + \Phi \left(-\frac{\lambda}{\eta_j} - \frac{r}{\sigma_{X_j} \eta_j} \right)} \right)^2 \right]. \end{aligned}$$

■

VII.13. Proof of Lemma S.11.

PROOF OF THE UPPER BOUND. The upper bound is fairly easy to establish using the conditional variance decomposition formula. In particular,

$$\begin{aligned} \sigma_{X_j}^2 \mathbb{V}[\widehat{b}_{j,\text{ini}} | S_j > 0] &= \sigma_{X_j}^2 \mathbb{V}[\widehat{b}_{j,\text{ini}}] = \sigma_{X_j}^2 \left(1 + \frac{1}{\eta_j^2} \right) \\ &= \sigma_{X_j}^2 \mathbb{V}[\mathbb{E}[\widehat{b}_{j,\text{ini}} | \widehat{\gamma}_j, S_j > 0] | S_j > 0] + \sigma_{X_j}^2 \mathbb{E}[\mathbb{V}[\widehat{b}_{j,\text{ini}} | \widehat{\gamma}_j, S_j > 0] | S_j > 0]. \end{aligned}$$

As a result, we have

$$\mathbb{V}[\widehat{\gamma}_{j,\text{RB}} | S_j > 0] \leq \sigma_{X_j}^2 \left(1 + \frac{1}{\eta_j^2} \right).$$

■

PROOF OF THE LOWER BOUND. To provide a lower bound, we consider the quantity

$$\begin{aligned} \sigma_{X_j}^2 \mathbb{E}[\mathbb{V}[\widehat{b}_{j,\text{ini}} | \widehat{\gamma}_j, S_j > 0] | S_j > 0] &= \frac{\sigma_{X_j}^2}{\eta_j^2} \mathbb{E} \left[\mathbb{V} \left[\frac{Z_i}{\eta_j} \mid \widehat{\gamma}_j, S_j > 0 \right] \mid S_j > 0 \right] \\ &= \frac{\sigma_{X_j}^2}{\eta_j^2} \mathbb{E} \left[1 + \frac{A_{j,+} \phi(A_{j,+}) + -A_{j,-} \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} - \left(\frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} \right)^2 \mid S_j > 0 \right]. \end{aligned}$$

Note that when γ_j/σ_{X_j} is large, the conditional distribution of Z_j/η_j will be very close to its unconditional distribution. Therefore, it suffices to focus on the $\gamma_j = 0$ scenario. In addition, we assume $\eta_j = 1$ in the following calculation for ease of presentation.

To start, recall from Lemma S.5 that the conditional density of $R_j = \widehat{\gamma}_j/\sigma_{X_j}$ is

$$f_{\frac{\widehat{\gamma}_j}{\sigma_{X_j}} | S_j > 0}(r) = \frac{1}{\mathbb{P}[S_j > 0]} \phi(r) \left[\Phi \left(\frac{-\lambda - r}{\eta_j} \right) + 1 - \Phi \left(\frac{\lambda - r}{\eta_j} \right) \right], \quad \mathbb{P}[S_j > 0] = 2\Phi \left(-\frac{\lambda}{\sqrt{2}} \right).$$

For λ large, it suffices to consider the approximation to the probability $\Phi \left(-\frac{\lambda}{\sqrt{2}} \right)$

$$\frac{\frac{\sqrt{2}}{\lambda} \phi \left(\frac{\lambda}{\sqrt{2}} \right)}{\Phi \left(-\frac{\lambda}{\sqrt{2}} \right)} = 1 + o(1).$$

Then

$$\begin{aligned}
 \text{(I)} &= \mathbb{E} \left[\frac{(\lambda - R_j)\phi(\lambda - R_j) + (\lambda + R_j)\phi(\lambda + R_j)}{1 - \Phi(\lambda - R_j) + \Phi(-\lambda - R_j)} \middle| S_j > 0 \right] \\
 &= \frac{2}{\mathbb{P}[S > 0]} \int_0^\infty [(\lambda - r)\phi(\lambda - r) + (\lambda + r)\phi(\lambda + r)] \phi(r) \, dr \\
 &= \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_0^\infty (\lambda - r)\phi(\lambda - r)\phi(r) \, dr}_{\text{(I.1)}} + \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_0^\infty (\lambda + r)\phi(\lambda + r)\phi(r) \, dr}_{\text{(I.2)}}.
 \end{aligned}$$

It is easy to show that (I.2) takes the form

$$\text{(I.2)} = \frac{\lambda^2 \Phi\left(-\frac{\lambda}{\sqrt{2}}\right)}{4} + \frac{\lambda e^{-\frac{\lambda^2}{4}}}{4\sqrt{\pi}} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty.$$

It is also straightforward to find a closed-form solution for (I.1). For reasons that will become clear later, however, we will consider the following decomposition

$$\begin{aligned}
 \text{(I.1)} &= \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r)\phi(\lambda - r)\phi(r) \, dr}_{\text{(I.1.1)}} \\
 &\quad + \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_0^{\lambda/10} (\lambda - r)\phi(\lambda - r)\phi(r) \, dr}_{\text{(I.1.2)}} + \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_{9\lambda/10}^\infty (\lambda - r)\phi(\lambda - r)\phi(r) \, dr}_{\text{(I.1.3)}}.
 \end{aligned}$$

It is straightforward to verify the following

$$\text{(I.1.2)} = \frac{\lambda e^{-\frac{\lambda^2}{4}} \left(\sqrt{\pi} \lambda e^{\frac{\lambda^2}{4}} \left(\Phi\left(\frac{\lambda}{\sqrt{2}}\right) - \Phi\left(\frac{2\sqrt{2}\lambda}{5}\right) \right) + e^{\frac{9\lambda^2}{100}} - 1 \right)}{4\sqrt{\pi}} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty$$

$$\text{(I.1.3)} = \frac{\lambda e^{-\frac{4\lambda^2}{25}} \left(\sqrt{\pi} \lambda e^{\frac{4\lambda^2}{25}} \Phi\left(-\frac{2\sqrt{2}\lambda}{5}\right) - 1 \right)}{4\sqrt{\pi}} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty.$$

As a result, we have shown that

$$\text{(I)} = \text{(I.1.1)} + o(1) = \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r)\phi(\lambda - r)\phi(r) \, dr + o(1).$$

Now consider the other term in the original expectation:

$$\begin{aligned}
 \text{(II)} &= \mathbb{E} \left[\left(\frac{\phi(\lambda - R_j) - \phi(\lambda + R_j)}{1 - \Phi(\lambda - R_j) + \Phi(-\lambda - R_j)} \right)^2 \middle| S_j > 0 \right] \\
 &= \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} [\phi(\lambda - r) - \phi(\lambda + r)]^2 \frac{1}{1 - \Phi(\lambda - r) + \Phi(-\lambda - r)} \phi(r) \, dr}_{\text{(II.1)}} \\
 &\quad + \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_0^{\lambda/10} [\phi(\lambda - r) - \phi(\lambda + r)]^2 \frac{1}{1 - \Phi(\lambda - r) + \Phi(-\lambda - r)} \phi(r) \, dr}_{\text{(II.2)}} \\
 &\quad + \underbrace{\frac{2}{\mathbb{P}[S > 0]} \int_{9\lambda/10}^\infty [\phi(\lambda - r) - \phi(\lambda + r)]^2 \frac{1}{1 - \Phi(\lambda - r) + \Phi(-\lambda - r)} \phi(r) \, dr}_{\text{(II.3)}}
 \end{aligned}$$

In the following, we will show that both (II.2) and (II.3) are negligible. To this end, we make an important observation:

$$\frac{\partial}{\partial r} \left[1 - \Phi(\lambda - r) + \Phi(-\lambda - r) \right] = \phi(\lambda - r) - \phi(-\lambda - r).$$

The above is nonnegative for all $r \geq 0$, which means $1 - \Phi(\lambda - r) + \Phi(-\lambda - r)$ is non-decreasing in r for $r \geq 0$. Now we bound (II.2) as

$$\begin{aligned} \text{(II.2)} &\leq \frac{1}{2\Phi(-\lambda)} \frac{2}{\mathbb{P}[S > 0]} \int_0^{\lambda/10} \left[\phi(\lambda - r) - \phi(\lambda + r) \right]^2 \phi(r) dr \\ &= (1 + o(1)) \frac{\lambda^2 e^{-\frac{\lambda^2}{4}} \left(e^{\frac{2\lambda^2}{3}} \left(\Phi\left(\frac{23\lambda}{10\sqrt{3}}\right) - \Phi\left(\frac{17\lambda}{10\sqrt{3}}\right) \right) - 2\Phi\left(\frac{\sqrt{3}}{10}\lambda\right) + 1 \right)}{2\sqrt{6}} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty, \end{aligned}$$

and

$$\begin{aligned} \text{(II.3)} &\leq \frac{1}{\Phi(-\lambda/10)} \frac{2}{\mathbb{P}[S > 0]} \int_{9\lambda/10}^{\infty} \left[\phi(\lambda - r) - \phi(\lambda + r) \right]^2 \phi(r) dr \\ &\leq 4\phi(0)^2 \frac{1}{\Phi(-\lambda/10)} \frac{2}{\mathbb{P}[S > 0]} \Phi\left(-\frac{9\lambda}{10}\right) \\ &= (1 + o(1)) \frac{8}{9} \phi(0)^2 \frac{\phi\left(\frac{9\lambda}{10}\right)}{\phi(\lambda/10)} \frac{1}{\mathbb{P}[S > 0]} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty. \end{aligned}$$

Therefore, it suffices to consider

$$\text{(II)} = \text{(II.1)} + o(1) = \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} \left[\phi(\lambda - r) - \phi(\lambda + r) \right]^2 \frac{1}{1 - \Phi(\lambda - r) + \Phi(-\lambda - r)} \phi(r) dr + o(1).$$

For $r \in [\lambda/10, 9\lambda/10]$, one has

$$1 - \Phi(\lambda - r) + \Phi(-\lambda - r) \leq \frac{1}{\lambda - r} \phi(\lambda - r) + \frac{1}{\phi(\lambda + r)} \phi(\lambda + r),$$

which means

$$\begin{aligned} \text{(II.1)} &\geq \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} \left[\phi(\lambda - r) - \phi(\lambda + r) \right]^2 \frac{1}{\frac{1}{\lambda - r} \phi(\lambda - r) + \frac{1}{\phi(\lambda + r)} \phi(\lambda + r)} \phi(r) dr \\ &= \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r) \left[\phi(\lambda - r) - \phi(\lambda + r) \right] \frac{\phi(\lambda - r) - \phi(\lambda + r)}{\phi(\lambda - r) + \frac{\lambda - r}{\lambda + r} \phi(\lambda + r)} \phi(r) dr. \end{aligned}$$

To make further progress, we note that

$$\frac{\phi(\lambda + r)}{\phi(\lambda - r)} = e^{-\lambda r},$$

which means

$$\begin{aligned} \text{(II.1)} &\geq \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r) \left[\phi(\lambda - r) - \phi(\lambda + r) \right] \frac{1 - e^{-\lambda r}}{1 + \frac{\lambda - r}{\lambda + r} e^{-\lambda r}} \phi(r) dr \\ &\geq \frac{1 - e^{-\lambda^2/10}}{1 + \frac{9}{11} e^{-\lambda^2/10}} \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r) \left[\phi(\lambda - r) - \phi(\lambda + r) \right] \phi(r) dr \\ &= \underbrace{\text{(I.1.1)} - \frac{1 - e^{-\lambda^2/10}}{1 + \frac{9}{11} e^{-\lambda^2/10}} \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r) \phi(\lambda + r) \phi(r) dr}_{\text{(II.1.1)}} \end{aligned}$$

$$+ \underbrace{\frac{\frac{10}{11}e^{-\lambda^2/10}}{1 + \frac{9}{11}e^{-\lambda^2/10}} \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r)\phi(\lambda - r)\phi(r) dr}_{(\text{II.1.2})}.$$

We will demonstrate in the following that both (II.1.1) and (II.1.2) are negligible.

$$\begin{aligned} (\text{II.1.1}) &= (1 + o(1)) \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r)\phi(\lambda + r)\phi(r) dr \\ &= (1 + o(1)) \frac{\lambda e^{-\frac{49\lambda^2}{25}} \left(3\sqrt{\pi}\lambda e^{\frac{49\lambda^2}{25}} \left(\Phi\left(\frac{7\sqrt{2}\lambda}{5}\right) - \Phi\left(\frac{3\sqrt{2}\lambda}{5}\right) \right) - 1e^{\frac{8\lambda^2}{5}} + 1 \right)}{4\sqrt{\pi}} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty, \end{aligned}$$

and

$$\begin{aligned} (\text{II.1.2}) &= (1 + o(1)) \frac{10}{11} e^{-\lambda^2/10} \frac{2}{\mathbb{P}[S > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - r)\phi(\lambda - r)\phi(r) dr \\ &= (1 + o(1)) \frac{1}{4} \lambda^2 e^{-\frac{\lambda^2}{10}} \left(2\Phi\left(\frac{2\sqrt{2}\lambda}{5}\right) - 1 \right) \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty. \end{aligned}$$

To summarize, we showed that

$$\frac{\sigma_{X_j}^2}{\eta_j^2} \mathbb{E} \left[\mathbb{V} \left[\frac{Z_j}{\eta_j} \middle| \hat{\gamma}_j, S_j > 0 \right] \middle| S_j > 0 \right] = \frac{\sigma_{X_j}^2}{\eta_j^2} [1 + (\text{I}) - (\text{II})] = \frac{\sigma_{X_j}^2}{\eta_j^2} [1 + (\text{I.1.1}) - (\text{II.1}) + o(1)] \leq \frac{\sigma_{X_j}^2}{\eta_j^2} [1 + o(1)].$$

■

VII.14. Proof of Lemma S.12.

PROOF. First,

$$\hat{\sigma}_{X_j, \text{RB}}^2 - \sigma_{X_j, \text{RB}}^2 = \frac{\sigma_{X_j}^2}{\eta_j^4} \left(\mathbb{V}[Z_j | \hat{\gamma}_j, S_j > 0] - \mathbb{E}[\mathbb{V}[Z_j | \hat{\gamma}_j, S_j > 0] | S_j > 0] \right).$$

Therefore, we have the following bound

$$\begin{aligned} \left| \sum_{j \in S_\lambda} \left(\hat{\sigma}_{X_j, \text{RB}}^2 - \sigma_{X_j, \text{RB}}^2 \right) \right| &= O_{\text{P}} \left(\sqrt{\sum_{j \in S_\lambda} \mathbb{E} \left[\left(\hat{\sigma}_{X_j, \text{RB}}^2 - \sigma_{X_j, \text{RB}}^2 \right)^2 \middle| S_j > 0 \right]} \right) \\ &= \frac{\sigma_{X_j}^2}{\eta_j^4} O_{\text{P}} \left(\sqrt{\sum_{j \in S_\lambda} \mathbb{E} \left[\left(\mathbb{V}[Z_j | \hat{\gamma}_j, S_j > 0] - \mathbb{E}[\mathbb{V}[Z_j | \hat{\gamma}_j, S_j > 0] | S_j > 0] \right)^2 \middle| S_j > 0 \right]} \right) \\ &= \frac{\sigma_{X_j}^2}{\eta_j^4} O_{\text{P}} \left(\sqrt{\sum_{j \in S_\lambda} \mathbb{E} \left[\left(\mathbb{V}[Z_j | \hat{\gamma}_j, S_j > 0] \right)^2 \middle| S_j > 0 \right]} \right) \\ &= \frac{\sigma_{X_j}^2}{\eta_j^4} O_{\text{P}} \left(\sqrt{\sum_{j \in S_\lambda} \mathbb{E} \left[\left(Z_j - \mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0] \right)^4 \middle| S_j > 0 \right]} \right). \end{aligned}$$

The last line follows from Jensen's inequality and the law of iterated expectation:

$$\begin{aligned} \mathbb{E} \left[\left(\mathbb{V}[Z_j | \hat{\gamma}_j, S_j > 0] \right)^2 \middle| S_j > 0 \right] &= \mathbb{E} \left[\left(\mathbb{E}[(Z_j - \mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0])^2 | \hat{\gamma}_j, S_j > 0] \right)^2 \middle| S_j > 0 \right] \\ &\leq \mathbb{E} \left[\mathbb{E}[(Z_j - \mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0])^4 | \hat{\gamma}_j, S_j > 0] \middle| S_j > 0 \right] = \mathbb{E} \left[\left(Z_j - \mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0] \right)^4 \middle| S_j > 0 \right]. \end{aligned}$$

To proceed, we first compute the conditional fourth central moment of Z_j . For convenience, let $R_j = \hat{\gamma}_j / \sigma_{X_j}$, and

$$m_\ell(r) = \frac{\left(\frac{\lambda}{\eta_j} - \frac{r}{\eta_j}\right)^\ell \phi\left(\frac{\lambda}{\eta_j} - \frac{r}{\eta_j}\right) - \left(-\frac{\lambda}{\eta_j} - \frac{r}{\eta_j}\right)^\ell \phi\left(\frac{\lambda}{\eta_j} + \frac{r}{\eta_j}\right)}{1 - \Phi\left(\frac{\lambda}{\eta_j} - \frac{r}{\eta_j}\right) + \Phi\left(-\frac{\lambda}{\eta_j} - \frac{r}{\eta_j}\right)},$$

and we start from

$$\begin{aligned}\mathbb{E}[Z_j | R_j = r, S_j > 0] &= \eta_j m_0(r) \\ \mathbb{E}[Z_j^2 | R_j = r, S_j > 0] &= \eta_j^2 + \eta_j^2 m_1(r) \\ \mathbb{E}[Z_j^3 | R_j = r, S_j > 0] &= 2\eta_j^3 m_0(r) + \eta_j^3 m_2(r) \\ \mathbb{E}[Z_j^4 | R_j = r, S_j > 0] &= 3\eta_j^4 + 9\eta_j^4 m_1(r) + \eta_j^4 m_3(r).\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}[(Z_j - \mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0])^4 | \hat{\gamma}_j, S_j > 0] \\ &= \mathbb{E}[Z_j^4 | \hat{\gamma}_j, S_j > 0] - 4\mathbb{E}[Z_j^3 | \hat{\gamma}_j, S_j > 0]\mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0] + 6\mathbb{E}[Z_j^2 | \hat{\gamma}_j, S_j > 0](\mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0])^2 \\ &\quad - 3(\mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0])^4 \\ &= \eta_j^4 \left[3 - 2m_0(R_j)^2 - 3m_0(R_j)^4 + 9m_1(R_j) + 6m_0(R_j)^2 m_1(R_j) - 4m_0(R_j)m_2(R_j) + m_3(R_j) \right].\end{aligned}$$

While the formula above seems complicated, one can adopt similar strategy used in the proof of Lemma S.11, and show that

$$\mathbb{E}[(Z_j - \mathbb{E}[Z_j | \hat{\gamma}_j, S_j > 0])^4 | S_j > 0] \lesssim \lambda^2.$$

See also the proof of Lemma S.14 for another example of fourth moment calculation. ■

VII.15. Proof of Lemma S.13.

PROOF. We write $\hat{\gamma}_{j,\text{RB}} = \gamma_j + u_{X_j,\text{RB}}$, which implies

$$\sum_{j \in \mathcal{S}_\lambda} \hat{\gamma}_{j,\text{RB}}^2 = \sum_{j \in \mathcal{S}_\lambda} (\gamma_j + u_{X_j,\text{RB}})^2 = \sum_{j \in \mathcal{S}_\lambda} \gamma_j^2 + \sum_{j \in \mathcal{S}_\lambda} u_{X_j,\text{RB}}^2 + 2 \sum_{j \in \mathcal{S}_\lambda} \gamma_j u_{X_j,\text{RB}}.$$

Asymptotic order of the second term is easily found by its expectation:

$$\sum_{j \in \mathcal{S}_\lambda} u_{X_j,\text{RB}}^2 = O_p \left(\sum_{j \in \mathcal{S}_\lambda} \mathbb{E} \left[u_{X_j,\text{RB}}^2 \middle| j \in \mathcal{S}_\lambda \right] \right) = O_p \left(\sum_{j \in \mathcal{S}_\lambda} \sigma_{X_j,\text{RB}}^2 \right) = O_p \left(p_\lambda \nu^2 \right).$$

In addition, using the same technique in the proof of Lemma S.12, it suggests that

$$\sum_{j \in \mathcal{S}_\lambda} u_{X_j,\text{RB}}^2 = \sum_{j \in \mathcal{S}_\lambda} \sigma_{X_j,\text{RB}}^2 + O_p \left(\sqrt{p_\lambda \nu^4 \lambda^2} \right).$$

Similarly, asymptotic order of the third term is given by the square root of its variance:

$$2 \sum_{j \in \mathcal{S}_\lambda} \gamma_j u_{X_j,\text{RB}} = O_p \left(\sqrt{\sum_{j \in \mathcal{S}_\lambda} \mathbb{V} \left[\gamma_j u_{X_j,\text{RB}} \middle| j \in \mathcal{S}_\lambda \right]} \right) = O_p \left(\sqrt{\sum_{j \in \mathcal{S}_\lambda} \gamma_j^2 \sigma_{X_j,\text{RB}}^2} \right) = O_p \left(\sqrt{p_\lambda \nu^4 \kappa_\lambda} \right). ■$$

VII.16. *Proof of Lemma S.14.*

PROOF. To start, we have the conditional variance of the first term

$$\mathbb{V} \left[\sum_{j \in \mathcal{S}_\lambda} \gamma_j (u_{Y_j} - \beta u_{X_j, \text{RB}}) \middle| \mathcal{S}_\lambda \right] = \sum_{j \in \mathcal{S}_\lambda} \gamma_j^2 (\sigma_{Y_j}^2 + \beta^2 \sigma_{X_j, \text{RB}}^2) \asymp_{\text{P}} p_\lambda \nu^4 \kappa_\lambda.$$

By Lemma S.12, we also have that

$$\sum_{j \in \mathcal{S}_\lambda} \beta (u_{X_j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \lesssim_{\text{P}} \sqrt{p_\lambda \nu^4 \lambda^2}.$$

To proceed, we consider the following conditional variance

$$\mathbb{V} \left[\sum_{j \in \mathcal{S}_\lambda} u_{X_j, \text{RB}} u_{Y_j} - \beta (u_{X_j, \text{RB}}^2 - \sigma_{X_j, \text{RB}}^2) \middle| \mathcal{S}_\lambda \right] = \sigma_{X_j, \text{RB}}^2 \sigma_{Y_j}^2 + \beta^2 \mathbb{V}[u_{X_j, \text{RB}}^2 | S_j > 0].$$

The above suggests that a bound for the fourth moment of $u_{X_j, \text{RB}}$ is needed, which we demonstrate below.

We write

$$\begin{aligned} u_{X_j, \text{RB}} &= \hat{\gamma}_{j, \text{RB}} - \gamma_j = \hat{\gamma}_j - \gamma_j - \frac{\sigma_{X_j}}{\eta_j} \frac{\phi(A_{j,+}) - \phi(A_{j,-})}{1 - \Phi(A_{j,+}) + \Phi(A_{j,-})} \\ &= u_{X_j} - \frac{\sigma_{X_j}}{\eta_j} \frac{\phi\left(\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j}\right) - \phi\left(-\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j}\right)}{1 - \Phi\left(\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j}\right) + \Phi\left(-\frac{\lambda}{\eta_j} - \frac{\gamma_j}{\sigma_{X_j} \eta_j} - \frac{u_{X_j}}{\sigma_{X_j} \eta_j}\right)}. \end{aligned}$$

From Lemma S.5, we also have the following conditional distribution:

$$f_{\frac{u_{X_j}}{\sigma_{X_j}} | S_j > 0}(y) = \frac{1}{\mathbb{P}[S_j > 0]} \phi(y) [1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y))],$$

where

$$\mathbb{P}[S_j > 0] = \Phi\left(\frac{-\lambda + \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}}\right) + \Phi\left(\frac{-\lambda - \frac{\gamma_j}{\sigma_{X_j}}}{\sqrt{1 + \eta_j^2}}\right).$$

Therefore,

$$\begin{aligned} \mathbb{E}[u_{X_j, \text{RB}}^4 | S_j > 0] &= \frac{\sigma_{X_j}^4}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \left(y - \frac{1}{\eta_j} \frac{\phi(B_{j,+}(y)) - \phi(B_{j,-}(y))}{1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y))} \right)^4 \\ &\quad \phi(y) [1 - \Phi(B_{j,+}(y)) + \Phi(B_{j,-}(y))] dy. \end{aligned}$$

Given that the calculations can be quite involved, we let $\eta_j = 1$ and $\gamma_j = 0$ to streamline the presentation. That is, we consider

$$\frac{\mathbb{E}[u_{X_j, \text{RB}}^4 | S_j > 0]}{\sigma_{X_j}^4} = \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \left(y - \frac{\phi(-\lambda + y) - \phi(-\lambda - y)}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right)^4 \phi(y) [\Phi(-\lambda + y) + \Phi(-\lambda - y)] dy.$$

We will further split the above into five terms

$$\begin{aligned} \text{(I)} &= \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y^4 \phi(y) [\Phi(-\lambda + y) + \Phi(-\lambda - y)] dy \\ \text{(II)} &= -\frac{4}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y^3 \frac{\phi(-\lambda + y) - \phi(-\lambda - y)}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \phi(y) [\Phi(-\lambda + y) + \Phi(-\lambda - y)] dy \end{aligned}$$

$$\begin{aligned}
(\text{III}) &= \frac{6}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y^2 \left[\frac{\phi(-\lambda + y) - \phi(-\lambda - y)}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^2 \phi(y) [\Phi(-\lambda + y) + \Phi(-\lambda - y)] dy \\
(\text{IV}) &= -\frac{4}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y \left[\frac{\phi(-\lambda + y) - \phi(-\lambda - y)}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^3 \phi(y) [\Phi(-\lambda + y) + \Phi(-\lambda - y)] dy \\
(\text{V}) &= \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \left[\frac{\phi(-\lambda + y) - \phi(-\lambda - y)}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^4 \phi(y) [\Phi(-\lambda + y) + \Phi(-\lambda - y)] dy.
\end{aligned}$$

Term (I). Using integration by parts, we have

$$\begin{aligned}
(\text{I}) &= \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} \phi(y) d \left\{ y^3 [\Phi(-\lambda + y) + \Phi(-\lambda - y)] \right\} \\
&= \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} 3y^2 \phi(y) [\Phi(-\lambda + y) + \Phi(-\lambda - y)] dy \\
&\quad + \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} y^3 \phi(y) [\phi(-\lambda + y) - \phi(-\lambda - y)] dy \\
&= 3 + \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} (y^3 + 3y) \phi(y) [\phi(-\lambda + y) - \phi(-\lambda - y)] dy.
\end{aligned}$$

Term (II). Due to symmetry, we can write

$$(\text{II}) = -\frac{8}{\mathbb{P}[S_j > 0]} \int_0^{\infty} y^3 [\phi(-\lambda + y) - \phi(-\lambda - y)] \phi(y) dy.$$

It is straightforward to show that

$$\frac{8}{\mathbb{P}[S_j > 0]} \int_0^{\infty} y^3 \phi(-\lambda - y) \phi(y) dy \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty,$$

and hence

$$(\text{II}) = o(1) - \frac{8}{\mathbb{P}[S_j > 0]} \int_0^{\infty} y^3 \phi(-\lambda + y) \phi(y) dy.$$

Using similar techniques as in the derivation of (I.2) and (I.3), one can show that

$$(\text{II}) = -\frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^3 \phi(y) \phi(-\lambda + y) dy + o(1).$$

Term (III). Due to symmetry, we can write

$$\begin{aligned}
(\text{III}) &= \frac{12}{\mathbb{P}[S_j > 0]} \int_0^{\infty} y^2 [\phi(-\lambda + y) - \phi(-\lambda - y)]^2 \frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \phi(y) dy \\
&= \underbrace{\frac{12}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 [\phi(-\lambda + y) - \phi(-\lambda - y)]^2 \frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \phi(y) dy}_{(\text{III.1})} \\
&\quad + \underbrace{\frac{12}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} y^2 [\phi(-\lambda + y) - \phi(-\lambda - y)]^2 \frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \phi(y) dy}_{(\text{III.2})} \\
&\quad + \underbrace{\frac{12}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^{\infty} y^2 [\phi(-\lambda + y) - \phi(-\lambda - y)]^2 \frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \phi(y) dy}_{(\text{III.3})}.
\end{aligned}$$

Then

$$(\text{III.2}) \leq \frac{12}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} y^2 [\phi(-\lambda + y) - \phi(-\lambda - y)]^2 \frac{1}{\Phi(-\lambda + y)} \phi(y) dy$$

$$\begin{aligned} &\leq \frac{24}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{\lambda - y}{\phi(-\lambda + y)} \phi(y) dy \\ &\rightarrow 0, \quad \text{as } \lambda \rightarrow \infty, \end{aligned}$$

and

$$\begin{aligned} \text{(III.3)} &\leq \frac{12}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^{\infty} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{1}{\Phi(-\lambda/10)} \phi(y) dy \\ &\leq \frac{24}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^{\infty} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{\lambda}{10\phi(-\lambda/10)} \phi(y) dy \\ &\rightarrow 0, \quad \text{as } \lambda \rightarrow \infty. \end{aligned}$$

Therefore,

$$\text{(III)} = \frac{12}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \phi(y) dy + o(1).$$

To further simplify,

$$\begin{aligned} &\frac{12}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \phi(y) dy \\ &= \frac{12}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{1}{\Phi(-\lambda + y)} \phi(y) dy \\ &\quad + O \left(\frac{1}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{\Phi(-\lambda - y)}{[\Phi(-\lambda + y)]^2} \phi(y) dy \right). \end{aligned}$$

The big-O term in the above has order

$$\begin{aligned} &O \left(\frac{1}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{(\lambda - y)^2}{\lambda + y} \frac{\phi(-\lambda - y)}{[\phi(-\lambda + y)]^2} \phi(y) dy \right) \\ &= O \left(\frac{\lambda}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{\phi(-\lambda - y)}{[\phi(-\lambda + y)]^2} \phi(y) dy \right) = o(1), \end{aligned}$$

so that

$$\begin{aligned} \text{(III)} &= \frac{12}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 (\lambda - y) \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{1}{\phi(-\lambda + y)} \phi(y) dy \\ &\quad + O \left(\frac{1}{\lambda \mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{1}{\phi(-\lambda + y)} \phi(y) dy \right) + o(1). \end{aligned}$$

It can be shown that the big-O term in the above has order λ^2 , which means

$$\text{(III)} = \frac{12}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 (\lambda - y) \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^2 \frac{1}{\phi(-\lambda + y)} \phi(y) dy + O(\lambda^2).$$

Term (IV). Due to symmetry, we can write

$$\begin{aligned} \text{(IV)} &= -\frac{8}{\mathbb{P}[S_j > 0]} \int_0^{\infty} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^2 \phi(y) dy \\ &= -\underbrace{\frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^2 \phi(y) dy}_{\text{(IV.1)}} \end{aligned}$$

$$\begin{aligned}
& - \underbrace{\frac{8}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^2 \phi(y) dy}_{(IV.2)} \\
& - \underbrace{\frac{8}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^{\infty} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^2 \phi(y) dy}_{(IV.3)}.
\end{aligned}$$

Next, we have

$$\begin{aligned}
(IV.2) & \leq \frac{8}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y)} \right]^2 \phi(y) dy \\
& \leq \frac{32}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{\lambda - y}{\phi(-\lambda + y)} \right]^2 \phi(y) dy \\
& \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty,
\end{aligned}$$

and

$$\begin{aligned}
(IV.3) & \leq \frac{8}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^{\infty} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda/10)} \right]^2 \phi(y) dy \\
& \leq \frac{32}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^{\infty} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{\lambda}{10\phi(-\lambda/10)} \right]^2 \phi(y) dy \\
& \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty.
\end{aligned}$$

Therefore,

$$(IV) = - \frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^2 \phi(y) dy + o(1).$$

To further simplify, we note that

$$\begin{aligned}
& \frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)} \right]^2 \phi(y) dy \\
& = \frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y)} \right]^2 \phi(y) dy \\
& + O \left(\frac{1}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \frac{\Phi(-\lambda - y)}{[\Phi(-\lambda + y)]^3} \phi(y) dy \right),
\end{aligned}$$

and the big-O term is further bounded by

$$\begin{aligned}
& O \left(\frac{1}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \frac{\Phi(-\lambda - y)}{[\Phi(-\lambda + y)]^2} \phi(y) dy \right) \\
& = O \left(\frac{\lambda^2}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \frac{\phi(-\lambda - y)}{[\phi(-\lambda + y)]^3} \phi(y) dy \right) = o(1).
\end{aligned}$$

Finally, we also have

$$\begin{aligned}
& \frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\Phi(-\lambda + y)} \right]^2 \phi(y) dy \\
& = \frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y(\lambda - y)^2 \left[\phi(-\lambda + y) - \phi(-\lambda - y) \right]^3 \left[\frac{1}{\phi(-\lambda + y)} \right]^2 \phi(y) dy
\end{aligned}$$

$$+ O\left(\frac{1}{\lambda^2 \mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y [\phi(-\lambda + y) - \phi(-\lambda - y)]^3 \left[\frac{1}{\phi(-\lambda + y)}\right]^2 \phi(y) dy\right),$$

and the big-O term in the above is actually $O(1)$, which means that

$$(IV) = -\frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y(\lambda - y)^2 [\phi(-\lambda + y) - \phi(-\lambda - y)]^3 \left[\frac{1}{\phi(-\lambda + y)}\right]^2 \phi(y) dy + O(1).$$

Term (V). Due to symmetry, we can write

$$\begin{aligned} (V) &= \frac{2}{\mathbb{P}[S_j > 0]} \int_0^\infty [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)}\right]^3 \phi(y) dy \\ &= \underbrace{\frac{2}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)}\right]^3 \phi(y) dy}_{(V.1)} \\ &\quad + \underbrace{\frac{2}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)}\right]^3 \phi(y) dy}_{(V.2)} \\ &\quad + \underbrace{\frac{2}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^\infty [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)}\right]^3 \phi(y) dy}_{(V.3)}. \end{aligned}$$

Next, we have

$$\begin{aligned} (V.2) &\leq \frac{2}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda + y)}\right]^3 \phi(y) dy \\ &\leq \frac{16}{\mathbb{P}[S_j > 0]} \int_0^{\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{\lambda - y}{\phi(-\lambda + y)}\right]^3 \phi(y) dy \\ &\rightarrow 0, \quad \text{as } \lambda \rightarrow \infty, \end{aligned}$$

and

$$\begin{aligned} (V.3) &\leq \frac{2}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^\infty [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda/10)}\right]^3 \phi(y) dy \\ &\leq \frac{16}{\mathbb{P}[S_j > 0]} \int_{9\lambda/10}^\infty [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{\lambda}{10\phi(-\lambda/10)}\right]^3 \phi(y) dy \\ &\rightarrow 0, \quad \text{as } \lambda \rightarrow \infty. \end{aligned}$$

Therefore,

$$(V) = \frac{2}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)}\right]^3 \phi(y) dy + o(1).$$

To further simplify,

$$\begin{aligned} &\frac{2}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\Phi(-\lambda + y) + \Phi(-\lambda - y)}\right]^3 \phi(y) dy \\ &= \frac{2}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - y)^3 [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\phi(-\lambda + y)}\right]^3 \phi(y) dy \\ &\quad + O\left(\frac{\lambda^3}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \frac{\phi(-\lambda - y)}{[\phi(-\lambda + y)]^4} \phi(y) dy\right) \end{aligned}$$

$$+ O\left(\frac{1}{\lambda^3 \mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\phi(-\lambda + y)}\right]^3 \phi(y) dy\right).$$

The two big-O terms above are both $o(1)$. Therefore,

$$(V) = \frac{2}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - y)^3 [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\phi(-\lambda + y)}\right]^3 \phi(y) dy + o(1).$$

Collecting pieces. Now we showed that

$$\begin{aligned} \frac{\mathbb{E}[u_{X_j, \text{RB}}^4 | S_j > 0]}{\sigma_{X_j}^4} &= O(\lambda^2) + 3 + \\ &+ \frac{1}{\mathbb{P}[S_j > 0]} \int_{-\infty}^{\infty} (y^3 + 3y) \phi(y) [\phi(-\lambda + y) - \phi(-\lambda - y)] dy \\ &- \frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^3 \phi(y) \phi(-\lambda + y) dy \\ &+ \frac{12}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y^2 (\lambda - y) [\phi(-\lambda + y) - \phi(-\lambda - y)]^2 \frac{1}{\phi(-\lambda + y)} \phi(y) dy \\ &- \frac{8}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} y (\lambda - y)^2 [\phi(-\lambda + y) - \phi(-\lambda - y)]^3 \left[\frac{1}{\phi(-\lambda + y)}\right]^2 \phi(y) dy \\ &+ \frac{2}{\mathbb{P}[S_j > 0]} \int_{\lambda/10}^{9\lambda/10} (\lambda - y)^3 [\phi(-\lambda + y) - \phi(-\lambda - y)]^4 \left[\frac{1}{\phi(-\lambda + y)}\right]^3 \phi(y) dy. \end{aligned}$$

All the integrals above have explicit solutions, and with tedious but still straightforward calculations,

$$\mathbb{E}[u_{X_j, \text{RB}}^4 | S_j > 0] \lesssim \sigma_{X_j}^4 \lambda^2.$$

■

VII.17. Proof of Theorem S.3.

PROOF. First, we can write

$$\hat{\beta}_{\text{RIVW}} - \beta = \frac{\sum_{j \in \mathcal{S}_\lambda} u_{j, \text{RIVW}} / \sigma_{Y_j}^2}{\sum_{j \in \mathcal{S}_\lambda} (\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) / \sigma_{Y_j}^2}.$$

Then it is easy to verify that the order of the asymptotic variance is

$$V_{\text{RIVW}} = \frac{\mathbb{V}\left[\sum_{j \in \mathcal{S}_\lambda} u_{j, \text{RIVW}} \middle| \mathcal{S}_\lambda\right]}{\left(\sum_{j \in \mathcal{S}_\lambda} \gamma_j^2\right)^2} \asymp_p \frac{p_\lambda \nu^4 \kappa_\lambda}{p_\lambda^2 \nu^4 \kappa_\lambda^2} \asymp_p \frac{1}{p_\lambda \kappa_\lambda}.$$

Finally, the asymptotic normality follows from the Lindeberg-Feller central limit theorem. See, for example, the proof of Lemma S.2. ■

VII.18. Proof of Theorem S.4.

PROOF. To estimate the leading variance of our RIVW estimator, that is, to provide a valid standard error construction, we first consider the following

$$\begin{aligned} \hat{\Gamma}_j \hat{\gamma}_{j, \text{RB}} - \beta(\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) &= (\hat{\gamma}_{j, \text{RB}} - u_{X_j, \text{RB}}) \beta \hat{\gamma}_{j, \text{RB}} + u_{Y_j} \hat{\gamma}_{j, \text{RB}} - \beta(\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) \\ &= \hat{\gamma}_{j, \text{RB}}^2 \beta - u_{X_j, \text{RB}} \hat{\gamma}_{j, \text{RB}} \beta - \beta(\hat{\gamma}_{j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2) + u_{Y_j} \hat{\gamma}_{j, \text{RB}} \\ &= -u_{X_j, \text{RB}} \hat{\gamma}_{j, \text{RB}} \beta + \beta \hat{\sigma}_{X_j, \text{RB}}^2 + u_{Y_j} \hat{\gamma}_{j, \text{RB}} \\ &= u_{j, \text{RIVW}}. \end{aligned}$$

Therefore, a valid estimator for $\mathbb{V} \left[\sum_{j \in \mathcal{S}_\lambda} u_{j,\text{RIVW}} \middle| \mathcal{S}_\lambda \right]$, which employs the null hypothesis, would be

$$\tilde{\mathbb{V}} \left[\sum_{j \in \mathcal{S}_\lambda} u_{j,\text{RIVW}} \middle| \mathcal{S}_\lambda \right] = \sum_{j \in \mathcal{S}_\lambda} u_{j,\text{RIVW}}^2 = \sum_{j \in \mathcal{S}_\lambda} \left[\hat{\Gamma}_j \hat{\gamma}_{j,\text{RB}} - \beta (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \right]^2.$$

It is also possible to replace β by $\hat{\beta}$, which leads to

$$\begin{aligned} \hat{\mathbb{V}} \left[\sum_{j \in \mathcal{S}_\lambda} u_{j,\text{RIVW}} \middle| \mathcal{S}_\lambda \right] &= \sum_{j \in \mathcal{S}_\lambda} \left[\hat{\Gamma}_j \hat{\gamma}_{j,\text{RB}} - \hat{\beta}_{\text{RIVW}} (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \right]^2 \\ &= \sum_{j \in \mathcal{S}_\lambda} \left[\hat{\Gamma}_j \hat{\gamma}_{j,\text{RB}} - \beta (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) + (\beta - \hat{\beta}_{\text{RIVW}}) (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \right]^2 \\ &= \tilde{\mathbb{V}} \left[\sum_{j \in \mathcal{S}_\lambda} u_{j,\text{RIVW}} \middle| \mathcal{S}_\lambda \right] + (\beta - \hat{\beta}_{\text{RIVW}})^2 \sum_{j \in \mathcal{S}_\lambda} \left[\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2 \right]^2 \\ &\quad + 2(\beta - \hat{\beta}_{\text{RIVW}}) \sum_{j \in \mathcal{S}_\lambda} \left[\hat{\Gamma}_j \hat{\gamma}_{j,\text{RB}} - \beta (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \right] \left[\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2 \right]. \end{aligned}$$

It is straightforward to show that both the second and the third term in the above are negligible relative to the leading term. $\blacksquare$

VII.19. Omitted Details of Remark S.1.

PROOF. Assume $\gamma_j / \sigma_{X_j} \sim \pi \mathcal{N}(0, \varepsilon^2) + (1 - \pi) \delta_0$, then we may compute the following expectation (for ease of presentation, we set $\eta_j = \eta$, as this will not affect the asymptotic order)

$$\begin{aligned} \mathbb{E} \left[\sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \omega_j \right] &= p\pi\varepsilon^2 \int_{-\infty}^{\infty} y^2 \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1 + \eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1 + \eta^2}} \right) \right) \phi(y) dy \\ &= p\pi\varepsilon^2 \int_{-\infty}^{\infty} \phi(y) d \left[y \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1 + \eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1 + \eta^2}} \right) \right) \right] \\ &= p\pi\varepsilon^2 \int_{-\infty}^{\infty} \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1 + \eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1 + \eta^2}} \right) \right) \phi(y) dy \\ &\quad + p\pi \frac{\varepsilon^3}{\sqrt{1 + \eta^2}} \int_{-\infty}^{\infty} y \left(\phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1 + \eta^2}} \right) - \phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1 + \eta^2}} \right) \right) \phi(y) dy. \end{aligned}$$

The first term in the above is asymptotically equivalent to $\varepsilon^2 \mathbb{E}[p_{\lambda, \mathbf{s}}]$. Since the second term is strictly positive, we have that

$$\kappa_{\lambda, \mathbf{s}} \succsim \varepsilon^2.$$

Similarly, we can compute

$$\begin{aligned} \mathbb{E} \left[\sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \omega_j^2 \right] &= p\pi\varepsilon^2 \int_{-\infty}^{\infty} y^2 \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1 + \eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1 + \eta^2}} \right) \right)^2 \phi(y) dy \\ &= p\pi\varepsilon^2 \int_{-\infty}^{\infty} \phi(y) d \left[y \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1 + \eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1 + \eta^2}} \right) \right)^2 \right] \\ &= p\pi\varepsilon^2 \int_{-\infty}^{\infty} \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1 + \eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1 + \eta^2}} \right) \right)^2 \phi(y) dy \end{aligned}$$

$$+ p\pi \frac{2\varepsilon^3}{\sqrt{1+\eta^2}} \int_{-\infty}^{\infty} y \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1+\eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1+\eta^2}} \right) \right) \left(\phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1+\eta^2}} \right) - \phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1+\eta^2}} \right) \right) \phi(y) dy.$$

If we assume $\varepsilon/\lambda \rightarrow \infty$, then the above has a lower bound of

$$\begin{aligned} \mathbb{E} \left[\sum_{j=1}^p \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \omega_j^2 \right] &\geq p\pi\varepsilon^2 \int_{(-\infty, -1) \cup (1, \infty)} \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1+\eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1+\eta^2}} \right) \right)^2 \phi(y) dy \\ &\geq \frac{p\pi\varepsilon^2}{2} \int_{(-\infty, -1) \cup (1, \infty)} \left(\Phi \left(\frac{-\lambda + \varepsilon y}{\sqrt{1+\eta^2}} \right) + \Phi \left(\frac{-\lambda - \varepsilon y}{\sqrt{1+\eta^2}} \right) \right) \phi(y) dy \gtrsim \frac{\varepsilon^2}{2} \mathbb{E}[p_{\lambda, \mathbf{s}}]. \end{aligned}$$

In other words,

$$\tilde{\kappa}_{\lambda, \mathbf{s}} \asymp \kappa_{\lambda, \mathbf{s}} \gtrsim \varepsilon^2.$$

■

VII.20. Proof of Theorem S.5.

PROOF. Recall that

$$u_{j, \text{RIVW}} = \gamma_j (u_{Y_j} - \beta u_{X_j, \text{RB}}) + (u_{X_j, \text{RB}} u_{Y_j} - \beta(u_{X_j, \text{RB}}^2 - \hat{\sigma}_{X_j, \text{RB}}^2)).$$

We consider the variance of the first term, since under our assumption it dominates in the asymptotic variance. That is, we compute

$$\mathbb{V} [\gamma_j (u_{Y_j} - \beta u_{X_j, \text{RB}}) \hat{\omega}_j] = \gamma_j^2 \mathbb{E} [(u_{Y_j} - \beta u_{X_j, \text{RB}})^2 \hat{\omega}_j^2].$$

Note that u_{Y_j} is independent of both $u_{X_j, \text{RB}}$ and $\hat{\omega}_j^2$, which means the above has a lower bound of

$$\begin{aligned} \mathbb{V} [\gamma_j (u_{Y_j} - \beta u_{X_j, \text{RB}}) \hat{\omega}_j] &= \gamma_j^2 \left(\mathbb{E} [u_{Y_j}^2 \hat{\omega}_j^2] + \beta^2 \mathbb{E} [u_{X_j, \text{RB}}^2 \hat{\omega}_j^2] \right) \\ &\geq \gamma_j^2 \mathbb{E} [u_{Y_j}^2 \hat{\omega}_j^2] = \gamma_j^2 \sigma_{Y_j}^2 \mathbb{E} [\hat{\omega}_j^2] \geq \gamma_j^2 \sigma_{Y_j}^2 \omega_j^2. \end{aligned}$$

It is also possible to provide an upper bound:

$$\begin{aligned} \mathbb{V} [\gamma_j (u_{Y_j} - \beta u_{X_j, \text{RB}}) \hat{\omega}_j] &= \gamma_j^2 \left(\mathbb{E} [u_{Y_j}^2 \hat{\omega}_j^2] + \beta^2 \mathbb{E} [u_{X_j, \text{RB}}^2 \hat{\omega}_j^2] \right) \\ &\leq \gamma_j^2 \left(\mathbb{E} [u_{Y_j}^2 \hat{\omega}_j] + \beta^2 \mathbb{E} [u_{X_j, \text{RB}}^2 \hat{\omega}_j] \right) \lesssim \gamma_j^2 \nu^2 \omega_j. \end{aligned}$$

As a result, we have

$$\mathbb{V} \left[\sum_{j=1}^p u_{j, \text{RIVW}} \hat{\omega}_j / \sigma_{Y_j}^2 \right] \asymp p_{\lambda, \mathbf{s}} \kappa_{\lambda, \mathbf{s}}.$$

With Lemma S.15, we have

$$\mathbf{V}_{\text{sRIVW}} \asymp \frac{1}{p_{\lambda, \mathbf{s}} \kappa_{\lambda, \mathbf{s}}}.$$

Both the asymptotic normality of $\hat{\beta}_{\text{sRIVW}}$ and the validity of the variance estimator $\hat{\mathbf{V}}_{\text{sRIVW}}$ can be established using similar techniques adopted in Theorems S.3 and S.4. ■

VII.21. Proof of Lemma S.15.

PROOF. To start, we write

$$\begin{aligned} \sum_{j=1}^p (\hat{\gamma}_{j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2) \hat{\omega}_j &= \sum_{j=1}^p \left((\gamma_j + u_{X_j,\text{RB}})^2 - \hat{\sigma}_{X_j,\text{RB}}^2 \right) \hat{\omega}_j \\ &= \sum_{j=1}^p \left(\gamma_j^2 + 2\gamma_j u_{X_j,\text{RB}} + u_{X_j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2 \right) \hat{\omega}_j. \end{aligned}$$

First consider the term $\sum_{j=1}^p \gamma_j^2 \hat{\omega}_j$. It is easy to verify the following crude bound:

$$\sum_{j=1}^p \gamma_j^2 \hat{\omega}_j = \sum_{j=1}^p \gamma_j^2 \omega_j + O_p \left(\sqrt{\sum_{j=1}^p \gamma_j^4 \omega_j} \right).$$

We will assume that the largest SNP-exposure association is dominated by the aggregate:

$$\max_{1 \leq j \leq p} \gamma_j^2 = o \left(\sum_{j=1}^p \gamma_j^2 \omega_j \right).$$

Then the above implies

$$\sum_{j=1}^p \gamma_j^2 \hat{\omega}_j = \left[\sum_{j=1}^p \gamma_j^2 \omega_j \right] (1 + o_p(1)).$$

Next, with the same techniques used to prove Lemma S.13, we have

$$\sum_{j=1}^p \left(2\gamma_j u_{X_j,\text{RB}} + u_{X_j,\text{RB}}^2 - \hat{\sigma}_{X_j,\text{RB}}^2 \right) \hat{\omega}_j = O_p \left(\sqrt{p_{\lambda,\mathbf{s}} \nu^4 (\kappa_{\lambda,\mathbf{s}} + \lambda^2)} \right),$$

where we recall

$$p_{\lambda,\mathbf{s}} = \sum_{1 \leq j \leq p} \omega_j, \quad \kappa_{\lambda,\mathbf{s}} = \frac{1}{p_{\lambda,\mathbf{s}}} \sum_{1 \leq j \leq p} \left(\frac{\gamma_j}{\sigma_{X_j}} \right)^2 \omega_j.$$

■

VII.22. Proof of Lemma S.16.

PROOF. Define

$$P_j = \hat{\gamma}_j + \sigma_{X_j} Z_j, \quad Q_j = \hat{\gamma}_j - \frac{\sigma_{X_j}}{\eta^2} Z_j.$$

Also define $\mathbf{P} = (P_1, P_2, \dots, P_J)'$. Note that $\mathbf{P} \sim \mathcal{N}(\boldsymbol{\gamma}, (1 + \eta^2) \boldsymbol{\Sigma}_X)$. To save notation, we will use $\boldsymbol{\Sigma}_P$ and $\boldsymbol{\Sigma}_Q$ to denote the covariances of $\mathbf{P}$ and $\mathbf{Q}$, respectively. As before, entries in the covariance matrices are $\sigma_{P_j P_{j'}}$ and $\sigma_{Q_j Q_{j'}}$.

Without loss of generality, consider the event that the first instrument is selected. We will first construct a sufficient statistic for γ_1 . Our construction follows closely the approach of [10]. The joint distribution of $\mathbf{P}$ and Q_1 given the selection event is (note that the following is also the joint distribution of $\mathbf{P}$ and Q_{g^*})

$$f_{\mathbf{P}, Q_1 | g^*=1}(\mathbf{p}, q) \propto \exp \left\{ -\frac{1}{2} (\mathbf{p} - \boldsymbol{\gamma})' \boldsymbol{\Sigma}_P^{-1} (\mathbf{p} - \boldsymbol{\gamma}) \right\} \exp \left\{ -\frac{(q - \gamma_1)^2}{2\sigma_{Q_1}^2} \right\} \psi(\mathbf{p}, q),$$

where $\psi(\mathbf{p}, q)$ captures the support of the conditional distribution, and takes the form

$$\psi(\mathbf{p}, q) = \mathbf{1} \left(\left| \frac{p_1}{\sigma_{X_1}} \right| \geq \max \left\{ \max_{2 \leq j \leq J} \left| \frac{p_j}{\sigma_{X_j}} \right|, \lambda \right\} \right).$$

By the factorization theorem, one can show that (Theorem 2.1 of Robertson, Prevost and Bowden 10) the complete and sufficient statistic for γ is

$$\mathbf{R} = (R_1, \dots, R_J)', \quad R_j = P_j + \eta^2 \frac{\sigma_{X_1 X_j}}{\sigma_{X_1}^2} Q_1.$$

Given the sufficient statistic, $\mathbf{R}$, our Rao-Blackwellized estimator for γ_1 , conditional on the selection event $g^* = 1$, is

$$\hat{\gamma}_{1, \text{RB}} = \mathbb{E}[Q_1 | \mathbf{R}, g^* = 1].$$

To find the conditional expectation, we will first compute the support. Notice that for some $j \neq 1$,

$$\left| \frac{P_1}{\sigma_{X_1}} \right| \geq \left| \frac{P_j}{\sigma_{X_j}} \right| \Leftrightarrow \left| \frac{R_1}{\sigma_{X_1}} - \eta^2 \frac{Q_1}{\sigma_{X_1}} \right| \geq \left| \frac{R_j}{\sigma_{X_j}} - \eta^2 \rho_{X_1 X_j} \frac{Q_1}{\sigma_{X_1}} \right|.$$

We will split the above into four events, $(A_{1j} \cap A_{2j}) \cup (A_{3j} \cap A_{4j})$:

$$\begin{aligned} A_{1j} : \frac{R_1}{\sigma_{X_1}} - \eta^2 \frac{Q_1}{\sigma_{X_1}} &\geq \frac{R_j}{\sigma_{X_j}} - \eta^2 \rho_{X_1 X_j} \frac{Q_1}{\sigma_{X_1}} \Leftrightarrow Q_1 \leq \frac{\sigma_{X_1}}{\eta^2(1 - \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} - \frac{R_j}{\sigma_{X_j}} \right) \\ A_{2j} : -\frac{R_1}{\sigma_{X_1}} + \eta^2 \frac{Q_1}{\sigma_{X_1}} &\leq \frac{R_j}{\sigma_{X_j}} - \eta^2 \rho_{X_1 X_j} \frac{Q_1}{\sigma_{X_1}} \Leftrightarrow Q_1 \leq \frac{\sigma_{X_1}}{\eta^2(1 + \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} + \frac{R_j}{\sigma_{X_j}} \right) \\ A_{3j} : -\frac{R_1}{\sigma_{X_1}} + \eta^2 \frac{Q_1}{\sigma_{X_1}} &\geq \frac{R_j}{\sigma_{X_j}} - \eta^2 \rho_{X_1 X_j} \frac{Q_1}{\sigma_{X_1}} \Leftrightarrow Q_1 \geq \frac{\sigma_{X_1}}{\eta^2(1 + \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} + \frac{R_j}{\sigma_{X_j}} \right) \\ A_{4j} : \frac{R_1}{\sigma_{X_1}} - \eta^2 \frac{Q_1}{\sigma_{X_1}} &\leq \frac{R_j}{\sigma_{X_j}} - \eta^2 \rho_{X_1 X_j} \frac{Q_1}{\sigma_{X_1}} \Leftrightarrow Q_1 \geq \frac{\sigma_{X_1}}{\eta^2(1 - \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} - \frac{R_j}{\sigma_{X_j}} \right). \end{aligned}$$

Therefore, we can define

$$A_j = (A_{1j} \cap A_{2j}) \cup (A_{3j} \cap A_{4j}) : Q_1 \in (-\infty, \tilde{R}_{j,-}] \cup [\tilde{R}_{j,+}, \infty),$$

and

$$\begin{aligned} \tilde{R}_{j,-} &= \min \left\{ \frac{\sigma_{X_1}}{\eta^2(1 - \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} - \frac{R_j}{\sigma_{X_j}} \right), \frac{\sigma_{X_1}}{\eta^2(1 + \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} + \frac{R_j}{\sigma_{X_j}} \right) \right\} \\ \tilde{R}_{j,+} &= \max \left\{ \frac{\sigma_{X_1}}{\eta^2(1 - \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} - \frac{R_j}{\sigma_{X_j}} \right), \frac{\sigma_{X_1}}{\eta^2(1 + \rho_{X_1 X_j})} \left(\frac{R_1}{\sigma_{X_1}} + \frac{R_j}{\sigma_{X_j}} \right) \right\}. \end{aligned}$$

We also need to take into account the selection via the threshold λ , which we will denote by $A_1 = A_{11} \cup A_{21}$:

$$\begin{aligned} A_{11} : \frac{R_1}{\sigma_{X_1}} - \eta^2 \frac{Q_1}{\sigma_{X_1}} &\geq \lambda \Leftrightarrow Q_1 \leq \frac{\sigma_{X_1}}{\eta^2} \left(\frac{R_1}{\sigma_{X_1}} - \lambda \right) \\ A_{21} : -\frac{R_1}{\sigma_{X_1}} + \eta^2 \frac{Q_1}{\sigma_{X_1}} &\geq \lambda \Leftrightarrow Q_1 \geq \frac{\sigma_{X_1}}{\eta^2} \left(\frac{R_1}{\sigma_{X_1}} + \lambda \right). \end{aligned}$$

Therefore, we have

$$A_1 = A_{11} \cup A_{21} : Q_1 \in (-\infty, \tilde{R}_{1,-}] \cup [\tilde{R}_{1,+}, \infty),$$

and

$$\begin{aligned} \tilde{R}_{1,-} &= \frac{\sigma_{X_1}}{\eta^2} \left(\frac{R_1}{\sigma_{X_1}} - \lambda \right) \\ \tilde{R}_{1,+} &= \frac{\sigma_{X_1}}{\eta^2} \left(\frac{R_1}{\sigma_{X_1}} + \lambda \right). \end{aligned}$$

Given $\tilde{R}_{j,\pm}$ for $j = 1, 2, \dots, J$, it is straightforward to see that the intersection can be written as the union of disjoint intervals:

$$\bigcap_{j=1}^J (-\infty, \tilde{R}_{j,-}] \cup [\tilde{R}_{j,+}, \infty) = \bigcup_{\ell=1}^L [a_\ell, b_\ell].$$

The final Rao-Blackwellized estimator is thus (Theorem 4.1 of Robertson, Prevost and Bowden 10)

$$\hat{\gamma}_{1,\text{RB}} = \mathbb{E}[Q_1 | \mathbf{R}, g^* = 1] = \frac{R_1}{\eta^2 + 1} - \frac{\sigma_{X_1}}{\eta} \frac{\sum_{\ell=1}^L (\phi(W_{\ell,-}) - \phi(W_{\ell,+}))}{\sum_{\ell=1}^L (\Phi(W_{\ell,-}) - \Phi(W_{\ell,+}))},$$

where

$$W_{\ell,-} = \frac{\eta}{\sigma_{X_1}} b_\ell - \frac{\eta}{\sigma_{X_1}(1 + \eta^2)} R_1, \quad W_{\ell,+} = \frac{\eta}{\sigma_{X_1}} a_\ell - \frac{\eta}{\sigma_{X_1}(1 + \eta^2)} R_1,$$

■

VII.23. Proof of Lemma S.17.

PROOF. Part (i). We use the bound that (for all $t \leq -1$)

$$\Phi(t) \leq \frac{1}{|t|} \phi(t) \leq \phi(t).$$

Therefore,

$$|\Phi^{-1}(b)| \leq |\phi^{-1}(b)| = \sqrt{-2 \log(\sqrt{2\pi}b)}.$$

Part (ii). We use the bound that (for all $t \leq -\sqrt{2}$)

$$\Phi(t) \geq \frac{1}{2|t|} \phi(t).$$

Since $t^2 > \log(2|t|)$ for all $t \leq -\sqrt{2}$, we have that ($t \leq -\sqrt{2}$)

$$\Phi(t) \geq \phi(\sqrt{3}t),$$

which implies that

$$|\Phi^{-1}(b)| \geq \sqrt{-\frac{2}{3} \log(\sqrt{2\pi}b)}.$$

Part (iii). For the second inequality, we note that

$$\frac{1}{2|t|} \phi(t) \leq \Phi(t) \leq \frac{1}{|t|} \phi(t)$$

for all $t \leq -\sqrt{2}$. Therefore,

$$\begin{aligned} \Phi\left(\frac{1}{\sqrt{a}} \Phi^{-1}(b)\right) &\leq \frac{\sqrt{a}}{|\Phi^{-1}(b)|} \phi\left(\frac{1}{\sqrt{a}} \Phi^{-1}(b)\right) = \frac{\sqrt{a}}{|\Phi^{-1}(b)|} \left[\frac{1}{\sqrt{2\pi}}\right]^{\frac{a-1}{a}} \left[\phi\left(\Phi^{-1}(b)\right)\right]^{\frac{1}{a}} \\ &= 2^{\frac{1}{a}} \left[\frac{1}{\sqrt{2\pi}}\right]^{\frac{a-1}{a}} \sqrt{a} \left[\frac{1}{|\Phi^{-1}(b)|}\right]^{\frac{a-1}{a}} \left[\frac{1}{2|\Phi^{-1}(b)|} \phi\left(\Phi^{-1}(b)\right)\right]^{\frac{1}{a}} \\ &\leq 2^{\frac{1}{a}} \left[\frac{1}{\sqrt{2\pi}}\right]^{\frac{a-1}{a}} \sqrt{a} \left[\frac{1}{|\Phi^{-1}(b)|}\right]^{\frac{a-1}{a}} \left[\Phi\left(\Phi^{-1}(b)\right)\right]^{\frac{1}{a}} = 2^{\frac{1}{a}} \left[\frac{1}{\sqrt{2\pi}}\right]^{\frac{a-1}{a}} \sqrt{a} \left[\frac{1}{|\Phi^{-1}(b)|}\right]^{\frac{a-1}{a}} b^{\frac{1}{a}} \\ &\leq 2\sqrt{a} \left[\frac{1}{|\Phi^{-1}(b)|}\right]^{\frac{a-1}{a}} b^{\frac{1}{a}}, \end{aligned}$$

where the last step holds for all $a \geq 1$. Then we apply the bound in part (ii), which implies that

$$\Phi\left(\frac{1}{\sqrt{a}}\Phi^{-1}(b)\right) \leq 3\sqrt{a}\left(\frac{1}{\log\left(\frac{1}{\sqrt{2\pi}b}\right)}\right)^{\frac{a-1}{2a}} b^{\frac{1}{a}}.$$

For the first inequality, we have

$$\begin{aligned}\Phi\left(\frac{1}{\sqrt{a}}\Phi^{-1}(b)\right) &\geq \frac{\sqrt{a}}{2|\Phi^{-1}(b)|}\phi\left(\frac{1}{\sqrt{a}}\Phi^{-1}(b)\right) = \frac{\sqrt{a}}{2|\Phi^{-1}(b)|}\left[\frac{1}{\sqrt{2\pi}}\right]^{\frac{a-1}{a}}\left[\phi\left(\Phi^{-1}(b)\right)\right]^{\frac{1}{a}} \\ &= \frac{\sqrt{a}}{2}\left[\frac{1}{\sqrt{2\pi}}\right]^{\frac{a-1}{a}}\left[\frac{1}{|\Phi^{-1}(b)|}\right]^{\frac{a-1}{a}}\left[\frac{1}{|\Phi^{-1}(b)|}\phi\left(\Phi^{-1}(b)\right)\right]^{\frac{1}{a}} \\ &\geq \frac{1}{2}\left[\frac{1}{\sqrt{2\pi}}\right]^{\frac{a-1}{a}}\sqrt{a}\left[\frac{1}{|\Phi^{-1}(b)|}\right]^{\frac{a-1}{a}}b^{\frac{1}{a}} \geq \frac{1}{2}\sqrt{a}\left[\frac{1}{|\Phi^{-1}(b)|}\right]^{\frac{a-1}{a}}b^{\frac{1}{a}}.\end{aligned}$$

Next, we use the bound derived in part (i), which implies that

$$\Phi\left(\frac{1}{\sqrt{a}}\Phi^{-1}(b)\right) \geq \frac{\sqrt{a}}{4}\left(\frac{1}{\log\left(\frac{1}{\sqrt{2\pi}b}\right)}\right)^{\frac{a-1}{2a}} b^{\frac{1}{a}}.$$

■

VII.24. Proof of Lemma S.18.

PROOF. Consider the moment generating function

$$\psi_{Z|a \leq Z \leq b}(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \left[\frac{\Phi\left(\frac{b-\mu}{\sigma} - \sigma t\right) - \Phi\left(\frac{a-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \right],$$

and we first calculate the derivatives of the moment generating function:

$$\begin{aligned}\psi_{Z|a \leq Z \leq b}^{(1)}(t) &= (\mu + \sigma^2 t)\psi_{Z|a \leq Z \leq b}(t) + \sigma e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi\left(\frac{a-\mu}{\sigma} - \sigma t\right) - \phi\left(\frac{b-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \\ \psi_{Z|a \leq Z \leq b}^{(2)}(t) &= \sigma^2 \psi_{Z|a \leq Z \leq b}(t) + (\mu + \sigma^2 t)\psi_{Z|a \leq Z \leq b}^{(1)}(t) \\ &\quad + \sigma(\mu + \sigma^2 t)e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi\left(\frac{a-\mu}{\sigma} - \sigma t\right) - \phi\left(\frac{b-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \\ &\quad + \sigma^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi^{(1)}\left(\frac{b-\mu}{\sigma} - \sigma t\right) - \phi^{(1)}\left(\frac{a-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\end{aligned}$$

Therefore, the first two moments are

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

As a by-product, we can also find the variance of the truncated distribution as

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

We now continue to find the third and fourth derivatives of the moment generating function.

$$\begin{aligned} \psi_{Z|a \leq Z \leq b}^{(3)}(t) &= 2\sigma^2 \psi_{Z|a \leq Z \leq b}^{(1)}(t) + (\mu + \sigma^2 t) \psi_{Z|a \leq Z \leq b}^{(2)}(t) \\ &\quad + \sigma(\mu^2 + \sigma^2 + 2\mu\sigma^2 t + \sigma^4 t^2) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi(\frac{a-\mu}{\sigma} - \sigma t) - \phi(\frac{b-\mu}{\sigma} - \sigma t)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 2\sigma^2(\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi^{(1)}(\frac{b-\mu}{\sigma} - \sigma t) - \phi^{(1)}(\frac{a-\mu}{\sigma} - \sigma t)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + \sigma^3 e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi^{(2)}(\frac{a-\mu}{\sigma} - \sigma t) - \phi^{(2)}(\frac{b-\mu}{\sigma} - \sigma t)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \end{aligned}$$

and

$$\begin{aligned} \psi_{Z|a \leq Z \leq b}^{(4)}(t) &= 3\sigma^2 \psi_{Z|a \leq Z \leq b}^{(2)}(t) + (\mu + \sigma^2 t) \psi_{Z|a \leq Z \leq b}^{(3)}(t) \\ &\quad + \sigma(\mu^3 + 3\mu\sigma^2 + 3\mu^2\sigma^2 t + 3\sigma^4 t + 3\mu\sigma^4 t^2 + \sigma^6 t^3) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi(\frac{a-\mu}{\sigma} - \sigma t) - \phi(\frac{b-\mu}{\sigma} - \sigma t)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 3\sigma^2(\mu^2 + \sigma^2 + 2\mu\sigma^2 t + \sigma^4 t^2) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi^{(1)}(\frac{b-\mu}{\sigma} - \sigma t) - \phi^{(1)}(\frac{a-\mu}{\sigma} - \sigma t)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 3\sigma^3(\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi^{(2)}(\frac{a-\mu}{\sigma} - \sigma t) - \phi^{(2)}(\frac{b-\mu}{\sigma} - \sigma t)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + \sigma^4 e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{\phi^{(3)}(\frac{b-\mu}{\sigma} - \sigma t) - \phi^{(3)}(\frac{a-\mu}{\sigma} - \sigma t)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}[Z^3|a \leq Z \leq b] &= \mu^3 + 3\mu\sigma^2 + (2\sigma^3 + 3\mu^2\sigma) \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 3\mu\sigma^2 \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} + \sigma^3 \frac{(\frac{a-\mu}{\sigma})^2 \phi(\frac{a-\mu}{\sigma}) - (\frac{b-\mu}{\sigma})^2 \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[Z^4|a \leq Z \leq b] &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 + 4(\mu^3\sigma + 2\mu\sigma^3) \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 3(2\mu^2\sigma^2 + 3\sigma^4) \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + 4\mu\sigma^3 \frac{(\frac{a-\mu}{\sigma})^2 \phi(\frac{a-\mu}{\sigma}) - (\frac{b-\mu}{\sigma})^2 \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &\quad + \sigma^4 \frac{(\frac{a-\mu}{\sigma})^3 \phi(\frac{a-\mu}{\sigma}) - (\frac{b-\mu}{\sigma})^3 \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}. \end{aligned}$$

■

VII.25. *Proof of Lemma S.19.*

PROOF. To start, consider

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

The rest moments can be found using a similar strategy.

■

VIII. Additional Simulation and Case Study Results.

VIII.1. *Simulation Results Corresponding to Lemma S.16.* In this section, we conduct real data based simulation studies to showcase the unbiasedness of the proposed estimator $\hat{\gamma}_{j,\text{RB}}$ in Section V when the instruments are correlated. In particular, we would like to verify the results provided in Lemma S.16. We generate realistic Monte Carlo samples that mimic the correlation structure and marginal effect sizes in the BMI GWAS data from the UK Biobank [5].

To generate simulated GWAS data following Assumption S.7, we first obtain correlation (i.e., linkage disequilibrium or LD) matrices, denoted as $\tilde{\Sigma}_{X_g}$, by using the 1000 Genomes samples of European ancestry (503 samples). Here, we divide the whole genome into 1,703 approximately independent blocks according to LDetect [1] to account for the high-dimensionality nature of the genotype data. Next, we estimate the covariance matrix for each block using the shrinkage estimator of the LD matrix [7, 16, 12], which shrinks the off-diagonal entries toward zero. Specifically, we calculate the sample LD correlation matrix from the 1000 Genomes samples of European ancestry. Each entry of the LD correlation matrix is then multiplied by the factor $\exp(-\frac{2N_e c_{ij}}{m})$, where N_e is the effective population size, m is the sample size of the data for generating the genetic map, and c_{ij} is the genetic distance between sites i and j in centiMorgans scale. The entries are set to zero if the factor $\exp(-\frac{2N_e c_{ij}}{m})$ is less than a pre-specified threshold c . Following others [7, 16], we use the genetic distance generated from the 1000 Genomes OMNI arrays with $N_e = 11,400$ and $m = 183$ and the pre-specified threshold c is set to 1×10^{-3} .

Next, to simulate exposure-SNP associations, we extract the BMI GWAS results (estimated effect sizes denoted by $\hat{\gamma}_g$ and standard deviations denoted by σ_{X_g} for each block) from the UK Biobank [N= 461,460, ID: ukb-b-19953; 5]. We then construct the covariance matrix $\Sigma_{X_g} = \text{diag}(\sigma_{X_g}) \tilde{\Sigma}_{X_g} \text{diag}(\sigma_{X_g})$, and generate $\hat{\gamma}_g^* \sim \mathcal{N}(\hat{\gamma}_g, \Sigma_{X_g})$.

To verify the unbiasedness of $\hat{\gamma}_g^*$, we present simulation results based on 2,000 Monte Carlo samples in Table S.1 below. Because the magnitude of the bias is rather small, to provide informative comparison, we compare the performance of the proposed Rao-Blackwellized estimator (with different η) and the performance of the estimator without adjustment (the last column) using standardized bias, which is calculated by Monte Carlo bias divided by Monte Carlo standard deviation.

Estimator	$\hat{\gamma}_g^* (\eta = 0.8)$	$\hat{\gamma}_g^* (\eta = 1.0)$	$\hat{\gamma}_g^* (\eta = 1.2)$	$\hat{\gamma}_{\text{max}}$
Standardized bias	0.023	0.000	-0.003	-0.271

TABLE S.1

Simulation result to verify Lemma S.16. Standardized bias is computed as the Monte Carlo bias divided by the Monte Carlo standard deviation of the corresponding estimator.

VIII.2. *Simulation Under Balanced Horizontal Pleiotropy.* In this section, we examine the performance of several estimators in the presence of balanced horizontal pleiotropy. We generate Monte Carlo samples with $p = 200,000$ independent genetic variants. We set the sample size $n_X = n_Y = 100,000$, and the standard deviations $\sigma_{X_j} = 1/\sqrt{n_X}$ and $\sigma_{Y_j} = 1/\sqrt{n_Y}$. Following [9], we generate the true SNP-exposure associations γ_j and the true pleiotropic effects α_j for all SNPs through:

$$\begin{pmatrix} \gamma_j \\ \alpha_j \end{pmatrix} \sim \pi_x \rho \begin{pmatrix} \mathcal{N}(0, \varepsilon_x^2) \\ \delta_0 \end{pmatrix} + \pi_x (1 - \rho) \begin{pmatrix} \mathcal{N}(0, \varepsilon_x^2) \\ \mathcal{N}(0, \tau^2) \end{pmatrix} + \pi_y \begin{pmatrix} \delta_0 \\ \mathcal{N}(0, \tau^2) \end{pmatrix} + (1 - \pi_x - \pi_y) \begin{pmatrix} \delta_0 \\ \delta_0 \end{pmatrix}.$$

We set $\varepsilon_x^2 = \tau^2 = 5 \times 10^{-5}$, $\pi_x = 0.02$, and $\pi_y = 0.01$. The value of ρ is given in the legend of Figures S.1–S.3.

In addition to comparisons with dIVW and IVW, we further compare our RIVW estimator with eight popular and recently proposed robust MR methods, including cML-MA-BIC-DP [13], MR-Egger [2], MR-Median [3], MRmix [8], MR-Mode [6], RAPS [15], contamination mixture [ConMix; 4], and PRESSO [11]. Because some robust MR methods are computationally intensive, we simulate 500 Monte Carlo samples under $\beta \neq 0$ and 1,000 Monte Carlo samples under $\beta = 0$ (to evaluate Type-I error rates).

The results are summarized in Figures S.1–S.3. First, when the proportion of invalid IVs $1 - \rho$ is 0.3, most methods yield well-controlled Type-I error rates (Figure S.1). When $1 - \rho$ increases to 0.7 (Figure S.3), the setting becomes more challenging. Encouragingly, the RIVW estimator continues to perform well. Furthermore, RIVW achieves higher power than competing methods. We also illustrate the empirical bias for all methods (bias proportion, which is defined as the absolute of Monte Carlo bias divided by the true β). Because most robust MR methods (including cML-MA-BIC-DP, MR-Egger, MR-Median, MRmix, MR-Mode, RAPS, MR-ConMix, and PRESSO) ignore the winner’s curse, they all yield estimates with rather large bias (unless the true causal effect is exactly zero). In contrast, both dIVW (which uses all IVs and thus circumvents selection bias) and RIVW (which corrects selection bias through a novel rerandomization procedure) achieve almost unbiased results. Third, we compare all the methods regarding their coverage. Perhaps due to ignoring the winner’s curse, many robust MR methods yield incorrect coverage probabilities.

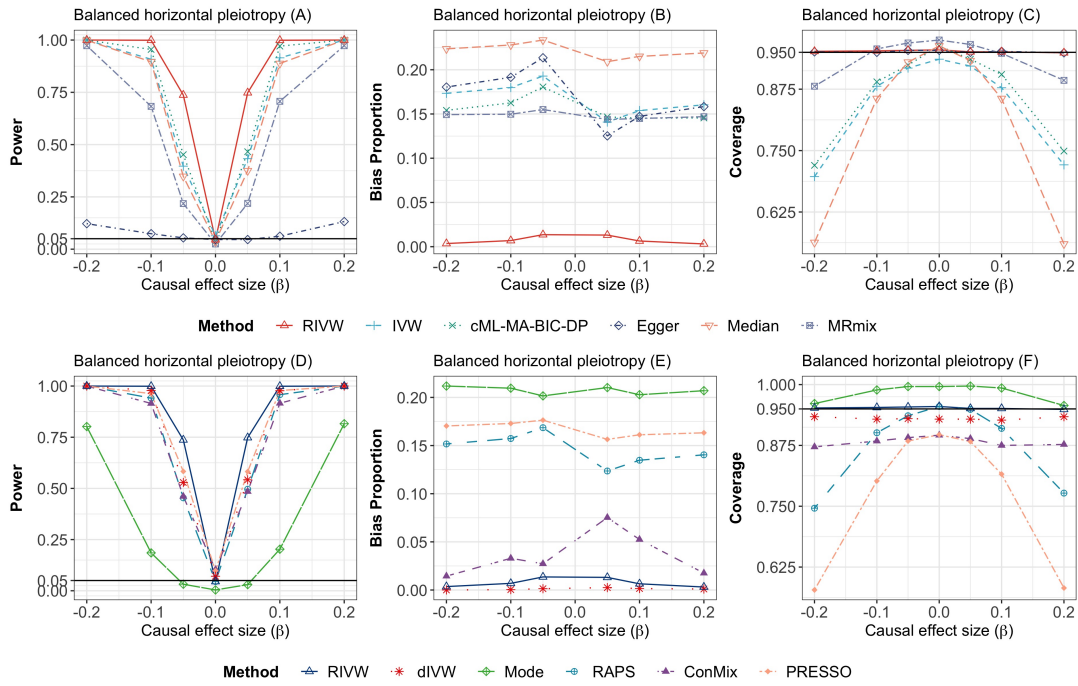


Fig S.1: Power, bias, and coverage of the RIVW estimator and several robust MR methods under the balanced horizontal pleiotropy setting with $1 - \rho = 0.3$. Power is the empirical power estimated by the proportion of p-values less than the significance threshold 0.05. Bias proportion is the absolute of Monte Carlo bias divided by true β (not computed when $\beta = 0$). Coverage is the empirical coverage probability of the 95% confidence interval.

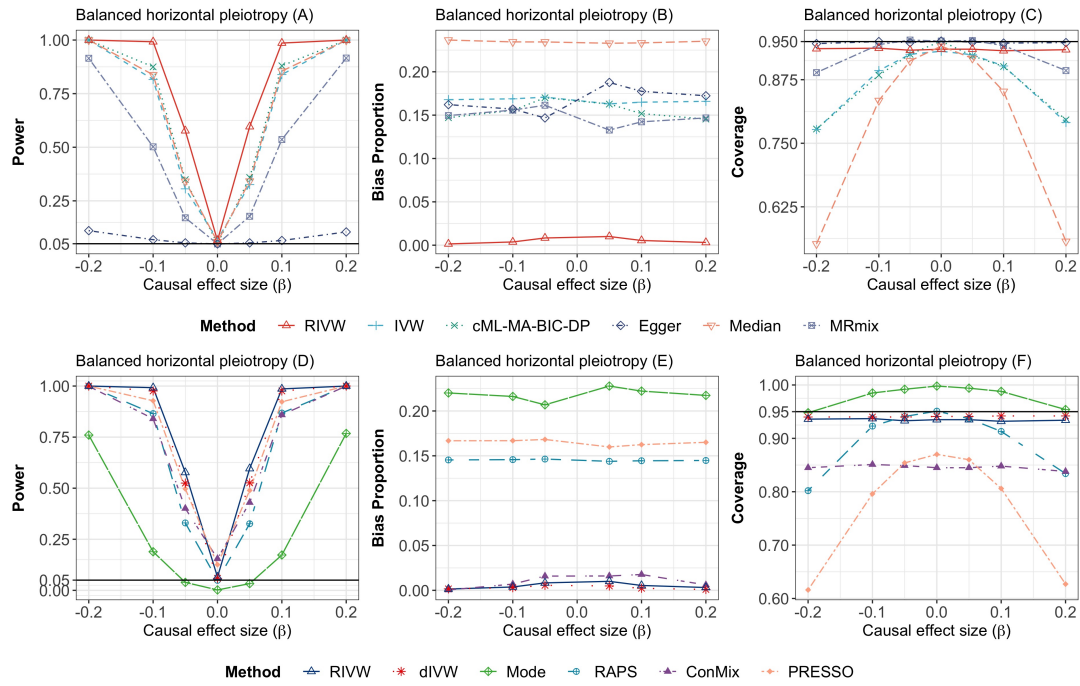


Fig S.2: Power, bias, and coverage of the RIVW estimator and several robust MR methods under the balanced horizontal pleiotropy setting with $1 - \rho = 0.5$. Power is the empirical power estimated by the proportion of p-values less than the significance threshold 0.05. Bias proportion is the absolute of Monte Carlo bias divided by true β (not computed when $\beta = 0$). Coverage is the empirical coverage probability of the 95% confidence interval.

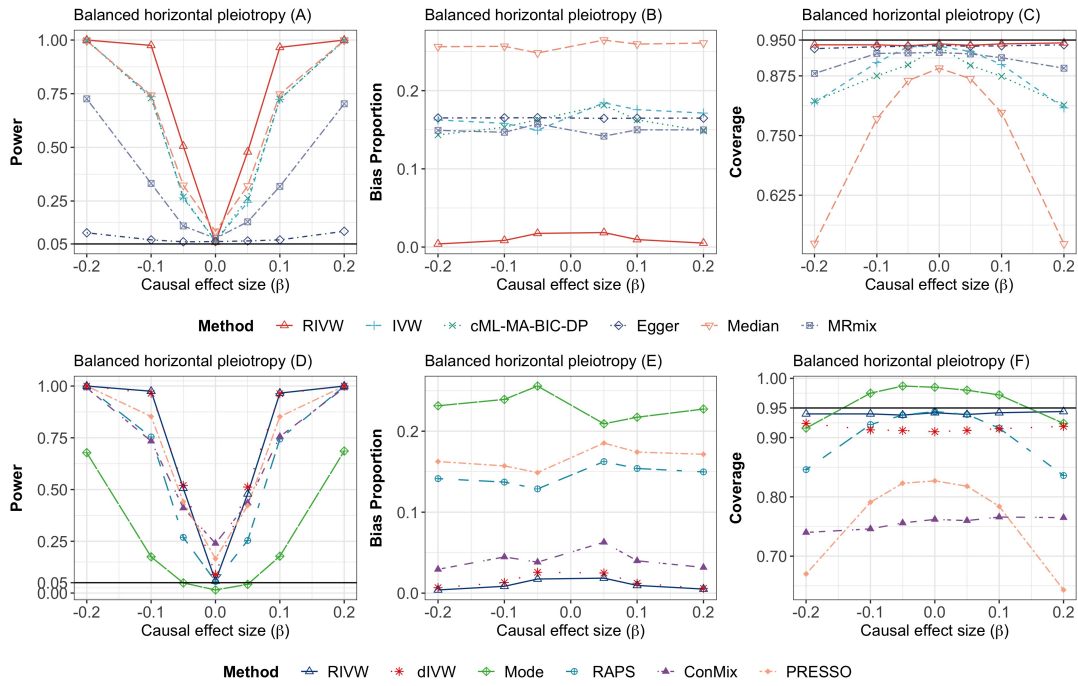


Fig S.3: Power, bias, and coverage of the RIVW estimator and several robust MR methods under the balanced horizontal pleiotropy setting with $1 - \rho = 0.7$. Power is the empirical power estimated by the proportion of p-values less than the significance threshold 0.05. Bias proportion is the absolute of Monte Carlo bias divided by true β (not computed when $\beta = 0$). Coverage is the empirical coverage probability of the 95% confidence interval.

VIII.3. *Same Trait Analyses with Standard Clumping and Pruning.* To further showcase the revised (sigma-based) pruning procedure introduced in the main paper, we repeat the same trait analyses under the same setting except that either standard clumping or pruning is used to pick independent SNPs. The results are presented in Table S.2 and Table S.3, respectively.

	$\hat{\beta}$	SE	95% CI	# IVs	F		$\hat{\beta}$	SE	95% CI	# IVs	F
	BMI-1—BMI-2 analysis						HDL-1—HDL-2 analysis				
RIVW	0.873	0.019	0.835-0.910	896	37.9		0.953	0.019	0.915-0.991	740	73.1
dIVW	0.777	0.010	0.757-0.796	1810	22.6		0.864	0.010	0.844-0.884	1816	32.1
IVW	0.834	0.014	0.806-0.861	405	42.5		0.924	0.013	0.898-0.950	311	68.1
	BMI-2—BMI-1 analysis						HDL-2—HDL-1 analysis				
RIVW	0.962	0.027	0.908-1.016	403	24.7		0.951	0.020	0.913-0.990	368	39.7
dIVW	0.642	0.008	0.626-0.658	1806	9.6		0.494	0.006	0.482-0.505	1801	12.5
IVW	0.856	0.021	0.815-0.898	78	42.0		0.934	0.018	0.899-0.969	96	58.1

TABLE S.2

Same trait analyses results using clumping to select independent IVs. The true causal effect $\beta = 1$. $\hat{\beta}$, SE, 95% CI, # IVs, and F stand for the estimated effect size, its standard error, 95% confidence interval, number of selected IVs, and mean F statistic, respectively.

	$\hat{\beta}$	SE	95% CI	# IVs	F		$\hat{\beta}$	SE	95% CI	# IVs	F
	BMI-1—BMI-2 analysis						HDL-1—HDL-2 analysis				
RIVW	0.905	0.054	0.799-1.010	63	29.4		0.984	0.079	0.830-1.139	74	34.9
dIVW	0.867	0.066	0.738-0.995	473	3.8		0.744	0.096	0.555-0.933	445	3.0
IVW	0.830	0.035	0.761-0.899	62	46.2		0.860	0.043	0.776-0.943	64	61.1
	BMI-2—BMI-1 analysis						HDL-2—HDL-1 analysis				
RIVW	1.174	0.144	0.891-1.456	59	18.8		1.125	0.136	0.859-1.390	63	19.2
dIVW	1.979	0.418	1.160-2.798	473	1.4		1.351	0.618	0.140-2.562	445	1.2
IVW	0.842	0.027	0.789-0.895	39	41.3		0.862	0.038	0.787-0.938	46	41.5

TABLE S.3

Same trait analyses results using pruning to select independent IVs. The true causal effect $\beta = 1$. $\hat{\beta}$, SE, 95% CI, # IVs, and F stand for the estimated effect size, its standard error, 95% confidence interval, number of selected IVs, and mean F statistic, respectively.

VIII.4. *Winner's Curse Bias and Measurement Error Bias Decomposition.* In this section, we consider the standard two-sample IVW estimator and decompose its bias into winner's curse bias and measurement error bias. We follow the same data generating process in Example 1 in the main paper, and corresponding model parameters are provided in Table S.4.

	$\hat{\beta}$	Winner's curse bias proportion	Measurement error bias proportion
Low heritability: $\pi_x = \pi_y = 0.002, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.040, h_y^2 = 0.042$			
IVW ($\lambda = 5.45$)	0.182	0.071	0.020
Medium heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.200, h_y^2 = 0.208$			
IVW ($\lambda = 5.45$)	0.182	0.066	0.025
High heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 3 \times 10^{-4}, h_x^2 = 0.600, h_y^2 = 0.624$			
IVW ($\lambda = 5.45$)	0.193	0.019	0.014

TABLE S.4

Bias decomposition for two-sample IVW estimators. The bias proportion is defined as the absolute Monte Carlo bias divided by true parameter $\beta = 0.2$. h_x^2 and h_y^2 are the heritability for exposure and outcome, respectively.

VIII.5. *Unequal Sample Sizes for the Exposure and Outcome GWAS.* In this section, we provide additional simulation results where (i) the sample size of the exposure GWAS is different from that of the outcome GWAS (Tables S.5 and S.6), and (ii) the sample size can be different for each SNP (Table S.7). Encouragingly, our proposed RIVW estimator continues to perform well in such settings.

	$\hat{\beta}$	monte SD	SE	CP	length	# IVs
Low heritability: $\pi_x = \pi_y = 0.002, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.040, h_y^2 = 0.042$						
IVW ($\lambda = 5.45$)	0.198	0.016	0.016	0.955	0.063	235
dIVW ($\lambda = 0$)	0.201	0.039	0.039	0.948	0.153	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.016	0.016	0.960	0.063	332
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.016	0.016	0.957	0.063	—
Three-sample IVW ($\lambda = 5.45$)	0.199	0.016	0.016	0.960	0.063	235
Three-sample IVW ($\lambda = 4.06$)	0.200	0.016	0.016	0.961	0.063	284
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.016	0.017	0.964	0.067	235
Three-sample RAPS ($\lambda = 4.06$)	0.200	0.016	0.017	0.967	0.066	284
Three-sample dIVW ($\lambda = 5.45$)	0.200	0.016	0.016	0.960	0.064	235
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.016	0.016	0.961	0.063	284
Medium heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.200, h_y^2 = 0.208$						
IVW ($\lambda = 5.45$)	0.198	0.007	0.007	0.937	0.028	1175
dIVW ($\lambda = 0$)	0.200	0.010	0.010	0.948	0.039	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.007	0.007	0.949	0.028	1430
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.007	0.007	0.949	0.028	—
Three-sample IVW ($\lambda = 5.45$)	0.199	0.007	0.007	0.942	0.028	1175
Three-sample IVW ($\lambda = 4.06$)	0.200	0.007	0.007	0.954	0.028	1383
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.007	0.008	0.947	0.029	1175
Three-sample RAPS ($\lambda = 4.06$)	0.200	0.007	0.007	0.951	0.029	1383
Three-sample dIVW ($\lambda = 5.45$)	0.200	0.007	0.007	0.949	0.028	1175
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.007	0.007	0.954	0.028	1383
High heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 3 \times 10^{-4}, h_x^2 = 0.600, h_y^2 = 0.624$						
IVW ($\lambda = 5.45$)	0.199	0.004	0.004	0.944	0.016	1507
dIVW ($\lambda = 0$)	0.200	0.005	0.005	0.944	0.019	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.004	0.004	0.951	0.016	1687
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.004	0.004	0.951	0.016	—
Three-sample IVW ($\lambda = 5.45$)	0.199	0.004	0.004	0.946	0.016	1507
Three-sample IVW ($\lambda = 4.06$)	0.200	0.004	0.004	0.953	0.016	1640
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.004	0.004	0.952	0.017	1507
Three-sample RAPS ($\lambda = 4.06$)	0.200	0.004	0.004	0.951	0.017	1640
Three-sample dIVW ($\lambda = 5.45$)	0.200	0.004	0.004	0.951	0.016	1507
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.004	0.004	0.953	0.016	1640

TABLE S.5

Simulation results under different settings with $n_X = 1,000,000$ and $n_Y = 100,000$. We set the true effect $\beta = 0.2$. The columns are: estimated effect size ($\hat{\beta}$), Monte Carlo standard deviation (monte SD), average standard error (SE), empirical coverage probability of 95% confidence intervals (CP), average confidence interval length (length), and number of selected instruments (# IVs). We do not report the number of selected instruments for the sRIVW estimator, as it employs a re-weighting strategy instead of IV selection.

	$\hat{\beta}$	monte SD	SE	CP	length	# IVs
Medium heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.200, h_y^2 = 0.208$						
IVW ($\lambda = 5.45$)	0.166	0.016	0.016	0.409	0.061	52
dIVW ($\lambda = 0$)	0.202	0.036	0.035	0.947	0.138	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.014	0.014	0.941	0.053	266
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.013	0.013	0.945	0.049	—
Three-sample IVW ($\lambda = 5.45$)	0.194	0.019	0.018	0.926	0.071	52
Three-sample IVW ($\lambda = 4.06$)	0.200	0.012	0.012	0.947	0.048	205
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.020	0.021	0.954	0.081	52
Three-sample RAPS ($\lambda = 4.06$)	0.199	0.013	0.013	0.954	0.051	205
Three-sample dIVW ($\lambda = 5.45$)	0.200	0.020	0.019	0.946	0.076	52
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.012	0.012	0.947	0.048	205
High heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 3 \times 10^{-4}, h_x^2 = 0.600, h_y^2 = 0.624$						
IVW ($\lambda = 5.45$)	0.187	0.005	0.005	0.321	0.020	346
dIVW ($\lambda = 0$)	0.200	0.013	0.012	0.928	0.048	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.005	0.005	0.952	0.020	685
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.005	0.005	0.951	0.019	—
Three-sample IVW ($\lambda = 5.45$)	0.196	0.006	0.005	0.871	0.021	345
Three-sample IVW ($\lambda = 4.06$)	0.200	0.005	0.005	0.961	0.019	631
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.006	0.006	0.949	0.023	345
Three-sample RAPS ($\lambda = 4.06$)	0.200	0.005	0.005	0.964	0.020	631
Three-sample dIVW ($\lambda = 5.45$)	0.200	0.006	0.006	0.953	0.022	346
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.005	0.005	0.961	0.019	631

TABLE S.6

Simulation results under different settings with $n_X = 50,000$ and $n_Y = 100,000$. We set the true effect $\beta = 0.2$. The columns are: estimated effect size ($\hat{\beta}$), Monte Carlo standard deviation (monte SD), average standard error (SE), empirical coverage probability of 95% confidence intervals (CP), average confidence interval length (length), and number of selected instruments (# IVs). We do not report the number of selected instruments for the sRIVW estimator, as it employs a re-weighting strategy instead of IV selection.

	$\hat{\beta}$	monte SD	SE	CP	length	# IVs
Low heritability: $\pi_x = \pi_y = 0.002, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.040, h_y^2 = 0.042$						
IVW ($\lambda = 5.45$)	0.182	0.024	0.023	0.869	0.090	40
dIVW ($\lambda = 0$)	0.210	0.128	0.127	0.963	0.497	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.023	0.022	0.952	0.088	147
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.021	0.021	0.945	0.082	—
Three-sample IVW ($\lambda = 5.45$)	0.196	0.025	0.025	0.943	0.097	40
Three-sample IVW ($\lambda = 4.06$)	0.200	0.021	0.021	0.955	0.081	98
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.027	0.028	0.963	0.109	40
Three-sample RAPS ($\lambda = 4.06$)	0.200	0.021	0.022	0.958	0.086	98
Three-sample dIVW ($\lambda = 5.45$)	0.201	0.026	0.026	0.955	0.102	40
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.021	0.021	0.955	0.081	98
Medium heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.200, h_y^2 = 0.208$						
IVW ($\lambda = 5.45$)	0.182	0.011	0.010	0.571	0.040	200
dIVW ($\lambda = 0$)	0.201	0.026	0.025	0.946	0.097	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.010	0.010	0.945	0.037	505
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.009	0.009	0.944	0.036	—
Three-sample IVW ($\lambda = 5.45$)	0.195	0.011	0.011	0.920	0.043	200
Three-sample IVW ($\lambda = 4.06$)	0.200	0.009	0.009	0.952	0.036	448
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.012	0.012	0.948	0.047	200
Three-sample RAPS ($\lambda = 4.06$)	0.200	0.009	0.010	0.954	0.038	448
Three-sample dIVW ($\lambda = 5.45$)	0.200	0.012	0.012	0.948	0.045	200
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.009	0.009	0.952	0.036	448
High heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 3 \times 10^{-4}, h_x^2 = 0.600, h_y^2 = 0.624$						
IVW ($\lambda = 5.45$)	0.193	0.005	0.004	0.675	0.018	648
dIVW ($\lambda = 0$)	0.200	0.010	0.009	0.924	0.035	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.005	0.004	0.949	0.018	985
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.200	0.005	0.004	0.951	0.017	—
Three-sample IVW ($\lambda = 5.45$)	0.197	0.005	0.005	0.900	0.018	648
Three-sample IVW ($\lambda = 4.06$)	0.200	0.004	0.004	0.956	0.017	935
Three-sample RAPS ($\lambda = 5.45$)	0.200	0.005	0.005	0.954	0.019	648
Three-sample RAPS ($\lambda = 4.06$)	0.200	0.005	0.005	0.956	0.018	935
Three-sample dIVW ($\lambda = 5.45$)	0.200	0.005	0.005	0.955	0.019	648
Three-sample dIVW ($\lambda = 4.06$)	0.200	0.004	0.004	0.956	0.017	935

TABLE S.7

Simulation results under different settings with varying sample size for each SNP. The exposure and outcome GWAS sample size n_X and n_Y are generated from a uniform distribution $n_X, n_Y \sim \text{Unif}(70000, 130000)$. We set the true effect $\beta = 0.2$. The columns are: estimated effect size ($\hat{\beta}$), Monte Carlo standard deviation (monte SD), average standard error (SE), empirical coverage probability of 95% confidence intervals (CP), average confidence interval length (length), and number of selected instruments (# IVs). We do not report the number of selected instruments for the sRIVW estimator, as it employs a re-weighting strategy instead of IV selection.

VIII.6. *Supplementary Simulation Evidence.* This section contains additional simulation results, including

- simulation results under no balanced horizontal pleiotropy (Table S.8 for the IVW estimator with different selection cutoff, Table S.9 with $\beta = 0$);
- simulation results with different choices of η (Figures S.4 and S.5)

	$\hat{\beta}$	monte SD	SE	CP	length	# IVs
Low heritability: $\pi_x = \pi_y = 0.002, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.040, h_y^2 = 0.042$						
IVW ($\lambda = 5.45$)	0.182	0.023	0.023	0.865	0.090	40
IVW ($\lambda = 4.06$)	0.172	0.018	0.018	0.635	0.070	99
Medium heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.200, h_y^2 = 0.208$						
IVW ($\lambda = 5.45$)	0.182	0.011	0.010	0.566	0.040	200
IVW ($\lambda = 4.06$)	0.179	0.008	0.008	0.287	0.032	453
High heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 3 \times 10^{-4}, h_x^2 = 0.600, h_y^2 = 0.624$						
IVW ($\lambda = 5.45$)	0.193	0.005	0.004	0.688	0.017	655
IVW ($\lambda = 4.06$)	0.193	0.004	0.004	0.601	0.016	942

TABLE S.8

Simulation results under different settings. We set the true effect $\beta = 0.2$. The columns are: estimated effect size ($\hat{\beta}$), Monte Carlo standard deviation (monte SD), average standard error (SE), empirical coverage probability of 95% confidence intervals (CP), average confidence interval length (length), and number of selected instruments (# IVs).

	$\hat{\beta}$	monte SD	SE	CP	Type 1 error	# IVs
Low heritability: $\pi_x = \pi_y = 0.002, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.040, h_y^2 = 0.042$						
IVW ($\lambda = 5.45$)	0.000	0.023	0.023	0.952	0.048	40
dIVW ($\lambda = 0$)	0.003	0.118	0.118	0.956	0.044	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	0.000	0.022	0.021	0.949	0.051	148
sRIVW ($\lambda = 4.06, \eta = 0.5$)	0.000	0.020	0.020	0.944	0.056	—
Three-sample IVW ($\lambda = 5.45$)	0.001	0.025	0.025	0.959	0.041	40
Three-sample IVW ($\lambda = 4.06$)	0.000	0.020	0.020	0.955	0.045	99
Three-sample RAPS ($\lambda = 5.45$)	0.000	0.026	0.027	0.967	0.033	40
Three-sample RAPS ($\lambda = 4.06$)	-0.000	0.020	0.021	0.959	0.041	99
Three-sample dIVW ($\lambda = 5.45$)	0.001	0.025	0.025	0.958	0.042	40
Three-sample dIVW ($\lambda = 4.06$)	0.000	0.020	0.020	0.955	0.045	99
Medium heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 1 \times 10^{-4}, h_x^2 = 0.200, h_y^2 = 0.208$						
IVW ($\lambda = 5.45$)	-0.000	0.011	0.010	0.949	0.051	200
dIVW ($\lambda = 0$)	0.000	0.024	0.024	0.941	0.059	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	-0.000	0.009	0.009	0.948	0.052	509
sRIVW ($\lambda = 4.06, \eta = 0.5$)	-0.000	0.009	0.009	0.949	0.051	—
Three-sample IVW ($\lambda = 5.45$)	-0.000	0.011	0.011	0.949	0.051	200
Three-sample IVW ($\lambda = 4.06$)	-0.000	0.009	0.009	0.952	0.048	452
Three-sample RAPS ($\lambda = 5.45$)	-0.000	0.012	0.012	0.951	0.049	200
Three-sample RAPS ($\lambda = 4.06$)	-0.000	0.009	0.009	0.956	0.044	452
Three-sample dIVW ($\lambda = 5.45$)	-0.000	0.011	0.011	0.949	0.051	200
Three-sample dIVW ($\lambda = 4.06$)	-0.000	0.009	0.009	0.952	0.048	452
High heritability: $\pi_x = \pi_y = 0.01, \varepsilon_x^2 = \tau^2 = 3 \times 10^{-4}, h_x^2 = 0.600, h_y^2 = 0.624$						
IVW ($\lambda = 5.45$)	-0.000	0.004	0.004	0.949	0.051	655
dIVW ($\lambda = 0$)	0.000	0.009	0.009	0.924	0.076	200000
RIVW ($\lambda = 4.06, \eta = 0.5$)	-0.000	0.004	0.004	0.950	0.050	993
sRIVW ($\lambda = 4.06, \eta = 0.5$)	-0.000	0.004	0.004	0.951	0.049	—
Three-sample IVW ($\lambda = 5.45$)	-0.000	0.004	0.005	0.957	0.043	655
Three-sample IVW ($\lambda = 4.06$)	-0.000	0.004	0.004	0.951	0.049	943
Three-sample RAPS ($\lambda = 5.45$)	-0.000	0.005	0.005	0.959	0.041	655
Three-sample RAPS ($\lambda = 4.06$)	-0.000	0.004	0.005	0.951	0.049	943
Three-sample dIVW ($\lambda = 5.45$)	-0.000	0.005	0.005	0.957	0.043	655
Three-sample dIVW ($\lambda = 4.06$)	-0.000	0.004	0.004	0.951	0.049	943

TABLE S.9

Simulation results under different settings. We set the true effect $\beta = 0$. The columns are: estimated effect size ($\hat{\beta}$), Monte Carlo standard deviation (monte SD), average standard error (SE), empirical coverage probability of 95% confidence intervals (CP), average confidence interval length (length), and number of selected instruments (# IVs). We do not report the number of selected instruments for the sRIVW estimator, as it employs a re-weighting strategy instead of IV selection.

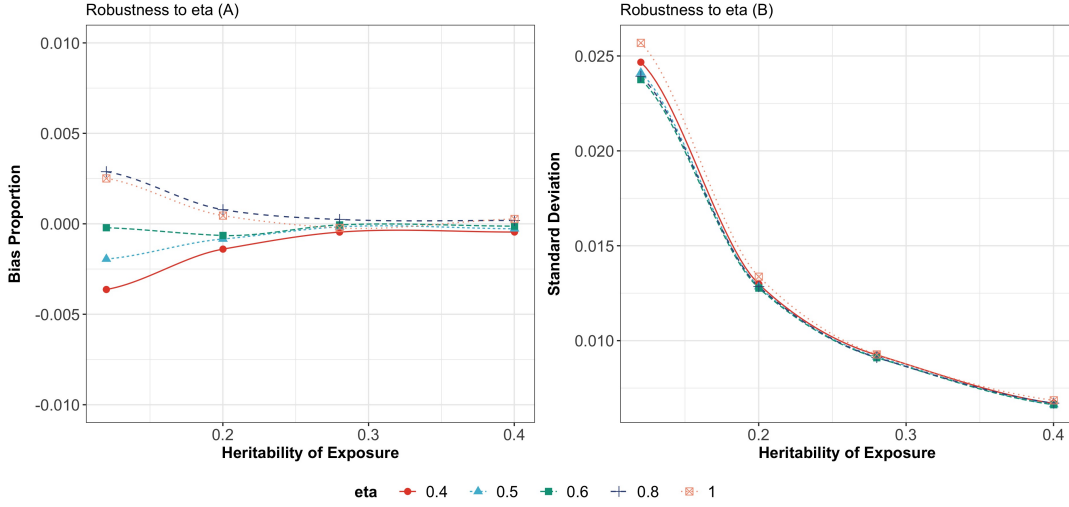


Fig S.4: Monte Carlo bias and standard deviation of the RIVW estimator for different η values. We generate Monte Carlo samples following the same simulation setup as in Example 1 except that we set the mixing probability $\pi_x = \pi_y = 0.02$ and vary $\varepsilon_x^2 = \tau^2$ in a set $\{3 \times 10^{-5}, 5 \times 10^{-5}, 7 \times 10^{-5}, 1 \times 10^{-4}\}$.

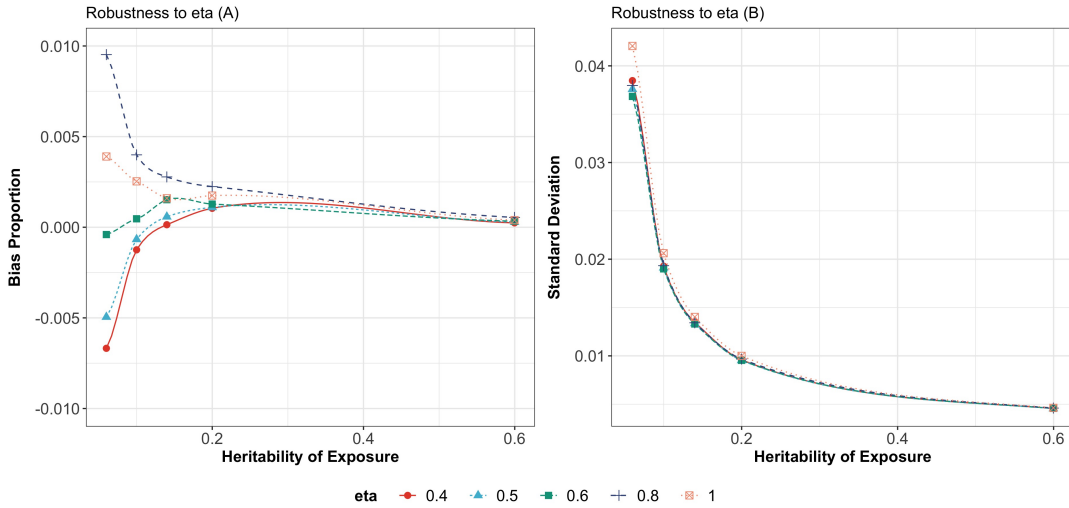


Fig S.5: Monte Carlo bias and standard deviation of the RIVW estimator for different η values. We generate Monte Carlo samples following the same simulation setup as in Example 1 except that we set the mixing probability $\pi_x = \pi_y = 0.01$ and vary $\varepsilon_x^2 = \tau^2$ in a set $\{3 \times 10^{-5}, 5 \times 10^{-5}, 7 \times 10^{-5}, 1 \times 10^{-4}, 3 \times 10^{-4}\}$.

REFERENCES

- [1] BERISA, T. and PICKRELL, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32** 283–285.
- [2] BOWDEN, J., DAVEY SMITH, G. and BURGESS, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44** 512–525.
- [3] BOWDEN, J., DAVEY SMITH, G., HAYCOCK, P. C. and BURGESS, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology* **40** 304–314.
- [4] BURGESS, S., FOLEY, C. N., ALLARA, E., STALEY, J. R. and HOWSON, J. M. (2020). A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications* **11** 1–11.
- [5] BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L. T., SHARP, K. et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562** 203–209.
- [6] HARTWIG, F. P., DAVEY SMITH, G. and BOWDEN, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology* **46** 1985–1998.
- [7] LLOYD-JONES, L. R., ZENG, J., SIDORENKO, J., YENGO, L., MOSER, G., KEMPER, K. E. et al. (2019). Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nature Communications* **10** 1–11.
- [8] QI, G. and CHATTERJEE, N. (2019). Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nature Communications* **10** 1–10.
- [9] QI, G. and CHATTERJEE, N. (2021). A comprehensive evaluation of methods for Mendelian randomization using realistic simulations and an analysis of 38 biomarkers for risk of type 2 diabetes. *International Journal of Epidemiology* **50** 1335–1349.
- [10] ROBERTSON, D. S., PREVOST, A. T. and BOWDEN, J. (2016). Accounting for selection and correlation in the analysis of two-stage genome-wide association studies. *Biostatistics* **17** 634–649.
- [11] VERBANCK, M., CHEN, C.-Y., NEALE, B. and DO, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* **50** 693–698.
- [12] WEN, X. and STEPHENS, M. (2010). Using linear predictors to impute allele frequencies from summary or pooled genotype data. *Annals of Applied Statistics* **4** 1158–1182.
- [13] XUE, H., SHEN, X. and PAN, W. (2021). Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *American Journal of Human Genetics* **108** 1251–1269.
- [14] YE, T., SHAO, J. and KANG, H. (2021). Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *Annals of Statistics* **49** 2079–2100.
- [15] ZHAO, Q., WANG, J., HEMANI, G., BOWDEN, J. and SMALL, D. S. (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics* **48** 1742–1769.
- [16] ZHU, X. and STEPHENS, M. (2017). Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *The Annals of Applied Statistics* **11** 1561–1592.