Supporting Information for “Conditional inference in cis-Mendelian randomization using weak genetic factors”

by Ashish Patel, Dipender Gill, Paul Newcombe, & Stephen Burgess

Web Appendix A  Summary data

Web Appendix A.1 Available summary data

We work with two-sample summary data on univariable genetic variant–trait associations as commonly reported from genetic association studies. For any $a$-vector $A$, and $b$-vector $B$, let $\text{var}(A)$ denote the $a \times a$ sample variance–covariance matrix of $A$, and $\text{cov}(A,B)$ denote the $a \times b$ sample covariance matrix of $A$ and $B$, that are calculated from an i.i.d. sample of size $n_Y$. Let $\text{var}(A)$ and $\text{cov}(A,B)$ denote analogous quantities calculated from an non-overlapping i.i.d. sample of size $n_X$. Let $Y$ denote the outcome, $X$ denote the exposure, and $Z = (Z_1, \ldots, Z_p)'$ denote a $p$-vector genetic variants, where all variables are mean-centered.

First, for $j \in [p]$, we have summary data from univariable $Y$ on $Z_j$ linear regressions, $\hat{\beta}_Y = \text{var}(Z_j)^{-1}\text{cov}(Z_j,Y)$ and $\hat{\sigma}_Y^2 = n_Y^{-1}\text{var}(Z_j)^{-1}(\text{var}(Y) - \text{cov}(Z_j,Y)^2\text{var}(Z_j)^{-1})$. Here, $\hat{\beta}_Y$ denotes the estimated association between $Z_j$ and $Y$, and $\hat{\sigma}_Y$ the corresponding standard error. Similarly, $\hat{\beta}_X = \text{var}(Z_j)^{-1}\text{cov}(Z_j,X)$ is the estimated association between $Z_j$ and $X$, and $\hat{\sigma}_X$ is the corresponding standard error, where $\hat{\sigma}_X^2 = n_X^{-1}\text{var}(Z_j)^{-1}(\text{var}(X) - \text{cov}(Z_j,X)^2\text{var}(Z_j)^{-1})$. We further assume knowledge of the sample variances of the traits, $\text{var}(X)$ and $\text{var}(Y)$, and knowledge of the pairwise correlation $\rho_{jk}$ between any genetic variants $Z_j$ and $Z_k$, $(j, k \in [p])$.

Let $\hat{\delta}_X = (\hat{\delta}_{X_1}, \ldots, \hat{\delta}_{X_p})'$ and $\hat{\delta}_Y = (\hat{\delta}_{Y_1}, \ldots, \hat{\delta}_{Y_p})'$ denote the vector of sample variant–trait covariances, and let $\delta_X = (\delta_{X_1}, \ldots, \delta_{X_p})'$ and $\delta_Y = (\delta_{Y_1}, \ldots, \delta_{Y_p})'$ denote their population counterparts. By similar arguments used in Wang and Kang (2021), we can recover the following relevant quantities from summary data:

1. Variance-covariance matrix of genetic variants: $\text{var}(Z)$ and $\text{var}(Z)$

Note that $\hat{v}_j \equiv (n_Y \hat{\sigma}_Y^2 + \hat{\beta}_Y^2)^{-1} = \text{var}(Y)^{-1}\text{var}(Z_j)$. Thus, $\text{var}(Y)\rho_{jk} \sqrt{\hat{v}_j \hat{v}_k} = \text{cov}(Z_j, Z_k)$. Therefore, we can construct the sample variance-covariance matrix of $Z$, $\text{var}(Z)$. We can also compute $\text{var}(Z)$ by identical arguments.
2. genetic variant–trait covariances: $\hat{\delta}_X$ and $\hat{\delta}_Y$

Let $\hat{\delta}_Y = \hat{\delta}_Y = \hat{\delta}_Y$. We can also compute $\hat{\delta}_X$ by identical arguments.

Web Appendix A.2 Approximation of log odds ratio estimates to linear regression coefficients

Here we describe the procedure for obtaining estimates of $\hat{\delta}_Y = \hat{\delta}_Y = \hat{\delta}_Y$ from available summary data on log odds ratio estimates from univariable $Y$ on $Z_j$ logit regression coefficients $\hat{\beta}_Y$, and corresponding standard errors $\hat{\sigma}_Y$. The procedure relies on individual variants $Z_j$ explaining a low proportion of variation in the outcome $Y$, which is plausible in typical MR applications. The standardized least squares estimate of a linear regression of $Y$ on $Z_j$ is approximately $\hat{\beta}_Y = \hat{\beta}_Y$. Thus, $\hat{\delta}_Y = \hat{\beta}_Y \hat{\beta}_Y$. For other summary data methods which use estimated coefficients of univariable linear regressions as inputs, we can compute $\bar{\beta}_Y = \bar{\beta}_Y \hat{\beta}_Y$ for approximated effect size estimates, and the approximated standard errors based on equating z-scores are $\bar{\sigma}_Y = \bar{\beta}_Y \hat{\sigma}_Y \hat{\beta}_Y^{-1}$.

Web Appendix A.3 Individual-level data model

Let $Y$ denote a continuous outcome, $X$ a continuous exposure, and $Z = (Z_1, \ldots, Z_p)'$ a $p$-vector of mean-centered genetic variants. Our summary-data model can be motivated from the following individual-level data model

$$
Y = \alpha_0 + X\theta_0 + U
$$
$$
X = \gamma_0 + (\Lambda'Z)'\gamma_1 + V,
$$

where $E[U|Z] = 0$, $E[U^2|\Lambda'Z] = \sigma_U^2$, $E[V|Z] = 0$, and $E[V|\Lambda'Z] = \sigma_V^2$. This means that the genetic variants $Z$ are valid instruments, and that when using the linear combination of variants $\Lambda'Z$ as instruments, the errors are homoscedastic. Note that the reduced form model for $Y$ can be written

$$
Y = \phi_0 + (\Lambda'Z)'\phi_1 + \epsilon,
$$

where $E[\epsilon|Z] = 0$, $E[\epsilon^2|\Lambda'Z] = \sigma_\epsilon^2$, $\phi_0 = \alpha_0 + \theta_0\gamma_0$, $\phi_1 = \gamma_1\theta_0$, and $\epsilon = U + \theta_0V$. Therefore, given an i.i.d. $n_X$-sample on $(X, Z)$, we have $E[\Lambda'\hat{\delta}_X] = \Lambda'E[ZX] = \Lambda'\delta_X$, and $var(\Lambda'\hat{\delta}_X) = E[(\Lambda'Z)(\Lambda'Z)\cdot var(X|\Lambda'Z)]n_X^{-1} = \Omega_0\sigma_X^2n_X^{-1}$. The mean and variance of $\Lambda'\hat{\delta}_Y$ from an non-overlapping i.i.d. $n_Y$-sample on $(Y, Z)$ can be calculated by identical arguments. Hence, a central limit theorem motivates Assumption 2.2.
Web Appendix B  Preparatory lemmata

We use the following abbreviations: ‘T’ denotes the triangle inequality; ‘CH’ denotes Chebyshev’s inequality; ‘CS’ denotes the Cauchy-Schwarz inequality; ‘M’ denotes the Markov inequality; ‘w.p.a.1’ denotes ‘with probability approaching 1’; ‘RHS’ denotes ‘right-hand side’; ‘LHS denotes ‘left-hand side’. For any vector or matrix $A$, $\|A\|$ denotes the Euclidean norm.

Web Appendix B.1 Bai (2003)’s results on factor loadings

Let $n_Z$ denote the sample size of the reference sample used to compute the genetic variant correlation matrix $\rho$. The reference sample is permitted to overlap with one of the genetic association studies. We further assume that $n_Z$ is $\Theta(n)$ so that, in our asymptotic analysis, we consider the sample sizes of the genetic association studies ($n_X$ and $n_Y$) and any reference sample ($n_Z$) to all be increasing at the same rate with the number of genetic variants $p$.

Let $z_{ki}$ denote the mean-centered $k$–th genetic variant for the $i$-th individual, $i \in [n_Z]$. Let $z_i = (z_{i1}, \ldots, z_{ip})'$, and let $z = (z_1, \ldots, z_{n_Z})'$ denote the $n_Z \times p$ matrix of mean-centered genotypes. Under Bai and Ng (2002)’s approximate factor model, we have

$$z = FA' + e,$$

where $\Lambda = (\lambda_1, \ldots, \lambda_p)'$ denotes a $p \times r$ matrix of factor loadings, $F = (F_1, \ldots, F_{n_Z})'$ denotes an $n_Z \times r$ matrix of factors, and $e = (e_1, \ldots, e_{n_Z})'$ denotes a $n_Z \times p$ matrix of idiosyncratic errors.

Let the columns of $\bar{\Lambda}$ denote the first $r$ eigenvectors of the $p \times p$ sample variance-covariance matrix $\bar{\sigma}(Z) = z'z/n_Z$ multiplied by $\sqrt{p}$. Let $\bar{F} = (z\Lambda)(\Lambda^\prime \Lambda)^{-1}$, and $\hat{F} = \bar{F}(\hat{F}'\hat{F}/n_Z)^{-\frac{1}{2}}$. Then, $\hat{F}'\hat{F}/n_Z = I_r$, and $\hat{\Lambda} = \hat{F}'z/n_Z = (\hat{F}'\hat{F}/n_Z)^{-\frac{1}{2}}(\hat{\Lambda}'\hat{\Lambda})^{-1}(\hat{\Lambda}'z/n_Z)$.

Let $H = (\Lambda'\Lambda/p)(\hat{F}'\hat{F}/n_Z)D_{np}^{-1}$ where $D_{np}$ is an $r \times r$ diagonal matrix with its diagonal entries equal to the first $r$ largest eigenvalues of $(zz'/(n_Zp))$ in decreasing order.

**Lemma 1.** As $n, p \to \infty$, (i) $\|H\| = O_P(1)$; (ii) $\|H^{-1}\| = O_P(1)$.

**Proof.** By Lemma A.3(i) of Bai (2003, p.161), $D_{np} \xrightarrow{P} D_0$, a diagonal matrix consisting of the eigenvalues of $\Sigma_{\Lambda}\Sigma_{F}$. Hence, w.p.a.1, $D_{np}^{-1}$ exists since $\Sigma_{\Lambda}$ and $\Sigma_{F}$ are positive definite, and $\|D_{np}^{-1}\| = O_P(1)$. By construction, $\hat{F}'\hat{F}/n_Z = I_r$, so that $\|\hat{F}'\hat{F}/n_Z\| = O_P(1)$. By CS, Assumption 1(iv), and $M$, $\|\hat{F}'F/n_Z\| \leq \sum_{i=1}^{n_Z} \|F_i\|^2/n_Z = O_P(1)$. Similarly, for all $k \in [p]$, by Assumption 1(v), $\|\lambda_k\| \leq C_\Lambda < \infty$. Therefore, $\|\Lambda'\Lambda/p\| \leq \sum_{k=1}^{p} \|\lambda_k\|^2/p = O(1)$. Also, by CS, $\|H\| \leq \|\Lambda'\Lambda/p\| \cdot \|F'F/n_Z\|^\frac{1}{2} \cdot \|\hat{F}'\hat{F}/n_Z\|^\frac{1}{2} \cdot \|D_{np}^{-1}\|$, so that $\|H\| = O_P(1)$. $H$ is invertible; see Bai (2003, p.145). Therefore, by Part (i), $\|H^{-1}\| = O_P(1)$.

**Lemma 2.** As $n, p \to \infty$, $\|\hat{\lambda}_k - H^{-1}\lambda_k\| = O_P(n^{-\frac{1}{2}}) + O_P(\min(n, p)^{-1})$, $k \in [p]$.
Proof. By Bai (2003, p.165), for any $k \in [p]$, we have

$$
\hat{\lambda}_k - H^{-1}\lambda_k = H' \frac{1}{nZ} \sum_{i=1}^{nZ} F_i e_{ki} + \frac{1}{nZ} \sum_{i=1}^{nZ} \hat{F}_i (F_i - H^{-1}\hat{F}_i)' \lambda_k + \frac{1}{nZ} \sum_{i=1}^{nZ} (\hat{F}_i - H'F_i)e_{ki},
$$

(S.1)

where $e_{ki}$ is the $(i, k)$-th element of the matrix of idiosyncratic errors $e$. Hence, by CS,

$$
||\hat{\lambda}_k - H^{-1}\lambda_k|| = \frac{1}{\sqrt{nZ}} ||H|| \left( \left| \frac{1}{\sqrt{nZ}} \sum_{i=1}^{nZ} F_i e_{ki} \right| + \left| \frac{1}{nZ} \sum_{i=1}^{nZ} \hat{F}_i (F_i - H^{-1}\hat{F}_i)' \right| \left| \lambda_k \right| + \left| \frac{1}{nZ} \sum_{i=1}^{nZ} (\hat{F}_i - H'F_i)e_{ki} \right| \right),
$$

where the first term on the RHS is $O_P(n^{-\frac{1}{2}})$ by Lemma 1(i), Assumption D of Bai (2003, p.141), and M. The second term on the RHS is $O_P(\min(n,p)^{-1})$ by Lemma B.3 of Bai (2003, p.165), Lemma 1(ii), and Assumption 1(v). The third term on the RHS is $O_P(\min(n,p)^{-1})$ by Lemma B.1 of Bai (2003, p.163). □

**Web Appendix B.2 Weak dependence of model errors**

Under our model assumptions, $\delta_Y = \delta_X \theta_0$. We can write $\delta_Y - \delta_X \theta_0 = \varepsilon_Y - \varepsilon_X \theta_0$, where $\varepsilon_X = \delta_X - \delta_X$ and $\varepsilon_Y = \delta_Y - \delta_Y$. Let $\varepsilon = \varepsilon_Y - \theta_0 \varepsilon_X$, and define $\varepsilon_{Xk} = \sqrt{n} \varepsilon_{Xk}$ and $\varepsilon_k = \sqrt{n} \varepsilon_k$, $k \in [p]$. Under an assumption of random sampling, and $E[\varepsilon_{Xk}^2] < \infty$, $E[\varepsilon_k^2] < \infty$, the re-scaled errors are bounded in probability. We maintain the following assumptions which are analogous to Assumptions A and B in Bai and Ng (2010, pp.1581-2), and they ensure the sampling errors from estimating the space spanned by the factor loadings are negligible as $n, p \to \infty$.

**Assumption S0.** There exists a universal positive constant $C$, such that, as $n_Z, n, p \to \infty$, (i) for each $i \in [n_Z]$, $\sum_{k=1}^{p} [E[\varepsilon_{ki}\varepsilon_{ki}^*]] \leq C$ and $\sum_{k=1}^{p} [E[\varepsilon_{ki}\varepsilon_{ki}^*]] \leq C$; (ii) $E\left[ \left| \frac{1}{\sqrt{nZ}} \sum_{i=1}^{nZ} \sum_{k=1}^{p} F_i (\varepsilon_{ki} - E[\varepsilon_{ki}\varepsilon_{ki}^*]) \right| \right] \leq C$ and $E\left[ \left| \frac{1}{\sqrt{nZ}} \sum_{i=1}^{nZ} \sum_{k=1}^{p} F_i (\varepsilon_{ki} - E[\varepsilon_{ki}\varepsilon_{ki}^*]) \right| \right] \leq C$; (iii) for each $i \in [n_Z]$, $E\left[ \left| \frac{1}{\sqrt{p}} \sum_{k=1}^{p} (\varepsilon_{ki} - E[\varepsilon_{ki}\varepsilon_{ki}^*]) \right| \right] \leq C$ and $E\left[ \left| \frac{1}{\sqrt{p}} \sum_{k=1}^{p} (\varepsilon_{ki} - E[\varepsilon_{ki}\varepsilon_{ki}^*]) \right| \right] \leq C$; (iv) for each $i \in [n_Z]$, $E\left[ \left| \frac{1}{\sqrt{p}} \sum_{k=1}^{p} (\varepsilon_{ki} - E[\varepsilon_{ki}\varepsilon_{ki}^*]) \right|^2 \right] \leq C$ and $E\left[ \left| \frac{1}{\sqrt{p}} \sum_{k=1}^{p} (\varepsilon_{ki} - E[\varepsilon_{ki}\varepsilon_{ki}^*]) \right|^2 \right] \leq C$.

Assumptions S0(i)–(iii) restrict the extent to which idiosyncratic errors of the approximate factor model can be correlated with sampling errors of genetic associations in the linear model. These assumptions are satisfied when a separate reference sample is used only to estimate the variance-covariance matrix of genetic variants, $var(Z)$. In general, these assumptions are also likely to hold when the latent factors explain a large proportion of genetic variation. Conversely, Assumptions S0(i)–(iii) may be violated when variants are in less structured linkage disequilibrium, and when the structural linear model used to describe genetic variant–trait associations is misspecified. For example, this could occur if there is selection bias from the recruitment of study participants having survived competing risks to exposure and outcome (Schooling et al., 2021). For Parts (ii)–(iv) of Assumption S0, we note that these are conditions on zero-mean sums, and are analogous to Assumptions C5, D, F1, and F2 of Bai (2003, p.141 and p.144).
Web Appendix B.3 Covariance terms

Lemma 3. As \( n, p \to \infty \), (i) \( \Omega_0 = O_P(p) \); (ii) \( \Omega_0^{-1} = O_P(p^{-1}) \); (iii) \( \hat{\Omega}_0 - H^{-1}\Omega_0H^{-1'} = o_P(p) \); (iv) \( \hat{\Omega}_0 = O_P(p) \); (v) \( \hat{\Omega}_0^{-1} = O_P(p^{-1}) \); (vi) \( \hat{\Omega}_0^{-1} - H'\Omega_0^{-1}H = o_P(p^{-1}) \).

Proof. First, for Parts (i) and (ii), \( \Omega_0 = O(\Lambda'\operatorname{var}(Z)\Lambda) = O(p) \) since \( \Lambda'\Lambda = \Theta(p) \), \( \operatorname{cov}(Z_j, Z_k) = \Theta(1) \) for all \( j, k \in [p] \), and \( \Omega_0 \) is invertible. For Part (iii), note that we can write \( \Lambda'\operatorname{var}(Z)\hat{\Lambda} - H^{-1}\Lambda'\operatorname{var}(Z)H^{-1'} = R_1 + 2R_2 + R_3 \), where \( R_1 = (\hat{\Lambda} - \Lambda H^{-1'})(\Lambda - \Lambda H^{-1}) \), \( R_2 = H^{-1}\Lambda'\operatorname{var}(Z)(\hat{\Lambda} - \Lambda H^{-1}) \), and \( R_3 = H^{-1}\Lambda'\operatorname{var}(Z) - \operatorname{var}(Z)\Lambda H^{-1'} \). Note that \( R_1 = O((\hat{\Lambda} - \Lambda H^{-1'})'(\hat{\Lambda} - \Lambda H^{-1})) \), since \( \operatorname{cov}(Z_j, Z_k) - \operatorname{cov}(\hat{Z}_j, \hat{Z}_k) = o_P(1) \) by the weak law of large numbers. Hence, \( \|R_1\| \leq \sum_{k=1}^p \|\hat{\lambda}_k - H^{-1}\lambda_k\|^2 = O_P(n^{-1}p) + O_P(\min(n, p)^{-2})p \) by Lemma 2. Similarly, we have \( \|R_2\| \leq \|\hat{\Lambda} - \Lambda H^{-1'}\| \cdot \|\Lambda H^{-1}\| \) by Assumption 1(v) and Lemma 2, \( \|\hat{\Lambda} - \Lambda H^{-1'}\|^2 \leq (\hat{\Lambda} - \Lambda H^{-1'})'(\hat{\Lambda} - \Lambda H^{-1}) \leq \sum_{k=1}^p \|\hat{\lambda}_k - H^{-1}\lambda_k\|^2 = O_P(n^{-1}p) + O_P(\min(n, p)^{-2})p \), and \( \|\Lambda H^{-1}\|^2 \leq \|H^{-1}\|^2 \cdot \|\Lambda'\Lambda\| = O_P(p) \) by Lemma 1(ii) and Assumption 1(v). Hence, by CS, \( \|R_2\| = O_P(n^{-\frac{3}{2}}p) + O_P(\min(n, p)^{-1}p) \). For \( R_3 \), we have that \( H^{-1}p^{-1}\Lambda'\operatorname{var}(Z) - \operatorname{var}(Z)\Lambda H^{-1} = o_P(1) \) since \( \hat{\Lambda}^{-1}p^{-1}\Lambda'\operatorname{var}(Z) - \operatorname{var}(Z)\Lambda H^{-1} \) is invertible by large numbers. Hence, \( \|R_3\| = o_P(1) \). For Part (iv), note that by T and CS, \( \|\hat{\Omega}_0\| \leq \|\hat{\Omega}_0 - H^{-1}\Omega_0H^{-1'}\| + \|H^{-1}\|\|\Omega_0\| = O_P(p) \) by Lemma 1(ii) and Parts (i) and (iii). For Part (v), by Part (iii), \( p^{-1}(\hat{\Omega}_0 - H^{-1}\Omega_0H^{-1'}) = o_P(1) \) and \( p^{-1}H^{-1}\Omega_0H^{-1'} \) is invertible. Hence, \( p^{-1}\hat{\Omega}^{-1}_0 = \hat{\Omega}_0^{-1} \) is invertible by w.p.a.1, and \( \hat{\Omega}_0^{-1} = O_P(p) \) by Parts (ii) and (iv). For (vi), we can write \( \hat{\Omega}_0^{-1} - H'\Omega_0^{-1}H = \hat{\Omega}_0^{-1}(H^{-1}\Omega_0H^{-1'} - \hat{\Omega}_0)H'\Omega_0^{-1}H \), so that by CS, \( \|\hat{\Omega}_0^{-1} - H'\Omega_0^{-1}H\| \leq \|\hat{\Omega}_0^{-1}\| \cdot \|\hat{\Omega}_0 - H^{-1}\Omega_0H^{-1'}\| \cdot \|H\|^2\|\Omega_0^{-1}\| = O_P(p^{-1}) \) by Lemma 1(i), and Parts (ii), (iii), and (v).

\[ \square \]

Web Appendix B.4 Derivative terms

Let \( \hat{G} = -\hat{\Lambda}'\hat{\delta}_X, G = -H^{-1}\Lambda'\delta_X \), and \( G = -H^{-1}\Lambda'\delta_X \). Similarly, let \( \hat{G}_Y = -\hat{\Lambda}'\hat{\delta}_Y, G_Y = -H^{-1}\Lambda'\delta_Y \), and \( G_Y = -H^{-1}\Lambda'\delta_Y \).

Lemma 4. As \( n, p \to \infty \), (i) \( \hat{G} - G = O_P(1) + o_P(n^{-\frac{3}{2}}p^\frac{1}{2}) \); (ii) \( \hat{G} - G = O_P(n^{-\frac{3}{2}}p^\frac{1}{2}) \); (iii) \( G = O_P(p^\frac{1}{2}) \) and \( \hat{G} = O_P(p^\frac{1}{2}) \).

Proof. For Part (i), note that \( \hat{G} - G = -\sum_{k=1}^p (\hat{\lambda}_k - H^{-1}\lambda_k)\delta_X - \sum_{k=1}^p (\hat{\lambda}_k - H^{-1}\lambda_k)\varepsilon_{X_k} \). The first term on the RHS is bounded by \( \sum_{k=1}^p \|\hat{\lambda}_k - H^{-1}\lambda_k\|^2 \). The second term on the RHS is \( O_P(n^{-\frac{3}{2}}p^{-\frac{3}{2}}) \) by identical arguments used in Proof of Lemma 8(ii) below. For Part (ii), by Assumption 2 and CH, \( \Lambda'\hat{\delta}_X - \delta_X = O_P(n^{-\frac{3}{2}}p^\frac{1}{2}) \), so that by Lemma 1(ii), \( \hat{G} - G = O_P(n^{-\frac{3}{2}}p^\frac{1}{2}) \). For Part (iii), by CS, \( \|G\| \leq \|H^{-1}\| \cdot \|\Lambda'\Lambda\| \cdot \|\delta_X\| = O_P(p^\frac{1}{2}) \) by Lemma 1(ii) and Assumptions 1(v) and 3. The final result of Part (iii) then follows by Parts (i), (ii), and T.

\[ \square \]

Lemma 5. As \( n, p \to \infty \), (i) \( \hat{G}_Y - G_Y = O_P(1) + o_P(n^{-\frac{3}{2}}p^\frac{1}{2}) \); (ii) \( \hat{G}_Y - G_Y = O_P(n^{-\frac{3}{2}}p^\frac{1}{2}) \); (iii) \( G_Y = O_P(p^\frac{1}{2}) \) and \( \hat{G}_Y = O_P(p^\frac{1}{2}) \).

Proof. The results follow by analogous arguments used in the Proof of Lemma 4, where here we also note that \( \delta_Y = \delta_X\theta_0 \), so that \( \|\delta_X\| = O(1) \) implies that \( \|\delta_Y\| = O(1) \).

\[ \square \]
Covariance terms II

Let $\Omega(\theta) = H^{-1}\Omega_0 H^{-1} \cdot (n^{-1}_Y \sigma^2 + \theta^2 n^{-1}_X \sigma^2_Y)$, $\Omega = \Omega(\theta_0)$, and for any consistent estimator $\hat{\theta}$ for $\theta_0$, let $\hat{\Omega}(\hat{\theta}) = \hat{\Omega}_0 \cdot (n^{-1}_Y \hat{\sigma}^2 + \hat{\theta}^2 n^{-1}_X \hat{\sigma}^2_Y)$. 

**Lemma 6.** For $n, p \to \infty$, (i) $\hat{\sigma}^2 - \sigma^2 = o_P(1)$; (ii) $\hat{\sigma}^2_Y - \sigma^2_Y = o_P(1)$.

Proof. For Part (i), note that

$$\hat{\sigma}^2 = \text{var}(Y) - \hat{G}'_Y \hat{\Omega}^{-1}_0 \hat{G}_Y$$

by consistency of $\hat{\Omega}$ by identical arguments.

Let $\hat{\sigma}^2_Y = \text{var}(Y) - \hat{G}'_Y \hat{\Omega}^{-1}_0 \hat{G}_Y - \hat{G}'_Y H^{-1}_0 \hat{H}^{-1}_Y \hat{G}_Y = \text{var}(Y) + o_P(1)$ by Lemmas 1(ii), 3(i), (iii), 6, and M. Part (iv) then follows by Parts (ii), (iii), and T. For Part (v), note that Part (iii) means that $\hat{\theta} - \theta_0 = o_P(n^{-1})$, and $\hat{\theta} - \theta_0$ is invertible, so that $p^{-1}n\hat{\Omega}(\hat{\theta})$ is invertible w.p.a.1, and $\hat{\Omega}(\hat{\theta})^{-1} = o_P(np^{-1})$. Finally, for Part (vi), note that we can write $\hat{\Omega}(\hat{\theta})^{-1} - \Omega^{-1} = \hat{\Omega}(\hat{\theta})^{-1}(\Omega - \hat{\Omega}(\hat{\theta}))\Omega^{-1} = O_P(n^2 p^{-2} o_P(n^{-1}) = o_P(np^{-1})$ by Parts (ii), (iii), and T.

**Web Appendix B.5 Residual term**

Let $\hat{g}(\theta) = \hat{\Lambda}'(\hat{\delta}_Y - \hat{\delta}_X \theta)$, and $g(\theta) = H^{-1}\Lambda'(\delta_Y - \delta_X \theta)$. 

[S-6]
Lemma 8. As \( n, p \to \infty \), (i) \( \hat{g}(\theta_0) - g(\theta_0) = o_p(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \); (ii) \( \hat{g}(\theta_0) = O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \); (iii) \( \hat{g}(\theta) = o_P(p^{\frac{3}{2}}) \) for any \( \theta \xrightarrow{p} \theta_0 \).

Proof. For Part (i), we can write

\[
\hat{g}(\theta_0) - g(\theta_0) = \sum_{k=1}^{p} (\hat{\lambda}_k - H^{-1}\lambda_k) \varepsilon_k
\]

\[
= H' \frac{1}{nZ} \sum_{i=1}^{nZ} \sum_{k=1}^{p} F_i e_{ki} \varepsilon_k + \frac{1}{nZ} \sum_{i=1}^{nZ} \sum_{k=1}^{p} \hat{F}_i (F_i - H^{-1} \hat{F}_i)' \lambda_k \varepsilon_k + \frac{1}{nZ} \sum_{i=1}^{nZ} \sum_{k=1}^{p} (\hat{F}_i - H' F_i) e_{ki} \varepsilon_k
\]

\[
:= H'R_1 + R_2 + R_3,
\]

where the second equality follows by Equation (S.1). First, note that

\[
\sqrt{n}R_1 = \sqrt{\frac{p}{nZ}} \sqrt{\frac{1}{nZ} \sum_{i=1}^{nZ} \sum_{k=1}^{p} F_i (e_{ki} \varepsilon_k^* - \mathbb{E}[e_{ki} \varepsilon_k^*])} + \frac{1}{nZ} \sum_{i=1}^{nZ} \sum_{k=1}^{p} \mathbb{E}[e_{ki} \varepsilon_k^*].
\]

By CS,

\[
\sqrt{n}\|R_1\| \leq \sqrt{\frac{p}{nZ}} \left\| \frac{1}{nZ} \sum_{i=1}^{nZ} \sum_{k=1}^{p} F_i (e_{ki} \varepsilon_k^* - \mathbb{E}[e_{ki} \varepsilon_k^*]) \right\| + \left( \frac{1}{nZ} \sum_{i=1}^{nZ} \|F_i\|^2 \right)^{\frac{1}{2}} \left( \frac{1}{nZ} \sum_{i=1}^{nZ} \left( \sum_{k=1}^{p} |\mathbb{E}[e_{ki} \varepsilon_k^*]| \right)^2 \right)^{\frac{1}{2}}
\]

\[
= O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) + O_P(1),
\]

by Assumption 1(iv), Assumptions S0(i) and (ii), CS, T, and M. Thus, by Lemma 1(i) and CS, \( H'R_1 = O_P(n^{-1}p^{\frac{1}{2}}) + O_P(n^{-\frac{1}{2}}) = o_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \).

For \( R_2 \), by CS, Lemma 1(ii), and Lemma B.3 of Bai (2003, p.165),

\[
\left\| \frac{1}{nZ} \hat{F}'(F - \hat{F}H^{-1}) \right\| \leq \left\| \frac{1}{nZ} \hat{F}'(FH - \hat{F}) \right\| \cdot \|H^{-1}\|
\]

\[
= O_P(\min(n,p)^{-1}).
\]

Also, \( \|A' \varepsilon\| = O_P(p^{\frac{1}{2}}n^{-\frac{1}{2}}) \) by Assumption 2. Hence, by CS, \( \|R_2\| = O_P(p^{\frac{3}{2}}n^{-\frac{1}{2}}) \times O_P(\min(n,p)^{-1}) = o_P(n^{-\frac{1}{2}}p^{\frac{3}{2}}) \).

Finally, for \( R_3 \), note that

\[
\sqrt{n}R_3 = \frac{1}{nZ} \sum_{i=1}^{nZ} (\hat{F}_i - H' F_i) \sum_{k=1}^{p} (e_{ki} \varepsilon_k^* - \mathbb{E}[e_{ki} \varepsilon_k^*]) + \frac{1}{nZ} \sum_{i=1}^{nZ} (\hat{F}_i - H' F_i) \sum_{k=1}^{p} \mathbb{E}[e_{ki} \varepsilon_k^*].
\]
Then, by CS,
\[
\|\mathcal{R}_3\| \leq \sqrt{\frac{p}{n} \left( \frac{1}{n} \sum_{i=1}^{n_Z} \|\hat{F}_i - H'F_i\|^2 \right)^\frac{1}{2}} \left( \frac{1}{n} \sum_{i=1}^{n_Z} \left| \frac{1}{\sqrt{p}} \sum_{k=1}^{p} (e_{ki} \epsilon_{ik}^* - \mathbb{E}[e_{ki} \epsilon_{ik}^*]) \right|^2 \right)^\frac{1}{2} + \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=1}^{n_Z} \|\hat{F}_i - H'F_i\|^2 \right)^\frac{1}{2} \left( \frac{1}{n} \sum_{i=1}^{n_Z} \left( \sum_{k=1}^{p} |\mathbb{E}[e_{ki} \epsilon_{ik}^*]| \right)^2 \right)^\frac{1}{2}
\]
\[
= O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}} \cdot \min(n, p)^{-\frac{1}{2}}) + O_P(n^{-\frac{1}{2}})
\]
\[
= o_P(n^{-\frac{1}{2}} p^{\frac{1}{2}}),
\]
by M, T, Lemma A.1 of Bai (2003, p.159), and Assumptions S0(i) and (iii). By the above results on the remainder terms \(\mathcal{R}_1, \mathcal{R}_2,\) and \(\mathcal{R}_3,\) Part (i) follows by T.

For Part (ii), note that \(g(\theta_0) = O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}})\) by Lemma 1(ii), Assumption 2, and CH. Hence, by Part (i) and T, \(\hat{g}(\theta_0) = O_P(n^{-\frac{1}{2}} p^{\frac{1}{2}})\). Similarly, \(\hat{g}(\hat{\theta}) - \hat{g}(\theta_0) = \hat{G}(\hat{\theta} - \theta_0) = o_P(p^{\frac{1}{2}})\) by Lemma 4(iii), CS, and consistency of \(\hat{\theta}\) for \(\theta_0\). Hence, Part (iii) follows by Part (ii) and T. \(\square\)

Web Appendix C  Proof of main results

Web Appendix C.1  Proof of Theorem 1

The F-LIML estimator is given by \(\hat{\theta}_F = \arg\max_\theta \hat{Q}(\theta)\), where \(\hat{Q}(\theta) = -\hat{g}(\theta)'\hat{\Omega}^{-1}\hat{g}(\theta)/2\). Under the conditions of Theorem 1, the estimator \(\hat{\theta}_F\) can be shown to be consistent for \(\theta_0\) by applying standard arguments used to establish consistency of extremum estimators; see, for example, Newey and McFadden (1994, Theorem 2.1, p.2121) and Zhao et al. (2020, Proof of Theorem 3.1). By a first-order Taylor expansion, there exists a \(\hat{\theta}\) in the line segment joining \(\hat{\theta}_F\) and \(\theta_0\) such that \(\nabla_{\theta_0} \hat{Q}(\hat{\theta}_F - \theta_0) = -\nabla_{\theta_0} \hat{Q}(\theta_0)\). Dividing both sides by \(-G'\Omega^{-1}G\), and by Lemmas 10(ii) and (iii) below, it follows that \((G'\Omega^{-1}G)^\frac{1}{2}(\hat{\theta}_F - \theta_0) \overset{D}{\rightarrow} N(0,1)\). Let \(\hat{\Omega} = \hat{\Omega}(\hat{\theta}_F)\). Now, since \(\hat{G}'\hat{\Omega}^{-1}\hat{G} - G'\Omega^{-1}G = \hat{G}'(\hat{\Omega}^{-1} - \Omega^{-1})\hat{G} + (\hat{G} - G)'\Omega^{-1}(\hat{G} - G) = o_P(n)\) by identical arguments used in the Proof of Lemma 9(ii) below. Therefore, \((G'\Omega^{-1}G)^{-1}(\hat{G}'\hat{\Omega}^{-1}\hat{G}) = o_P(1)\), and the result of the theorem follows by Slutsky’s lemma. \(\square\)

Lemma 9. Under Assumptions 1–3, and Equations (1)–(2), as \(n, p \rightarrow \infty\), (i) \(\nabla_{\theta_0} \hat{Q}(\theta_0) = -G'\Omega^{-1}\hat{g}(\theta_0) + o_P(n^{-\frac{1}{2}})\); (ii) \(\nabla_{\theta_0} \hat{Q}(\hat{\theta}) = -G'\Omega^{-1}G + o_P(n)\).

Proof. For Part (i), we have

\[
\nabla_{\theta_0} \hat{Q}(\theta_0) = -\hat{G}'\hat{\Omega}^{-1}\hat{g}(\theta_0) + \theta_0 n^{-\frac{1}{2}} \hat{\sigma}_F^2 \hat{g}(\theta_0)'\hat{\Omega}_0^{-1}\hat{g}(\theta_0)
\]
\[
= -(\hat{G}' - G)'\hat{\Omega}^{-1}\hat{g}(\theta_0) - G'(\hat{\Omega}^{-1} - \Omega^{-1})\hat{g}(\theta_0) - G'\Omega^{-1}\hat{g}(\theta_0) + O_P(1)
\]
\[
= -G'\Omega^{-1}\hat{g}(\theta_0) + O_P(1) + O_P(n^{-\frac{1}{2}} p^{-\frac{1}{2}}) + o_P(n^{\frac{1}{2}})
\]
\[
= -G'\Omega^{-1}\hat{g}(\theta_0) + o_P(n^{\frac{1}{2}}),
\]

[S-8]
where the second equality follows by Lemmas 3(iv), 6(ii), 7(v), 8(ii), and M, and the third equality follows by Lemmas 4(i), 4(ii), 4(iii), 7(v), 7(vi), and 8(ii).

Similarly, for Part (ii),

\[
\nabla_{\theta} Q(\theta) = G'\hat{\Omega}(\theta)^{-1}G' + 4\hat{g}(\theta) + n_{\theta}^{-1}\sigma_{\theta}^2 G'\hat{\Omega}(\theta)^{-1}\hat{g}(\theta) + n_{\theta}^{-1}\sigma_{\theta}^2 \hat{g}(\theta)
\]

\[
-4\hat{g}(\theta) + \hat{g}(\theta)^{\prime}\hat{\Omega}(\theta)\hat{\Omega}(\theta)^{-1}\hat{g}(\theta)
\]

\[
= G'\Omega^{-1}G - \hat{G}(\hat{\theta}) - \hat{G}(\hat{\theta})^{-1}\hat{G}(\hat{\theta})^{-1}\hat{G}(G - \hat{\theta}) + o_{P}(n)
\]

where the second equality follows by Lemmas 3(iv), 6(ii), 7(v), 8(ii), and M, and the third equality follows by Lemmas 4, 7(v), and 7(vi).

Lemma 10. As \(n,p \to \infty\), (i) \(G'\Omega^{-1}G = \Theta(n)\); (ii) \((G'\Omega^{-1}G)^{-\frac{1}{2}}\nabla_{\theta} Q(\theta) \overset{D}{\to} N(0,I_r)\); (iii) \((G'\Omega^{-1}G)^{-1}\nabla_{\theta} Q(\theta) \overset{P}{\to} -I_r\).

Proof. For Part (i), note that \(G'\Omega^{-1}G = G'\Omega^{-1}G - \hat{G}(\hat{\theta}) - \hat{G}(\hat{\theta})^{-1}\hat{G}(\hat{\theta})^{-1}\hat{G}(G - \hat{\theta}) + o_{P}(1) \overset{D}{\to} N(0,I_r)\). By the above, \(G'\Omega^{-1}G \overset{D}{\to} 0\) since \(G'\Omega^{-1}G \overset{P}{\to} -I_r\). By the above, \(G'\Omega^{-1}G \overset{D}{\to} 0\) since \(G'\Omega^{-1}G \overset{P}{\to} -I_r\).

Web Appendix C.2 Proof of Theorem 2

Let \(\Delta_{GG} = H^{-1}\Omega_{0}H^{-1}n_{\theta}^{-1}\sigma_{\theta}^2\) and \(\Delta_{G} = \Delta_{GG}\theta_{0}\).

Lemma 11. As \(n,p \to \infty\), (i) \(\hat{\Delta}_{GG} - \Delta_{GG} = o_{P}(n^{-1}p)\); (ii) \(\hat{\Delta}_{G} - \Delta_{G} = o_{P}(n^{-1}p)\).

Proof. We can write \(\hat{\Delta}_{GG} - \Delta_{GG} = (\hat{\Omega}_{0} - H^{-1}\Omega_{0}H^{-1})n_{\theta}^{-1}\sigma_{\theta}^2 + H^{-1}\Omega_{0}H^{-1}n_{\theta}^{-1}(\sigma_{\theta}^2 + \sigma_{\theta}^2)\), where first term on the RHS is \(o_{P}(n^{-1}p)\) by Lemmas 3(iii), 6(ii), and M, and the second term on the RHS is \(o_{P}(n^{-1}p)\) by Lemmas 1(ii), 3(i), and 6(ii). The result for Part (ii) directly follows since \(\hat{\Delta}_{G} - \Delta_{G} = (\hat{\Delta}_{GG} - \Delta_{GG})\theta_{0}\). Let \(\bar{S}_{0} = \Omega^{-\frac{1}{2}}g(\theta_{0})\) and \(\bar{T}_{0} = (\Delta_{GG} - \Delta_{G}\Omega^{-1}\Delta_{G})^{-\frac{1}{2}}(\hat{G} - \Delta_{G}\Omega^{-1}g(\theta_{0}))\). Under our weak instrument asymptotics, \(N'\delta_{X} = \Theta(n^{-1}p)\), so that \(G = -H^{-1}N'\delta_{X} = O_{P}(n^{-\frac{1}{2}}p)\) by Lemma 1(ii).

By Assumption 1, under \(H_{0} : \theta = \theta_{0}\),

\[
\left[\begin{array}{c}
g(\theta_{0}) \\
\hat{G} - \Delta_{G}\Omega^{-1}g(\theta_{0})
\end{array}\right] \sim N\left(\begin{array}{c}
0 \\
\Omega \\
0 \\
\Delta_{GG} - \Delta_{G}\Omega^{-1}\Delta_{G}
\end{array}\right).
\]

Let \(S_{0} = \Omega^{-\frac{1}{2}}g(\theta_{0})\) and \(T_{0} = (\Omega_{X} - \Delta_{G}\Omega^{-1}\Delta_{G})^{-\frac{1}{2}}(\hat{G} - \Delta_{G}\Omega^{-1}g(\theta_{0}))\). By identical arguments used in Smith (2007, Proof of Theorem 3.2, p.252), conditional on \(Z_{T} \sim N((\Delta_{GG} - \Delta_{G}\Omega^{-1}\Delta_{G})^{-\frac{1}{2}}(\hat{G} - \Delta_{G}\Omega^{-1}g(\theta_{0})), I_{r})\), we have that \(\bar{Q}_{S,0} \overset{D}{\to} \chi_{r}^{2}\),

\[\text{[S-9]}\]
\[ Q_{ST,0}^{-1}Q_{ST,0} \overset{D}{\to} \chi^2(1), \text{ and } (Q_{S,0} - Q_{T,0} + \sqrt{(Q_{S,0} - Q_{T,0})^2 + 4Q_{ST,0}^2})/2 \overset{D}{\to} (\chi^2(1) + \chi^2(r-1) - Z_T^2Z_T + \\
\sqrt{\chi^2(1) + \chi^2(r-1) - Z_T^2Z_T^2 + 4\chi^2(1)(Z_T^2Z_T)})/2. \]

We are left to show that \( \hat{Q}_S = \hat{Q}_{S,0} + o_P(1), \hat{Q}_{ST} = \hat{Q}_{ST,0} + o_P(1), \) and \( \hat{Q}_T = \hat{Q}_{T,0} + o_P(1), \) so that the result of the theorem follows by Slutsky’s lemma.

First, note that \((\hat{G} - \hat{\Delta}_G\hat{\Omega}(\theta_0)^{-1}\hat{g}(\theta_0)) - (\hat{G} - \Delta_G\Omega^{-1}g(\theta_0)) = (\hat{G} - \hat{G}) - (\hat{\Delta}_G - \Delta_G)\hat{\Omega}(\theta_0)^{-1}\hat{g}(\theta_0) - \Delta_G(\hat{\Omega}(\theta_0)^{-1} - \Omega^{-1})\hat{g}(\theta_0) - \Delta_G\Omega^{-1}(\hat{g}(\theta_0) - g(\theta_0)) = (\hat{G} - \hat{G}) + o_P(n^{-\frac{1}{2}}p^\frac{1}{2}) \), by Lemmas 1(i), 3(i), 6(ii), 7(ii), 7(v), 7(vi), 8(i), 8(ii), 11(ii), and M.

Using the expansion of \( \hat{\lambda}_k - H^{-1}\lambda_k, k \in [p] \) from Equation (S.1), we can write

\[
\hat{G} - G = -\sum_{k=1}^p (\hat{\lambda}_k - H^{-1}\lambda_k)\delta_{X_k} - \sum_{k=1}^p (\hat{\lambda}_k - H^{-1}\lambda_k)\varepsilon_{X_k} \\
= H\frac{1}{n_Z} \sum_{i=1}^{n_Z} \sum_{k=1}^p F_i e_{ki} \delta_{X_k} - \left( \frac{1}{n_Z} \sum_{i=1}^{n_Z} \hat{F}_i(F_i - H^{-1}\hat{F}_i) \right) (\Lambda'\delta_X) \\
- \frac{1}{n_Z} \sum_{i=1}^{n_Z} (\hat{F}_i - H'F_i) \left( \sum_{k=1}^p e_{ki} \delta_{X_k} \right) - \sum_{k=1}^p (\hat{\lambda}_k - H^{-1}\lambda_k)\varepsilon_{X_k} \\
:= \mathcal{R}_4 + \mathcal{R}_5 + \mathcal{R}_6 + \mathcal{R}_7.
\]

By CS, \( \|\mathcal{R}_4\| \leq \|H\| C_1 n^{-a} n_Z^{-\frac{1}{2}} \|n_Zp\|^{-\frac{1}{2}} \sum_{i=1}^{n_Z} \sum_{k=1}^p F_i e_{ki} \|O_P(n^{-\frac{1}{2}+a}) = o_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \) for some constant \( C_1 \), by Lemma 1(i), \( \delta_{X_k} = O(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \) for all \( k \in [p] \), and Assumption S0(iv). For \( \mathcal{R}_5 \), note that \( \|\mathcal{R}_5\| \leq \|H^{-1}\| \|n_Z^{-1}\hat{F}_i(FH - \hat{F})\| \|\Lambda'\delta_X\| = O_P(\min(n,p)^{-1}) O_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) = o_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \) by Lemma 1(ii), and Lemma B.3 of Bai (2003, p.165), \( \Lambda'\delta_X = \Theta(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \), and CS. Similarly, \( \|\mathcal{R}_6\| \leq C_1 n^{-a} \left( \sum_{i=1}^{n_Z} \|\hat{F}_i - H'F_i\|^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^{n_Z} \sum_{k=1}^p e_{ki} \right)^{\frac{1}{2}} = o_P(n^{-a} \min(n,p)^{-\frac{1}{2}}) \) for some constant \( C_1 \), where the first equality follows by \( \delta_{X_k} = O(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \) for all \( k \in [p] \), Lemma A.1 of Bai (2003, p.159), Assumption S0(iv), M, and CS. The second equality follows the rate restriction \( pn^{-\frac{1}{2}} \to \infty \), as \( n,p \to \infty \). Finally, \( \|\mathcal{R}_7\| = o_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \) by identical arguments used in Proof of Lemma 8(i). Hence, \((\hat{G} - \hat{\Delta}_G\hat{\Omega}(\theta_0)^{-1}\hat{g}(\theta_0)) - (\hat{G} - \Delta_G\Omega^{-1}g(\theta_0)) = o_P(n^{-\frac{1}{2}}p^{\frac{1}{2}}) \).

Also, \((\Delta_{GG} - \Delta_G\hat{\Omega}(\theta_0)^{-1}\Delta_G) - (\Delta_{GG} - \Delta_G\Omega^{-1}\Delta_G) = (\Delta_{GG} - \Delta_{GG}) - (\Delta_G - \Delta_G)\hat{\Omega}(\theta_0)^{-1}\Delta_G - \Delta_G(\hat{\Omega}(\theta_0)^{-1} - \Omega^{-1})\Delta_G - \Delta_G\Omega^{-1}(\Delta_G - \Delta_G) = o_P(n^{-p}) \) by Lemmas 1(ii), 3(i), 7(ii), 7(vi), 7(vi), 11, and M. Then, by similar arguments used in the proof of Lemmas 7(v) and 7(vi), we have that \((\Delta_{GG} - \hat{\Delta}_G\hat{\Omega}(\theta_0)^{-1}\Delta_G)^{-1} = O_P(np^{-1}) \), \((\Delta_{GG} - \Delta_G\Omega^{-1}\Delta_G)^{-1} = O_P(np^{-1}) \), and \((\hat{\Delta}_{GG} - \hat{\Delta}_G\hat{\Omega}(\theta_0)^{-1}\Delta_G)^{-1} - (\Delta_{GG} - \Delta_G\Omega^{-1}\Delta_G)^{-1} = o_P(np^{-1}) \). Now, note that \( \hat{T} - \hat{T}_0 = ((\hat{\Delta}_{GG} - \hat{\Delta}_G\hat{\Omega}(\theta_0)^{-1}\Delta_G)^{-\frac{1}{2}} - (\Delta_{GG} - \Delta_G\Omega^{-1}\Delta_G)^{-\frac{1}{2}})(\hat{\Gamma}_G - \Delta_G\hat{\Omega}(\theta_0)^{-1}\hat{g}(\theta_0)) + (\Delta_{GG} - \Delta_G\Omega^{-1}\Delta_G)^{-\frac{1}{2}}((\hat{\Delta}_{GG} - \hat{\Delta}_G\hat{\Omega}(\theta_0)^{-1}\Delta_G)^{-\frac{1}{2}} - (\Delta_{GG} - \Delta_G\Omega^{-1}\Delta_G)^{-\frac{1}{2}})\). Therefore, \( \hat{T} - \hat{T}_0 = o_P(1) \).

Next, \( \hat{S} - \hat{S}_0 = (\hat{\Omega}(\theta_0)^{-1} - \Omega^{-1})\hat{g}(\theta_0) + \Omega^{-\frac{1}{2}}(\hat{g}(\theta_0) - g(\theta_0)) = o_P(1) \) by Lemmas 7(v), 7(vi), 8(i), and 8(ii).

Having established the consistency of \( \hat{S} \) for \( S_0 \), and \( \hat{T} \) for \( T_0 \), we have that \( Q_S - Q_{S,0} = (\hat{S} - \hat{S}_0) + \hat{S}_0(\hat{S} - \hat{S}_0) + (\hat{S} - \hat{S}_0)'\hat{S}_0 = o_P(1) \), \( Q_{ST} - Q_{ST,0} = S'T - S_0'T_0 = (\hat{S} - \hat{S}_0)'(\hat{T} - T_0) + \hat{S}_0'(\hat{T} - T_0) = o_P(1) \), and \( Q_T - Q_{T,0} = (\hat{T} - T_0)'(\hat{T} - T_0) + T_0'(T - T_0) + (T - T_0)'T_0 = o_P(1) \).
Web Appendix C.3  Proof of Theorem 3

Let \( \hat{g}_S(\theta) = \Gamma'_S \hat{g}(\theta), \hat{G}_S = \Gamma'_S \hat{G}, \hat{\Omega}_S(\theta) = \Gamma'_S \hat{\Omega}(\theta) \Gamma_S, g_S(\theta) = \Gamma'_S g(\theta), G_S = \Gamma'_S G, \Omega_S(\theta) = \Gamma'_S \Omega(\theta) \Gamma_S, \Omega_S = \Omega_S(\theta_0), \) and \( D \) the \( p \times p \) diagonal matrix with its \( (j,j)\)-th element given by \( (\Delta_{GG})_{jj} \).

We work with the joint normality model

\[
\begin{bmatrix} \hat{\theta}_S \\ \hat{T} \end{bmatrix} \sim N \left( \begin{bmatrix} \theta_0 \\ D^{-\frac{1}{2}}G \end{bmatrix}, \begin{bmatrix} V_S & C'_G \\ C_G & V_G \end{bmatrix} \right),
\]

where \( V_S = (G'_S \Omega_S G_S)^{-1}, C_G = -D^{-\frac{1}{2}} \Delta_{GG} G_S \Omega_S^{-1} G_S V_S \theta_0, \) and \( V_G = D^{-\frac{1}{2}} \Delta_{GG} D^{-\frac{1}{2}}. \)

To remove dependence of \( \hat{\theta}_S \) on \( D^{-\frac{1}{2}} \), we condition on a sufficient statistic for the unknown nuisance parameter \( D^{-\frac{1}{2}} \), which is given by \( U = \hat{T} - C_G V_S^{-1} \hat{\theta}_S. \) Note that \( \hat{\theta}_S \) and \( U \) are uncorrelated. Moreover, conditional on \( U = u \), and under \( \mathcal{H}_0 : \theta = \theta_0, \) we have that \( \hat{T} = u + C_G V_S^{-1} \hat{\theta}_S, \) which is a random variable with distribution \( u + C_G V_S^{-1} (\theta_0 + V_S^{\frac{1}{2}} K) \) where \( K \sim N(0,1) \). Thus, if \( S \) is the selection event, and \( R \) is the set of indices of the estimated factor loadings corresponding to the retained instruments, under \( \mathcal{H}_0 : \theta = \theta_0, \)

\[
P(\hat{\theta}_S \leq w | S, U = u) = \frac{P(\{\hat{\theta}_S \leq w \} \cap S | U = u)}{P(S | U = u)}
= \frac{P(\{\theta_0 + V_S^{\frac{1}{2}} K \leq w \} \cap \{(\hat{u})_j > c_\delta, j \in R\} \cap \{|(\hat{u})_k| \leq c_\delta, k \in [r] \setminus R\})}{P(\{(\hat{u})_j > c_\delta, j \in R\} \cap \{|(\hat{u})_k| \leq c_\delta, k \in [r] \setminus R\})},
\]

where \( \hat{u} = u + C_G V_S^{-1} (\theta_0 + V_S^{\frac{1}{2}} K), \) and \( K \sim N(0,1). \)

The joint normality assumption with variance components \( V_S, C_G, \) and \( V_G \) are motivated by the following arguments. By Proof of Theorem 1, we have that \( \hat{\theta}_S - \theta_0 \approx -V_S G'_S \Omega_S^{-1} \Gamma'_S H^{-1} \Lambda'_Y - \hat{\delta}_Y - \delta_X \theta_0. \) By Proof of Lemma 4(ii), we have that \( D^{-\frac{1}{2}} \hat{G} - D^{-\frac{1}{2}} \hat{G} \approx -H^{-1} \Lambda'(\hat{\delta}_Y - \delta_Y). \) Under Assumption 2, we have that \( \text{Cov}(\Lambda'(\hat{\delta}_Y - \delta_Y), \Lambda'(\hat{\delta}_X - \delta_X)) = -\Omega_0 n^{-1}_X \sigma^2_Y, \) which motivates the expression for \( C_G. \) Also, \( \text{Var}(\Lambda'(\hat{\delta}_X - \delta_X)) = \Omega_0 n^{-1}_X \sigma^2_Y, \) which motivates the expression for \( V_G. \)

To make the result of Theorem 3 operational, we take the empirical distributions for probabilities in the numerator and denominator on the RHS of the expression for \( P(\hat{\theta}_S \leq w | S, U = u) \) above, based on draws of \( K \sim N(0,1). \) Also, we replace the variance components \( V_S, C_G, \) and \( V_G \) with estimates \( \hat{V}_S = (G'_S \hat{\Omega}(\hat{\theta}_S)^{-1} G_S)^{-1}, \hat{C}_G = -D^{-\frac{1}{2}} \hat{\Delta}_{GG} G_S \hat{\Omega}_S(\hat{\theta}_S)^{-1} G_S \hat{V}_S \hat{\theta}_S, \) and \( \hat{V}_G = D^{-\frac{1}{2}} \hat{\Delta}_{GG} D^{-\frac{1}{2}}, \) where \( \hat{D} = \hat{D} \) is a \( p \times p \) diagonal matrix with \( (j,j)\)-th element equal to \( (\hat{\Delta}_{GG})_{jj}. \) We note that the errors from estimating the variance components \( V_S \) and \( C_G, \) and estimation error \( \hat{T} - D^{-\frac{1}{2}} \hat{G}, \) are not ignorable. Therefore, the joint normality approximation used above is not exact.
Web Appendix D  Further simulation results

In this section, we present additional simulation results from the models discussed in Section 4. We briefly describe the alternative summary data methods that we used. The “CLR” test is the conditional likelihood ratio test proposed in Wang and Kang (2021). The “DecorrIVW” test is the de-correlated IVW test when variants are pruned at $R^2 \leq 0.1$ unless otherwise stated. The “RAPS” test is the robust adjusted profile score test proposed in Zhao et al., 2020 under a robust loss function with no overdispersion parameter. The “DIVW” test is the de-biased IVW test proposed in Ye et al. (2021). The “RAD” test is a radial IVW regression approach with second-order weights proposed in Bowden et al. (2019).

For indicating the level of pruning used, the tests CLR-0, CLR-1, CLR-2, and CLR-4 will denote the CLR test pruned to near-independence after selecting the strongest associated variant with LDL-C ($R^2 \leq 0.01$), $R^2 \leq 0.1$, $R^2 \leq 0.2$, and $R^2 \leq 0.4$, respectively. Likewise, RAPS-0, RAPS-1, and other tests are defined analogously.

For the power plots in Figures S1–S5, the first rows (Model 1) present the performance of tests under correct model specification and weak instruments. The second and third rows present the performance of tests when variants have local-to-zero direct effects $\tau$ on the outcome; $\delta_Y = \delta_X \theta_0 + n^{-\frac{1}{2}} \tau$. The second rows (Model 2) describe the setting where all variants have equal effects $\tau$ on the outcome, and the third rows (Model 3) describe the setting $\tau = \tau_2$, where $\tau_2$ is a fixed effect, and each element of $\tau_2$ is drawn from the uniform distribution $U(-1,1)$. The fourth and fifth rows describe the setting where the variant correlation matrix available to the researcher $\bar{\rho}$ is different from the true variant correlation matrix $\rho$ used to generate the summary genetic association data. For the fourth row (Model 4), we assume the correlation estimates available to the researcher satisfies $\bar{\rho}_{jk} = \rho_{jk} - \kappa_0$, for some $\kappa_0 > 0$, and in the fifth rows (Model 5), we assume the correlation estimates available to the researcher satisfies $\bar{\rho}_{jk} = \text{sgn}(\rho_{jk}) \cdot |\rho_{jk}|^{\kappa_2}$, for some $\kappa_2 > 0$.

In Figure S1, the results are presented for other methods not included in the plot of Figure 1 in the main text, for $p = 180$ variants. Figures S2–S3 present power results under $p = 90$ variants, and Figures S4–S5 present power results under $p = 360$ variants. Figure S6 presents histograms (under Model 1) of the standardized S-LIML estimates based on conventional standard errors, varying with the ratio $p/\sqrt{n_Z}$, where $n_Z$ is the sample size used to generate an estimated variance-covariance matrix of variants $\hat{\text{var}}(Z)$. The root mean square error (RMSE) of the S-LIML estimates is also indicated. Figure S7 presents histograms (under Model 1) of the standardized F-LIML estimates varying with the ratio $p/\sqrt{n_Z}$, and RMSE of methods which provide point estimates. We note that the RMSE of alternative summary data methods are not varying with $p/\sqrt{n_Z}$, since they did not require an estimate of $\text{var}(Z)$. Finally, Figure S8 presents the RMSE results of estimators under Model 3 (local-to-zero direct
variant effects on the outcome) when varying the number of estimated factors \( \hat{r} \) away from the true number of factors \( r = 11 \). Figure S9 presents the power performance of the factor-based methods under misspecified correlation structure (Model 5) and for varied number of estimated factors \( \hat{r} \).

Figure S1. Power results when testing the null hypothesis \( H_0 : \theta_0 = 0 \) for a 5% level test (\( p = 180 \) variants). The first row of panels correspond to Model 1 in Section 4.1. The second and third rows of panels
correspond to Models 2 and 3 in Section 4.2. The fourth and last rows of panels correspond to Models 4 and 5 in Section 4.3.

Figure S2. Power results when testing the null hypothesis $H_0 : \theta_0 = 0$ for a 5% level test ($p = 90$ variants).

The first row of panels correspond to Model 1 in Section 4.1. The second and third rows of panels correspond to Models 2 and 3 in Section 4.2. The fourth and last rows of panels correspond to Models 4 and 5 in Section 4.3.
Figure S3. Power results when testing the null hypothesis $H_0 : \theta_0 = 0$ for a 5% level test ($p = 90$ variants).

The first row of panels correspond to Model 1 in Section 4.1. The second and third rows of panels correspond to Models 2 and 3 in Section 4.2. The fourth and last rows of panels correspond to Models 4 and 5 in Section 4.3.
Figure S4. Power results when testing the null hypothesis $H_0 : \theta_0 = 0$ for a 5% level test ($p = 360$ variants).

The first row of panels correspond to Model 1 in Section 4.1. The second and third rows of panels correspond to Models 2 and 3 in Section 4.2. The fourth and last rows of panels correspond to Models 4 and 5 in Section 4.3.
Figure S5. Power results when testing the null hypothesis $H_0 : \theta_0 = 0$ for a 5% level test ($p = 360$ variants).

The first row of panels correspond to Model 1 in Section 4.1. The second and third rows of panels correspond to Models 2 and 3 in Section 4.2. The fourth and last rows of panels correspond to Models 4 and 5 in Section 4.3.
Figure S6. The performance of S-LIML under conventional standard errors for $p = 180$ (Model 1). Here, the $\sqrt{nZ}$-order sampling error in estimates $\hat{\overline{\text{var}}}(Z_j), j \in [p]$, is chosen to maintain a particular value of the ratio $p/\sqrt{nZ}, \text{var}(r)$ notes the variance of the number of selected factors by S-LIML, and the solid black line is the $N(0,1)$ density curve.
Figure S7. RMSE results under correct specification (Model 1) for \( p = 180 \). The histograms correspond to the standardized F-LIML estimates, and the solid black line is the \( N(0, 1) \) density curve. Here, the \( \sqrt{n_Z} \)-order sampling error in estimates \( \hat{\var{Z_j}} \), \( j \in [p] \), is chosen to maintain a particular value of the ratio \( p/\sqrt{n_Z} \).
Figure S8. Root-mean squared error (RMSE) results under locally invalid instruments (Model 3) for $p = 180$ and misspecification $\hat{r} \neq r$. The true number of factors is $r = 11$. 
Figure S9. Power results when testing the null hypothesis $H_0: \theta_0 = 0$ for a 5% level test ($p = 180$ variants) under Model 5 in Section 4.3 and varied number of estimated factors $\hat{r}$ selected as instruments.
Web Appendix D.1 S-LIML tests for non-zero values of the causal effect

Our inference strategy with S-LIML attempts to account for the first stage screening of relevant factors as instruments. There are, however, two cases where this may not be necessary; i.e. where we may be able to make correct inferences by ignoring the instrument selection step, and by simply using the usual GMM-based standard errors (as in Theorem 1). First, the procedure with S-LIML is more of a finite-sample correction and therefore for large enough sample sizes (and thus strong enough instruments) the correction may not have much of an influence. The second case is when we are interested only in testing the null hypothesis that there is no causal effect. However, if we are interested in obtaining confidence intervals for potentially non-zero effects (for example, by inverting the test), then ignoring the correction step may lead to relatively poor coverage under weak instruments.

To illustrate this, we simulated from the same “correct specification” setting in Model 1 described in Section 4.1, and we present the coverage results of confidence intervals with the correction step, denoted S-LIML, and without the correction step, denoted S-LIML (naive). For comparison, we also note the results of the F-LIML method (which does not screen out any weak instruments, and therefore uses the full set of genetic factors as instruments), and F-CLR (which is an identification-robust test, and so asymptotically guarantees correct coverage).

Figure S10 shows that when the instruments are strong enough, it is not necessary to account for the uncertainty in instrument selection based on their relevance. Moreover, even when the instruments are weak, and when the true causal effect is 0, there appears to be no coverage loss from ignoring the post-selection correction step. However, as the true causal effect moves away from 0, our inference strategy with S-LIML is able to achieve better coverage under weak instruments than if the uncertainty in instrument selection is ignored.
Figure S10. Coverage probabilities of confidence intervals under Model 1 and $H_0 : \theta = \theta_0$ for varying $\theta_0$.

The dashed black line indicates nominal coverage was set to 0.95.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71(1), 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.

Bai, J. and S. Ng (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory* 26(6), 1577–1606.

Bowden, J., F. Del Greco M, C. Minelli, Q. Zhao, D. A. Lawlor, and N. A. Sheehan et al. (2019). Improving the accuracy of two-sample summary-data Mendelian randomization: Moving beyond the NOME assumption. *International Journal of Epidemiology* 48(3), 728–742.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4(1), 2111–2245.

Schooling, C. M., P. M. Lopez, Z. Yang, J. V. Zhao, S. L. Au Yeung, and J. V. Huang (2021). Use of multivariable Mendelian randomization to address biases due to competing risk before recruitment. *Frontiers in Genetics* 11(January), 1–10.

Smith, R. J. (2007). Weak instruments and empirical likelihood: a discussion of the papers by D.W.K. Andrews and J.H. Stock and Y. Kitamura. In *Advances in Economics and Econometrics,* [S-23]
Wang, S. and H. Kang (2021). Weak-instrument robust tests in two-sample summary-data Mendelian randomization. *Biometrics (forthcoming)*, 1–15.

Ye, B. T., J. Shao, and H. Kang (2021). Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *Annals of Statistics* 49(4), 2079–2100.

Zhao, Q., J. Wang, G. Hemani, J. Bowden, and D. S. Small (2020). Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals of Statistics* 48(3), 1742 – 1769.