HGGA, Volume *5*

Supplemental information

MRBEE: A bias-corrected multivariable Mendelian randomization method

Noah Lorincz-Comi, Yihe Yang, Gen Li, and Xiaofeng Zhu

Supplementary material 1 of MRBEE: simulation and data analysis

Noah Lorincz-Comi, Yihe Yang, Gen Li, Xiaofeng Zhu

Contents

1 Supplemental Multivariable Simulations

1.1 Simulation settings for MVMR analysis in the main body

We consider the following statistical model which has the same representation as Lin (2023):

$$\boldsymbol{U} = \mathbf{G}\boldsymbol{\gamma}_U + \boldsymbol{e}_U, \tag{1}$$

$$\boldsymbol{X}_k = \mathbf{G}\boldsymbol{\gamma}_{X_k} + 0.25\boldsymbol{U} + \boldsymbol{e}_{X_j}, \quad j = 1, ..., 4 \tag{2}$$

$$\boldsymbol{Y} = \sum_{k=1}^{4} \theta_j \boldsymbol{X}_k + \mathbf{G}\boldsymbol{\alpha} + \boldsymbol{U} + \boldsymbol{e}_Y. \tag{3}$$

To make $\gamma_{X_1}, ..., \gamma_{X_4}$ to have correlation, we generate it from the Gaussian-Uniform copula model:

$$\begin{pmatrix} z_{j1} \\ z_{j2} \\ z_{j3} \\ z_{j4} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & -0.5 & 0.5 \\ 0.5 & 1 & -0.5 & 0.5 \\ -0.5 & -0.5 & 1 & -0.5 \\ 0.5 & 0.5 & -0.5 & 1 \end{pmatrix} \right), \tag{4}$$

$$\gamma_{X_k, j} = \Phi(z_{jk}) \times 0.22, \tag{5}$$

where $\Phi(\cdot)$ is the CDF of standard normal distribution. In this simulation, we consider the compound symmetric structure with a correlation $\mathrm{cor}(z_{jk}, z_{js}) = 0.5$ for all $k \neq s$. As for $\boldsymbol{\gamma}_u$, each element γ_{uj} are independently generated from

$$\gamma_{uj}^* \sim 0.3\mathrm{Unif}(0, 0.1) + 0.7\delta \tag{6}$$

where δ is a point mass at zero. As for α, each element α_j are independently generated from

$$\alpha_j \sim 0.3\mathcal{N}(0.1, 0.2^2) + 0.7\delta \tag{7}$$

where δ is a point mass at zero. The next part is fixing the heritability, which is achieved by

$$\sigma_e^2 = \frac{\mathrm{var}(\mathbf{G}\boldsymbol{\gamma}_{X_k})}{h^2} - 1, \tag{8}$$

where $h^2 = 0.1$ in this simulation. Finally, the random error is generated from

$$\begin{pmatrix} \boldsymbol{e}_U \\ \boldsymbol{e}_{X_1} \\ \boldsymbol{e}_{X_2} \\ \boldsymbol{e}_{X_3} \\ \boldsymbol{e}_{X_4} \\ \boldsymbol{e}_Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \sigma_e^2 \begin{pmatrix} 1 & 0.5 & 0.5 & -0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & -0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & -0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & -0.5 & 1 & -0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 & -0.5 & 1 & 0.6 \\ 0.5 & 0.5 & 0.5 & -0.5 & 0.5 & 1 \end{pmatrix} \right) \tag{9}$$

In Lin (2023), they did not consider the correlations among $\{\gamma_{X_k}\}$ and the error terms, and did not fix the heritability. These are two major adjustments me made.

1.2 Root Mean Square Error

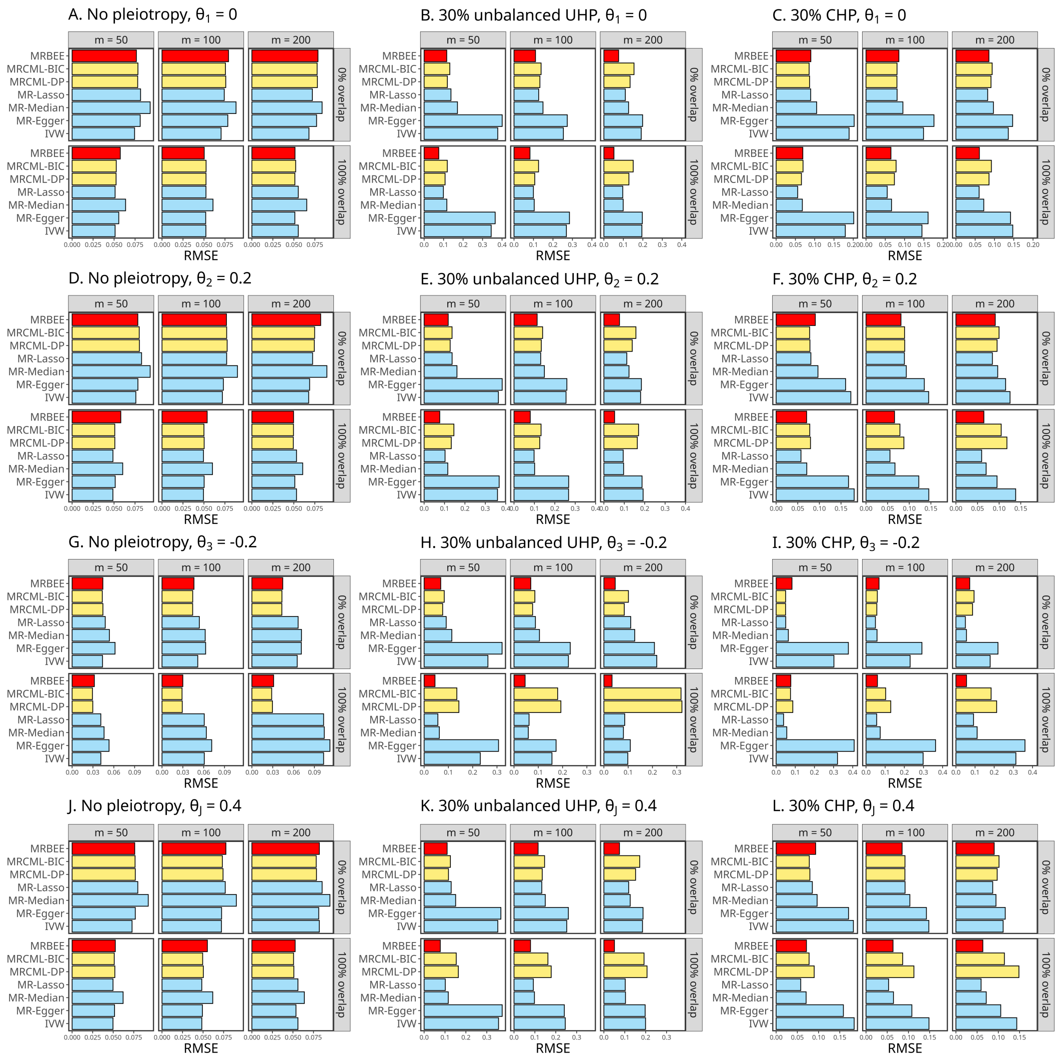


Figure S1: Barplot of the square-root of mean square error (RMSE). Panel A - L displays the barplots of the values of RMSE from seven methods in the MVMR simulation. The four rows represent the four causal effects θ_j, $j = 1, 2, 3, 4$. Each column corresponds to one of the three scenarios. The x-axis indicates the value of RMSE, while the y-axis lists the seven methods.

1.3 Standard Error Evaluation

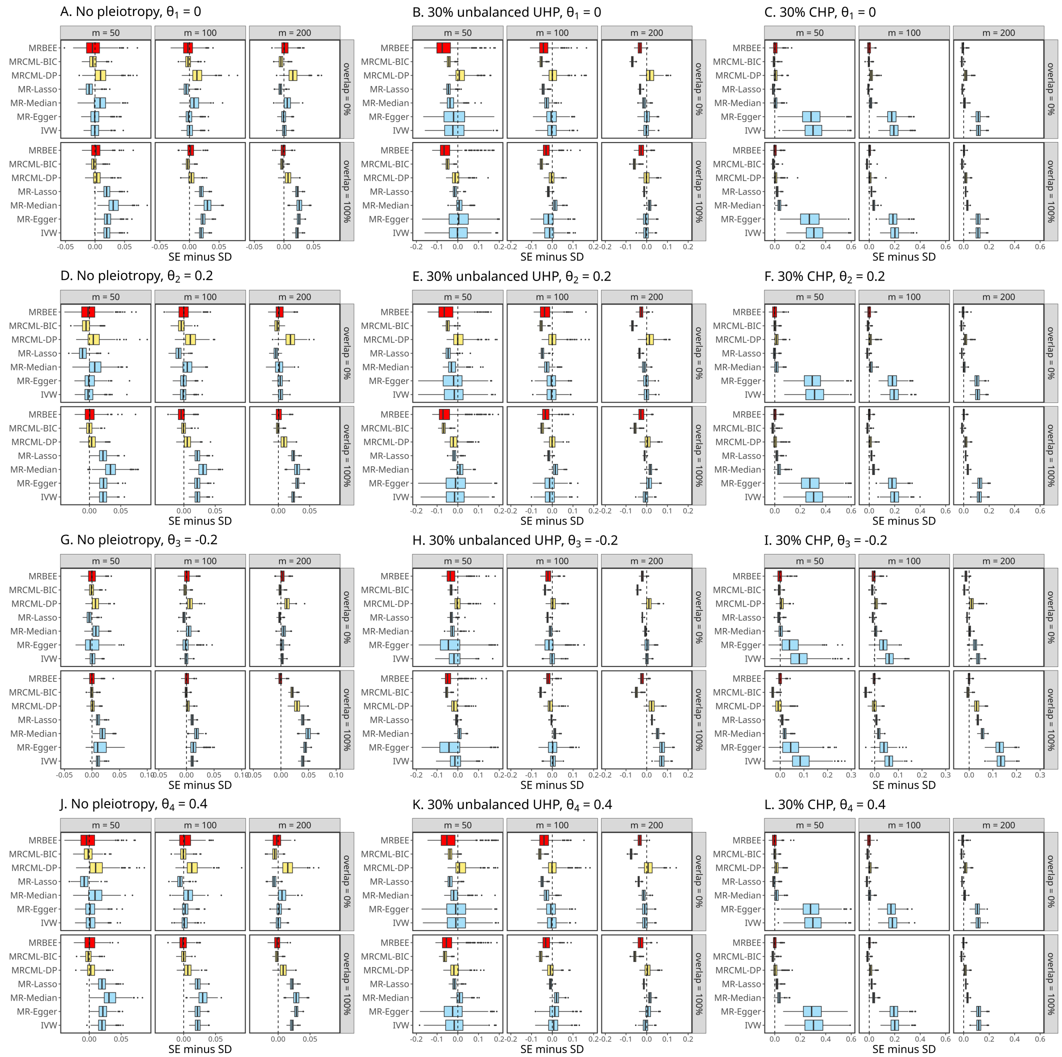


Figure S2: Boxplot of SE minus SD. Panel A - L displays the boxplots of the values of SE minus SD from seven methods in the MVMR simulation. The four rows represent the four causal effects θ_j, $j = 1, 2, 3, 4$. Each column corresponds to one of the three scenarios. The x-axis indicates the value of SE minus SD, while the y-axis lists the seven methods. If SE is correctly estimated, the mean of SE minus SD should be close to zero, which is indicated by a dashed line.

1.4 Coverage Frequency

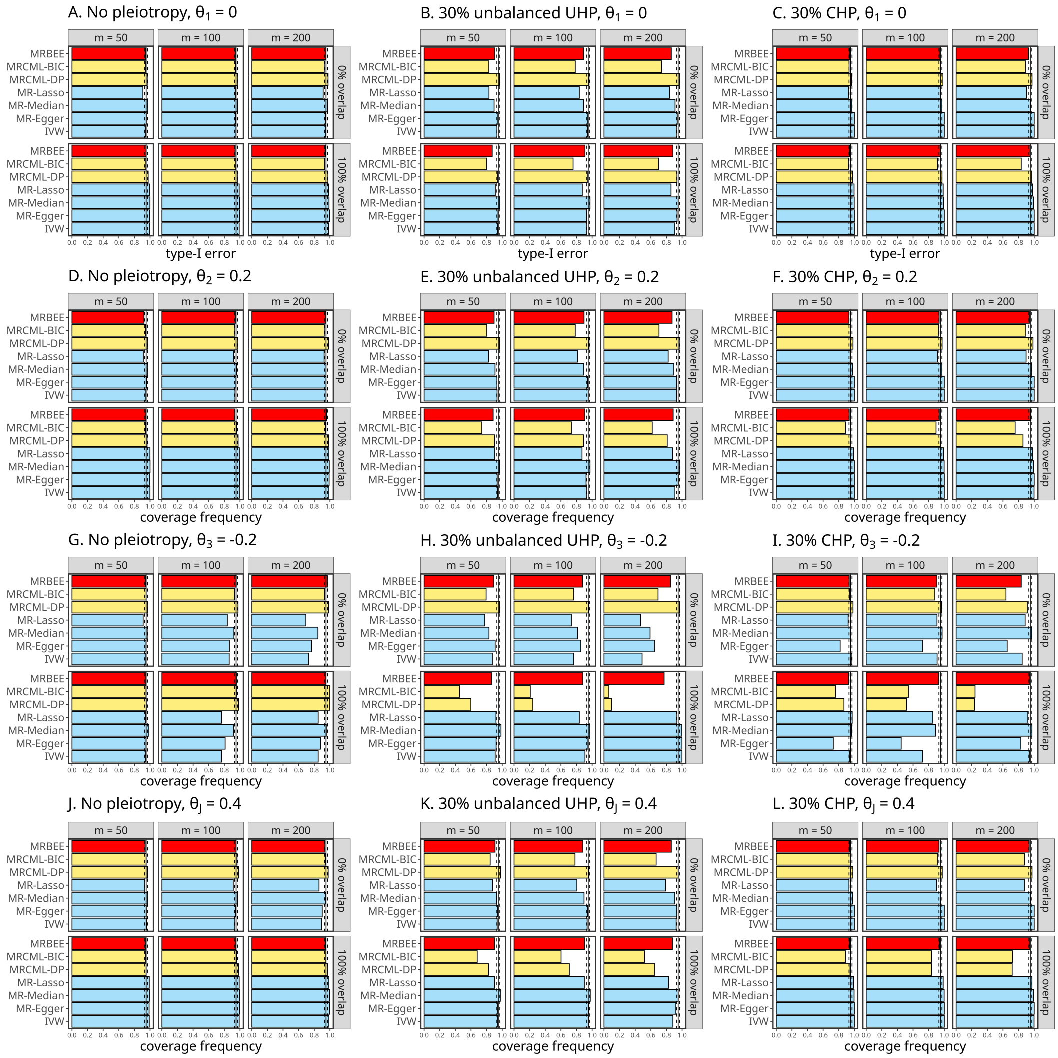


Figure S3: Boxplot of the coverage frequency. Panel A - L displays the boxplots of the values of coverage frequency from seven methods in the MVMR simulation. The four rows represent the four causal effects θ_j, $j = 1, 2, 3, 4$. Each column corresponds to one of the three scenarios. The x-axis indicates the coverage frequency, while the y-axis lists the seven methods. If SE is correctly estimated, the mean of coverage frequency should be around 95

1.5 Summary Table

Table S1. 0% sample overlap, theta1=0, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.003	0.074	0.075	0.940	0.060
	MR-Egger	0.003	0.081	0.082	0.952	0.048
	MR-Median	0.001	0.093	0.102	0.970	0.030
	MR-Lasso	0.006	0.081	0.072	0.908	0.092
	MRCML-DP	0.002	0.078	0.089	0.964	0.036
	MRCML-BIC	0.003	0.078	0.075	0.940	0.060
	MRBEE	0.008	0.081	0.079	0.942	0.058
30% unbalanced UHP	IVW	0.096	0.366	0.351	0.932	0.068
	MR-Egger	0.058	0.397	0.381	0.944	0.056
	MR-Median	0.031	0.169	0.135	0.896	0.104
	MR-Lasso	0.021	0.136	0.093	0.830	0.170
	MRCML-DP	0.005	0.119	0.134	0.966	0.034
	MRCML-BIC	0.008	0.132	0.089	0.828	0.172
	MRBEE	0.004	0.204	0.146	0.904	0.096
30% CHP	IVW	0.094	0.165	0.479	1.000	0.000
	MR-Egger	-0.097	0.178	0.471	1.000	0.000
	MR-Median	0.011	0.105	0.112	0.966	0.034
	MR-Lasso	0.000	0.090	0.080	0.924	0.076
	MRCML-DP	-0.001	0.087	0.095	0.964	0.036
	MRCML-BIC	0.000	0.086	0.080	0.932	0.068
	MRBEE	0.005	0.080	0.087	0.940	0.060

Table S2. 100% sample overlap, theta1=0, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.012	0.050	0.070	0.992	0.008
	MR-Egger	0.008	0.055	0.077	0.994	0.006
	MR-Median	0.012	0.063	0.095	0.998	0.002
	MR-Lasso	0.012	0.050	0.070	0.992	0.008
	MRCML-DP	0.002	0.052	0.056	0.976	0.024
	MRCML-BIC	0.003	0.053	0.051	0.946	0.054
	MRBEE	-0.002	0.051	0.054	0.950	0.050
30% unbalanced UHP	IVW	0.071	0.337	0.342	0.942	0.058
	MR-Egger	0.037	0.363	0.372	0.946	0.054
	MR-Median	0.020	0.115	0.124	0.968	0.032
	MR-Lasso	0.012	0.098	0.085	0.912	0.088
	MRCML-DP	-0.012	0.107	0.099	0.938	0.062
	MRCML-BIC	-0.013	0.118	0.069	0.798	0.202
	MRBEE	-0.004	0.152	0.100	0.872	0.128
30% CHP	IVW	0.099	0.150	0.470	1.000	0.000
	MR-Egger	-0.100	0.175	0.459	1.000	0.000
	MR-Median	0.016	0.066	0.103	0.998	0.002
	MR-Lasso	0.008	0.055	0.074	0.996	0.004
	MRCML-DP	-0.011	0.065	0.073	0.978	0.022
	MRCML-BIC	-0.005	0.070	0.059	0.924	0.076
	MRBEE	-0.005	0.057	0.061	0.938	0.062

Table S3. 0% sample overlap, theta2=0.2, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.011	0.075	0.075	0.946	0.724
	MR-Egger	-0.012	0.078	0.079	0.958	0.676
	MR-Median	-0.014	0.092	0.102	0.966	0.460
	MR-Lasso	-0.013	0.082	0.072	0.914	0.726
	MRCML-DP	-0.006	0.080	0.088	0.966	0.600
	MRCML-BIC	-0.006	0.080	0.075	0.944	0.748
	MRBEE	-0.005	0.078	0.077	0.922	0.718
30% unbalanced UHP	IVW	0.043	0.362	0.348	0.934	0.124
	MR-Egger	0.013	0.383	0.362	0.932	0.122
	MR-Median	-0.002	0.161	0.135	0.908	0.348
	MR-Lasso	-0.004	0.137	0.092	0.824	0.558
	MRCML-DP	-0.002	0.128	0.135	0.966	0.344
	MRCML-BIC	-0.001	0.137	0.089	0.802	0.582
	MRBEE	0.022	0.193	0.145	0.896	0.432
30% CHP	IVW	0.084	0.151	0.476	1.000	0.002
	MR-Egger	-0.036	0.157	0.462	1.000	0.002
	MR-Median	0.009	0.096	0.111	0.978	0.490
	MR-Lasso	0.002	0.080	0.079	0.946	0.730
	MRCML-DP	0.006	0.078	0.094	0.982	0.618
	MRCML-BIC	0.005	0.077	0.080	0.950	0.742
	MRBEE	0.005	0.087	0.088	0.930	0.660

Table S4. 100% sample overlap, theta2=0.2, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.008	0.048	0.070	0.998	0.928
	MR-Egger	0.008	0.051	0.074	0.996	0.900
	MR-Median	0.006	0.060	0.094	1.000	0.638
	MR-Lasso	0.008	0.048	0.070	0.998	0.928
	MRCML-DP	-0.004	0.051	0.055	0.962	0.950
	MRCML-BIC	-0.002	0.051	0.051	0.944	0.964
	MRBEE	0.000	0.053	0.054	0.946	0.940
30% unbalanced UHP	IVW	0.069	0.353	0.342	0.940	0.138
	MR-Egger	0.044	0.365	0.357	0.936	0.118
	MR-Median	0.023	0.114	0.125	0.966	0.468
	MR-Lasso	0.010	0.102	0.085	0.904	0.682
	MRCML-DP	-0.063	0.118	0.101	0.902	0.364
	MRCML-BIC	-0.050	0.137	0.070	0.738	0.608
	MRBEE	-0.001	0.157	0.101	0.884	0.632
30% CHP	IVW	0.098	0.152	0.472	1.000	0.000
	MR-Egger	-0.026	0.166	0.454	1.000	0.000
	MR-Median	0.010	0.070	0.105	1.000	0.542
	MR-Lasso	0.003	0.057	0.075	0.990	0.828
	MRCML-DP	-0.039	0.069	0.075	0.954	0.626
	MRCML-BIC	-0.025	0.074	0.059	0.886	0.814
	MRBEE	-0.001	0.058	0.061	0.934	0.892

Table S5. 0% sample overlap, theta3=-0.2, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.014	0.041	0.043	0.946	0.988
	MR-Egger	0.012	0.061	0.062	0.948	0.846
	MR-Median	0.014	0.052	0.059	0.964	0.904
	MR-Lasso	0.014	0.046	0.041	0.912	0.984
	MRCML-DP	0.002	0.044	0.051	0.970	0.984
	MRCML-BIC	-0.001	0.043	0.043	0.950	0.994
	MRBEE	0.001	0.044	0.044	0.942	0.990
30% unbalanced UHP	IVW	0.152	0.215	0.199	0.874	0.098
	MR-Egger	0.086	0.310	0.281	0.910	0.080
	MR-Median	0.054	0.101	0.076	0.830	0.568
	MR-Lasso	0.038	0.083	0.053	0.776	0.800
	MRCML-DP	0.005	0.076	0.077	0.966	0.718
	MRCML-BIC	0.000	0.083	0.051	0.794	0.918
	MRBEE	0.014	0.109	0.081	0.892	0.678
30% CHP	IVW	-0.249	0.171	0.254	0.960	0.352
	MR-Egger	-0.317	0.204	0.248	0.818	0.546
	MR-Median	-0.019	0.061	0.063	0.952	0.930
	MR-Lasso	0.001	0.050	0.045	0.920	0.972
	MRCML-DP	-0.008	0.049	0.055	0.982	0.978
	MRCML-BIC	-0.011	0.049	0.045	0.940	0.992
	MRBEE	-0.013	0.051	0.050	0.934	0.982

Table S6. 100% sample overlap, theta3=-0.2, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.030	0.029	0.040	0.940	1.000
	MR-Egger	-0.031	0.043	0.058	0.948	0.948
	MR-Median	-0.030	0.035	0.054	0.984	1.000
	MR-Lasso	-0.030	0.029	0.040	0.940	1.000
	MRCML-DP	0.002	0.030	0.031	0.952	1.000
	MRCML-BIC	-0.001	0.029	0.029	0.948	1.000
	MRBEE	0.004	0.030	0.031	0.946	0.998
30% unbalanced UHP	IVW	0.109	0.204	0.194	0.910	0.110
	MR-Egger	0.062	0.300	0.274	0.922	0.108
	MR-Median	-0.007	0.062	0.071	0.984	0.844
	MR-Lasso	-0.016	0.054	0.048	0.920	0.968
	MRCML-DP	0.117	0.082	0.067	0.598	0.342
	MRCML-BIC	0.097	0.094	0.041	0.454	0.664
	MRBEE	0.029	0.098	0.056	0.864	0.830
30% CHP	IVW	-0.276	0.162	0.250	0.940	0.476
	MR-Egger	-0.357	0.197	0.243	0.730	0.672
	MR-Median	-0.039	0.039	0.058	0.964	0.992
	MR-Lasso	-0.021	0.033	0.042	0.972	1.000
	MRCML-DP	0.064	0.060	0.055	0.866	0.690
	MRCML-BIC	0.039	0.064	0.034	0.760	0.896
	MRBEE	0.003	0.035	0.034	0.926	0.998

Table S7. 0% sample overlap, theta4=0.4, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.008	0.073	0.075	0.958	0.996
	MR-Egger	-0.009	0.076	0.079	0.952	0.996
	MR-Median	-0.004	0.093	0.103	0.970	0.970
	MR-Lasso	-0.008	0.080	0.072	0.932	0.994
	MRCML-DP	0.004	0.077	0.089	0.968	0.994
	MRCML-BIC	0.003	0.077	0.076	0.950	0.998
	MRBEE	-0.003	0.080	0.079	0.938	0.994
30% unbalanced UHP	IVW	0.021	0.352	0.349	0.936	0.236
	MR-Egger	-0.006	0.368	0.365	0.942	0.204
	MR-Median	-0.005	0.151	0.136	0.928	0.826
	MR-Lasso	-0.014	0.129	0.092	0.876	0.940
	MRCML-DP	0.004	0.117	0.136	0.980	0.868
	MRCML-BIC	0.006	0.126	0.089	0.846	0.968
	MRBEE	0.011	0.182	0.145	0.908	0.794
30% CHP	IVW	0.082	0.163	0.474	1.000	0.008
	MR-Egger	-0.035	0.167	0.460	1.000	0.004
	MR-Median	0.002	0.097	0.111	0.978	0.962
	MR-Lasso	-0.001	0.085	0.079	0.930	0.994
	MRCML-DP	0.005	0.079	0.095	0.980	0.986
	MRCML-BIC	0.006	0.078	0.079	0.952	0.998
	MRBEE	0.000	0.088	0.087	0.938	0.978

Table S8. 100% sample overlap, theta4=0.4, number of IVs = 50

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.010	0.049	0.069	0.990	1.000
	MR-Egger	0.009	0.051	0.073	0.988	1.000
	MR-Median	0.011	0.061	0.094	0.992	1.000
	MR-Lasso	0.010	0.049	0.069	0.990	1.000
	MRCML-DP	-0.004	0.052	0.055	0.950	1.000
	MRCML-BIC	-0.001	0.051	0.050	0.940	1.000
	MRBEE	-0.002	0.052	0.053	0.940	1.000
30% unbalanced UHP	IVW	0.046	0.354	0.341	0.942	0.242
	MR-Egger	0.026	0.373	0.354	0.938	0.242
	MR-Median	0.013	0.115	0.126	0.976	0.922
	MR-Lasso	0.014	0.101	0.085	0.900	0.982
	MRCML-DP	-0.113	0.119	0.105	0.824	0.742
	MRCML-BIC	-0.079	0.132	0.070	0.680	0.906
	MRBEE	-0.001	0.142	0.099	0.902	0.900
30% CHP	IVW	0.097	0.157	0.473	1.000	0.018
	MR-Egger	-0.029	0.156	0.455	1.000	0.000
	MR-Median	0.015	0.069	0.104	0.996	0.994
	MR-Lasso	0.011	0.057	0.075	0.990	1.000
	MRCML-DP	-0.052	0.073	0.080	0.944	0.962
	MRCML-BIC	-0.024	0.074	0.060	0.888	0.986
	MRBEE	0.000	0.059	0.060	0.938	0.998

Table S9. 0% sample overlap, theta1=0, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.003	0.070	0.071	0.948	0.052
	MR-Egger	0.001	0.078	0.078	0.942	0.058
	MR-Median	-0.005	0.088	0.097	0.968	0.032
	MR-Lasso	-0.003	0.074	0.069	0.940	0.060
	MRCML-DP	-0.006	0.076	0.090	0.972	0.028
	MRCML-BIC	-0.004	0.076	0.074	0.942	0.058
	MRBEE	-0.003	0.079	0.077	0.940	0.060
30% unbalanced UHP	IVW	0.061	0.246	0.246	0.940	0.060
	MR-Egger	0.032	0.272	0.270	0.944	0.056
	MR-Median	0.012	0.148	0.122	0.890	0.110
	MR-Lasso	0.007	0.127	0.083	0.836	0.164
	MRCML-DP	0.005	0.133	0.139	0.960	0.040
	MRCML-BIC	0.013	0.139	0.086	0.784	0.216
	MRBEE	0.009	0.175	0.135	0.890	0.110
30% CHP	IVW	0.082	0.125	0.322	1.000	0.000
	MR-Egger	-0.104	0.143	0.323	1.000	0.000
	MR-Median	0.013	0.095	0.107	0.968	0.032
	MR-Lasso	0.007	0.080	0.074	0.952	0.048
	MRCML-DP	0.000	0.080	0.097	0.982	0.018
	MRCML-BIC	0.003	0.080	0.078	0.952	0.048
	MRBEE	0.017	0.086	0.085	0.944	0.056

Table S10. 100% sample overlap, theta1=0, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.022	0.047	0.067	0.990	0.010
	MR-Egger	0.012	0.051	0.074	0.996	0.004
	MR-Median	0.019	0.058	0.088	0.996	0.004
	MR-Lasso	0.023	0.047	0.067	0.990	0.010
	MRCML-DP	0.001	0.052	0.056	0.966	0.034
	MRCML-BIC	0.004	0.052	0.050	0.946	0.054
	MRBEE	-0.002	0.050	0.052	0.938	0.062
30% unbalanced UHP	IVW	0.088	0.254	0.241	0.926	0.074
	MR-Egger	0.044	0.282	0.264	0.928	0.072
	MR-Median	0.032	0.099	0.113	0.972	0.028
	MR-Lasso	0.024	0.097	0.078	0.874	0.126
	MRCML-DP	-0.032	0.102	0.099	0.938	0.062
	MRCML-BIC	-0.033	0.122	0.070	0.756	0.244
	MRBEE	0.000	0.117	0.091	0.908	0.092
30% CHP	IVW	0.097	0.108	0.311	1.000	0.000
	MR-Egger	-0.105	0.122	0.310	1.000	0.000
	MR-Median	0.022	0.062	0.098	0.990	0.010
	MR-Lasso	0.016	0.053	0.071	0.986	0.014
	MRCML-DP	-0.026	0.069	0.076	0.970	0.030
	MRCML-BIC	-0.020	0.076	0.059	0.912	0.088
	MRBEE	0.000	0.054	0.058	0.954	0.046

Table S11. 0% sample overlap, theta2=0.2, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.008	0.072	0.071	0.936	0.784
	MR-Egger	-0.006	0.073	0.073	0.948	0.764
	MR-Median	-0.006	0.090	0.096	0.958	0.540
	MR-Lasso	-0.009	0.077	0.069	0.920	0.784
	MRCML-DP	0.001	0.078	0.089	0.972	0.662
	MRCML-BIC	0.003	0.077	0.073	0.934	0.794
	MRBEE	-0.001	0.077	0.077	0.940	0.734
30% unbalanced UHP	IVW	0.049	0.250	0.248	0.950	0.178
	MR-Egger	0.028	0.258	0.254	0.942	0.134
	MR-Median	0.012	0.149	0.123	0.892	0.434
	MR-Lasso	0.007	0.131	0.084	0.812	0.648
	MRCML-DP	0.016	0.133	0.137	0.960	0.376
	MRCML-BIC	0.018	0.139	0.086	0.784	0.656
	MRBEE	0.031	0.166	0.135	0.896	0.436
30% CHP	IVW	0.073	0.125	0.323	1.000	0.000
	MR-Egger	-0.022	0.133	0.316	1.000	0.000
	MR-Median	0.007	0.093	0.107	0.968	0.504
	MR-Lasso	0.000	0.089	0.074	0.912	0.736
	MRCML-DP	0.010	0.089	0.097	0.972	0.600
	MRCML-BIC	0.012	0.088	0.078	0.928	0.756
	MRBEE	0.005	0.084	0.084	0.934	0.686

Table S12. 100% sample overlap, theta2=0.2, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.019	0.046	0.068	0.992	0.970
	MR-Egger	0.016	0.048	0.069	0.996	0.952
	MR-Median	0.018	0.057	0.088	0.994	0.770
	MR-Lasso	0.019	0.046	0.068	0.992	0.970
	MRCML-DP	-0.002	0.050	0.056	0.976	0.950
	MRCML-BIC	0.000	0.050	0.050	0.956	0.980
	MRBEE	-0.001	0.056	0.052	0.934	0.950
30% unbalanced UHP	IVW	0.088	0.253	0.242	0.926	0.244
	MR-Egger	0.065	0.260	0.248	0.922	0.202
	MR-Median	0.031	0.096	0.113	0.972	0.536
	MR-Lasso	0.023	0.096	0.079	0.870	0.758
	MRCML-DP	-0.079	0.098	0.099	0.890	0.258
	MRCML-BIC	-0.059	0.119	0.070	0.734	0.518
	MRBEE	0.007	0.120	0.090	0.906	0.662
30% CHP	IVW	0.092	0.112	0.312	1.000	0.006
	MR-Egger	-0.008	0.122	0.303	1.000	0.002
	MR-Median	0.022	0.063	0.097	0.998	0.706
	MR-Lasso	0.016	0.053	0.071	0.992	0.920
	MRCML-DP	-0.054	0.069	0.077	0.950	0.506
	MRCML-BIC	-0.030	0.073	0.059	0.894	0.774
	MRBEE	0.000	0.057	0.057	0.936	0.928

Table S13. 0% sample overlap, theta3=-0.2, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.032	0.040	0.040	0.862	0.986
	MR-Egger	0.035	0.052	0.051	0.864	0.890
	MR-Median	0.034	0.052	0.056	0.922	0.852
	MR-Lasso	0.032	0.043	0.039	0.838	0.984
	MRCML-DP	0.003	0.044	0.051	0.974	0.988
	MRCML-BIC	0.002	0.044	0.042	0.930	1.000
	MRBEE	-0.002	0.043	0.045	0.952	0.994
30% unbalanced UHP	IVW	0.174	0.141	0.141	0.764	0.062
	MR-Egger	0.132	0.190	0.179	0.854	0.072
	MR-Median	0.070	0.078	0.070	0.814	0.476
	MR-Lasso	0.054	0.070	0.048	0.734	0.782
	MRCML-DP	-0.011	0.076	0.081	0.958	0.754
	MRCML-BIC	-0.016	0.085	0.050	0.764	0.926
	MRBEE	0.024	0.095	0.076	0.876	0.646
30% CHP	IVW	-0.203	0.109	0.172	0.908	0.760
	MR-Egger	-0.260	0.133	0.170	0.720	0.834
	MR-Median	-0.016	0.055	0.060	0.972	0.958
	MR-Lasso	0.004	0.049	0.041	0.906	0.990
	MRCML-DP	-0.024	0.051	0.058	0.964	0.984
	MRCML-BIC	-0.024	0.054	0.044	0.878	0.998
	MRBEE	-0.028	0.056	0.052	0.904	0.990

Table S14. 100% sample overlap, theta3=-0.2, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.054	0.028	0.038	0.764	1.000
	MR-Egger	-0.062	0.036	0.049	0.810	0.996
	MR-Median	-0.054	0.034	0.052	0.916	1.000
	MR-Lasso	-0.054	0.028	0.038	0.764	1.000
	MRCML-DP	0.004	0.029	0.031	0.982	1.000
	MRCML-BIC	0.000	0.029	0.028	0.954	1.000
	MRBEE	0.006	0.030	0.031	0.954	1.000
30% unbalanced UHP	IVW	0.081	0.134	0.136	0.908	0.138
	MR-Egger	0.034	0.170	0.173	0.946	0.180
	MR-Median	-0.027	0.052	0.065	0.970	0.956
	MR-Lasso	-0.039	0.049	0.044	0.836	0.994
	MRCML-DP	0.177	0.077	0.068	0.240	0.134
	MRCML-BIC	0.151	0.097	0.041	0.208	0.488
	MRBEE	0.029	0.069	0.052	0.880	0.828
30% CHP	IVW	-0.278	0.109	0.169	0.722	0.934
	MR-Egger	-0.340	0.127	0.166	0.448	0.948
	MR-Median	-0.063	0.039	0.056	0.888	1.000
	MR-Lasso	-0.045	0.033	0.040	0.852	1.000
	MRCML-DP	0.115	0.059	0.058	0.516	0.376
	MRCML-BIC	0.073	0.072	0.034	0.544	0.792
	MRBEE	0.006	0.036	0.034	0.928	0.996

Table S15. 0% sample overlap, theta4=0.4, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.021	0.069	0.071	0.934	0.998
	MR-Egger	-0.019	0.070	0.073	0.942	1.000
	MR-Median	-0.023	0.088	0.095	0.968	0.984
	MR-Lasso	-0.021	0.074	0.068	0.914	1.000
	MRCML-DP	0.001	0.074	0.088	0.980	0.996
	MRCML-BIC	0.000	0.074	0.073	0.962	1.000
	MRBEE	0.005	0.075	0.078	0.950	0.998
30% unbalanced UHP	IVW	0.039	0.250	0.248	0.936	0.434
	MR-Egger	0.021	0.260	0.255	0.942	0.404
	MR-Median	-0.003	0.150	0.123	0.896	0.856
	MR-Lasso	-0.014	0.132	0.084	0.804	0.954
	MRCML-DP	0.015	0.135	0.138	0.958	0.840
	MRCML-BIC	0.014	0.146	0.086	0.780	0.956
	MRBEE	0.017	0.172	0.136	0.880	0.814
30% CHP	IVW	0.063	0.134	0.321	1.000	0.122
	MR-Egger	-0.030	0.139	0.314	1.000	0.056
	MR-Median	0.001	0.103	0.107	0.950	0.954
	MR-Lasso	-0.009	0.092	0.074	0.900	0.990
	MRCML-DP	0.014	0.091	0.097	0.974	0.986
	MRCML-BIC	0.013	0.091	0.078	0.916	0.996
	MRBEE	0.007	0.086	0.085	0.932	0.998

Table S16. 100% sample overlap, theta4=0.4, number of IVs = 100

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.018	0.045	0.067	0.990	1.000
	MR-Egger	0.014	0.047	0.069	0.994	1.000
	MR-Median	0.022	0.058	0.088	0.994	1.000
	MR-Lasso	0.018	0.045	0.067	0.990	1.000
	MRCML-DP	-0.008	0.050	0.056	0.980	1.000
	MRCML-BIC	-0.003	0.049	0.049	0.958	1.000
	MRBEE	-0.002	0.053	0.052	0.946	1.000
30% unbalanced UHP	IVW	0.067	0.236	0.243	0.952	0.514
	MR-Egger	0.044	0.237	0.249	0.954	0.432
	MR-Median	0.031	0.093	0.114	0.976	0.978
	MR-Lasso	0.031	0.087	0.079	0.900	0.998
	MRCML-DP	-0.140	0.109	0.102	0.708	0.710
	MRCML-BIC	-0.103	0.126	0.071	0.602	0.878
	MRBEE	-0.009	0.118	0.091	0.904	0.944
30% CHP	IVW	0.100	0.109	0.311	1.000	0.192
	MR-Egger	-0.002	0.108	0.303	1.000	0.052
	MR-Median	0.019	0.062	0.097	0.998	0.998
	MR-Lasso	0.015	0.051	0.071	0.986	1.000
	MRCML-DP	-0.090	0.068	0.080	0.836	0.964
	MRCML-BIC	-0.047	0.073	0.059	0.836	0.996
	MRBEE	-0.007	0.055	0.057	0.938	1.000

Table S17. 0% sample overlap, theta1=0, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.005	0.068	0.069	0.946	0.054
	MR-Egger	0.010	0.076	0.076	0.942	0.058
	MR-Median	0.003	0.084	0.090	0.970	0.030
	MR-Lasso	0.005	0.072	0.066	0.914	0.086
	MRCML-DP	-0.002	0.078	0.094	0.976	0.024
	MRCML-BIC	0.003	0.078	0.073	0.928	0.072
	MRBEE	0.002	0.076	0.078	0.948	0.052
30% unbalanced UHP	IVW	0.070	0.179	0.176	0.926	0.074
	MR-Egger	0.043	0.194	0.195	0.938	0.062
	MR-Median	0.027	0.124	0.111	0.910	0.090
	MR-Lasso	0.013	0.109	0.078	0.840	0.160
	MRCML-DP	0.005	0.135	0.153	0.966	0.034
	MRCML-BIC	0.010	0.154	0.085	0.738	0.262
	MRBEE	0.010	0.140	0.107	0.862	0.138
30% CHP	IVW	0.085	0.106	0.222	1.000	0.000
	MR-Egger	-0.092	0.115	0.229	1.000	0.000
	MR-Median	0.027	0.094	0.100	0.952	0.048
	MR-Lasso	0.015	0.081	0.070	0.902	0.098
	MRCML-DP	0.009	0.091	0.107	0.972	0.028
	MRCML-BIC	0.014	0.093	0.079	0.890	0.110
	MRBEE	0.003	0.089	0.083	0.926	0.074

Table S18. 100% sample overlap, theta1=0, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.036	0.042	0.065	0.982	0.018
	MR-Egger	0.022	0.046	0.072	0.994	0.006
	MR-Median	0.036	0.055	0.081	0.988	0.012
	MR-Lasso	0.036	0.042	0.065	0.982	0.018
	MRCML-DP	-0.003	0.051	0.058	0.974	0.026
	MRCML-BIC	-0.001	0.052	0.049	0.932	0.068
	MRBEE	-0.001	0.051	0.053	0.940	0.060
30% unbalanced UHP	IVW	0.098	0.169	0.165	0.906	0.094
	MR-Egger	0.064	0.187	0.183	0.936	0.064
	MR-Median	0.047	0.087	0.101	0.950	0.050
	MR-Lasso	0.046	0.085	0.073	0.858	0.142
	MRCML-DP	-0.064	0.112	0.112	0.934	0.066
	MRCML-BIC	-0.067	0.135	0.076	0.700	0.300
	MRBEE	-0.001	0.120	0.095	0.882	0.118
30% CHP	IVW	0.108	0.100	0.213	1.000	0.000
	MR-Egger	-0.096	0.104	0.218	1.000	0.000
	MR-Median	0.036	0.063	0.091	0.990	0.010
	MR-Lasso	0.030	0.052	0.068	0.978	0.022
	MRCML-DP	-0.050	0.070	0.087	0.968	0.032
	MRCML-BIC	-0.049	0.078	0.065	0.834	0.166
	MRBEE	0.000	0.057	0.059	0.948	0.052

Table S19. 0% sample overlap, theta2=0.2, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.020	0.065	0.069	0.950	0.744
	MR-Egger	-0.018	0.067	0.070	0.950	0.748
	MR-Median	-0.018	0.088	0.089	0.950	0.524
	MR-Lasso	-0.019	0.070	0.066	0.926	0.764
	MRCML-DP	0.001	0.074	0.094	0.982	0.592
	MRCML-BIC	0.002	0.075	0.073	0.928	0.794
	MRBEE	0.000	0.076	0.078	0.946	0.742
30% unbalanced UHP	IVW	0.037	0.176	0.176	0.930	0.266
	MR-Egger	0.027	0.182	0.179	0.930	0.248
	MR-Median	-0.005	0.123	0.111	0.894	0.432
	MR-Lasso	-0.005	0.112	0.077	0.822	0.656
	MRCML-DP	0.030	0.136	0.152	0.964	0.326
	MRCML-BIC	0.034	0.153	0.085	0.704	0.656
	MRBEE	0.005	0.133	0.108	0.872	0.462
30% CHP	IVW	0.056	0.112	0.222	1.000	0.046
	MR-Egger	-0.008	0.115	0.220	1.000	0.014
	MR-Median	0.002	0.097	0.101	0.964	0.542
	MR-Lasso	-0.002	0.084	0.070	0.896	0.768
	MRCML-DP	0.024	0.092	0.108	0.984	0.586
	MRCML-BIC	0.026	0.097	0.079	0.892	0.774
	MRBEE	0.015	0.087	0.083	0.938	0.740

Table S20. 100% sample overlap, theta2=0.2, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.034	0.041	0.065	0.986	0.998
	MR-Egger	0.030	0.041	0.072	0.990	0.988
	MR-Median	0.033	0.051	0.081	0.992	0.936
	MR-Lasso	0.034	0.041	0.065	0.986	0.998
	MRCML-DP	-0.005	0.049	0.058	0.982	0.940
	MRCML-BIC	0.000	0.050	0.049	0.944	0.980
	MRBEE	0.000	0.052	0.053	0.956	0.956
30% unbalanced UHP	IVW	0.087	0.172	0.165	0.908	0.414
	MR-Egger	0.075	0.171	0.183	0.952	0.326
	MR-Median	0.049	0.084	0.101	0.966	0.742
	MR-Lasso	0.045	0.083	0.073	0.878	0.882
	MRCML-DP	-0.125	0.107	0.112	0.810	0.108
	MRCML-BIC	-0.108	0.132	0.076	0.618	0.388
	MRBEE	-0.005	0.121	0.096	0.886	0.534
30% CHP	IVW	0.104	0.091	0.213	0.998	0.112
	MR-Egger	0.028	0.091	0.218	1.000	0.016
	MR-Median	0.037	0.059	0.091	0.990	0.830
	MR-Lasso	0.029	0.052	0.068	0.980	0.976
	MRCML-DP	-0.095	0.070	0.087	0.856	0.190
	MRCML-BIC	-0.071	0.077	0.065	0.756	0.520
	MRBEE	-0.003	0.058	0.060	0.958	0.916

Table S21. 0% sample overlap, theta3=-0.2, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.054	0.036	0.039	0.728	0.972
	MR-Egger	0.057	0.043	0.045	0.764	0.886
	MR-Median	0.053	0.048	0.053	0.846	0.814
	MR-Lasso	0.054	0.040	0.037	0.692	0.970
	MRCML-DP	0.000	0.043	0.055	0.982	0.980
	MRCML-BIC	-0.003	0.043	0.042	0.932	0.998
	MRBEE	0.001	0.042	0.045	0.968	0.994
30% unbalanced UHP	IVW	0.195	0.098	0.099	0.488	0.048
	MR-Egger	0.175	0.113	0.115	0.646	0.052
	MR-Median	0.106	0.070	0.063	0.590	0.356
	MR-Lasso	0.090	0.065	0.044	0.468	0.648
	MRCML-DP	-0.028	0.079	0.091	0.968	0.740
	MRCML-BIC	-0.039	0.093	0.049	0.692	0.952
	MRBEE	-0.020	0.083	0.062	0.852	0.890
30% CHP	IVW	-0.159	0.082	0.120	0.846	0.962
	MR-Egger	-0.199	0.094	0.119	0.654	0.980
	MR-Median	-0.008	0.055	0.057	0.968	0.958
	MR-Lasso	0.013	0.048	0.039	0.884	0.990
	MRCML-DP	-0.066	0.057	0.069	0.912	0.996
	MRCML-BIC	-0.069	0.065	0.044	0.638	1.000
	MRBEE	-0.027	0.062	0.049	0.834	0.978

Table S22. 100% sample overlap, theta3=-0.2, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.100	0.025	0.065	0.850	1.000
	MR-Egger	-0.109	0.028	0.072	0.882	1.000
	MR-Median	-0.099	0.032	0.081	0.954	1.000
	MR-Lasso	-0.100	0.025	0.065	0.850	1.000
	MRCML-DP	0.007	0.029	0.058	0.998	0.990
	MRCML-BIC	0.001	0.029	0.049	0.998	0.996
	MRBEE	0.006	0.033	0.032	0.948	1.000
30% unbalanced UHP	IVW	0.033	0.093	0.165	0.998	0.058
	MR-Egger	0.010	0.108	0.183	0.998	0.072
	MR-Median	-0.064	0.049	0.101	0.996	0.894
	MR-Lasso	-0.071	0.048	0.073	0.926	0.996
	MRCML-DP	0.309	0.087	0.112	0.096	0.072
	MRCML-BIC	0.292	0.125	0.076	0.062	0.244
	MRBEE	0.063	0.079	0.057	0.770	0.658
30% CHP	IVW	-0.303	0.079	0.213	0.938	0.852
	MR-Egger	-0.350	0.088	0.218	0.828	0.908
	MR-Median	-0.105	0.035	0.091	0.970	1.000
	MR-Lasso	-0.087	0.031	0.068	0.918	1.000
	MRCML-DP	0.206	0.054	0.087	0.234	0.002
	MRCML-BIC	0.170	0.070	0.065	0.242	0.114
	MRBEE	0.009	0.037	0.037	0.944	0.996

Table S23. 0% sample overlap, theta4=0.4, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	-0.046	0.068	0.069	0.892	0.998
	MR-Egger	-0.045	0.068	0.070	0.902	0.996
	MR-Median	-0.045	0.084	0.090	0.946	0.986
	MR-Lasso	-0.046	0.072	0.066	0.860	0.998
	MRCML-DP	-0.006	0.078	0.094	0.978	0.996
	MRCML-BIC	-0.004	0.078	0.073	0.940	0.998
	MRBEE	-0.003	0.080	0.078	0.936	1.000
30% unbalanced UHP	IVW	0.011	0.185	0.176	0.924	0.638
	MR-Egger	0.002	0.188	0.178	0.932	0.620
	MR-Median	-0.018	0.125	0.111	0.906	0.898
	MR-Lasso	-0.025	0.116	0.077	0.788	0.972
	MRCML-DP	0.053	0.142	0.152	0.950	0.858
	MRCML-BIC	0.062	0.160	0.085	0.668	0.974
	MRBEE	0.008	0.141	0.107	0.862	0.912
30% CHP	IVW	0.029	0.108	0.222	1.000	0.476
	MR-Egger	-0.033	0.112	0.219	1.000	0.278
	MR-Median	-0.020	0.092	0.100	0.956	0.968
	MR-Lasso	-0.027	0.083	0.070	0.874	0.996
	MRCML-DP	0.032	0.092	0.108	0.972	0.984
	MRCML-BIC	0.035	0.096	0.079	0.874	1.000
	MRBEE	0.012	0.091	0.083	0.936	0.998

Table S24. 100% sample overlap, theta4=0.4, number of IVs = 200

Scenario	Method	Bias	SD	SE	CovFreq	RJF
no pleiotropy	IVW	0.036	0.043	0.065	0.986	1.000
	MR-Egger	0.032	0.043	0.072	0.994	1.000
	MR-Median	0.037	0.052	0.081	0.994	1.000
	MR-Lasso	0.036	0.043	0.065	0.986	1.000
	MRCML-DP	-0.010	0.050	0.058	0.972	1.000
	MRCML-BIC	0.000	0.051	0.049	0.948	1.000
	MRBEE	-0.003	0.055	0.053	0.942	1.000
30% unbalanced UHP	IVW	0.102	0.171	0.165	0.882	0.852
	MR-Egger	0.090	0.176	0.183	0.920	0.774
	MR-Median	0.058	0.086	0.101	0.950	1.000
	MR-Lasso	0.054	0.087	0.073	0.826	1.000
	MRCML-DP	-0.177	0.107	0.112	0.650	0.538
	MRCML-BIC	-0.139	0.133	0.076	0.520	0.806
	MRBEE	0.012	0.125	0.096	0.876	0.952
30% CHP	IVW	0.108	0.095	0.213	0.998	0.812
	MR-Egger	0.034	0.100	0.218	1.000	0.532
	MR-Median	0.037	0.061	0.091	0.990	1.000
	MR-Lasso	0.030	0.051	0.068	0.970	1.000
	MRCML-DP	-0.129	0.074	0.087	0.718	0.886
	MRCML-BIC	-0.080	0.082	0.065	0.722	0.980
	MRBEE	-0.005	0.061	0.060	0.944	1.000

1.6 Replication of Lin et al

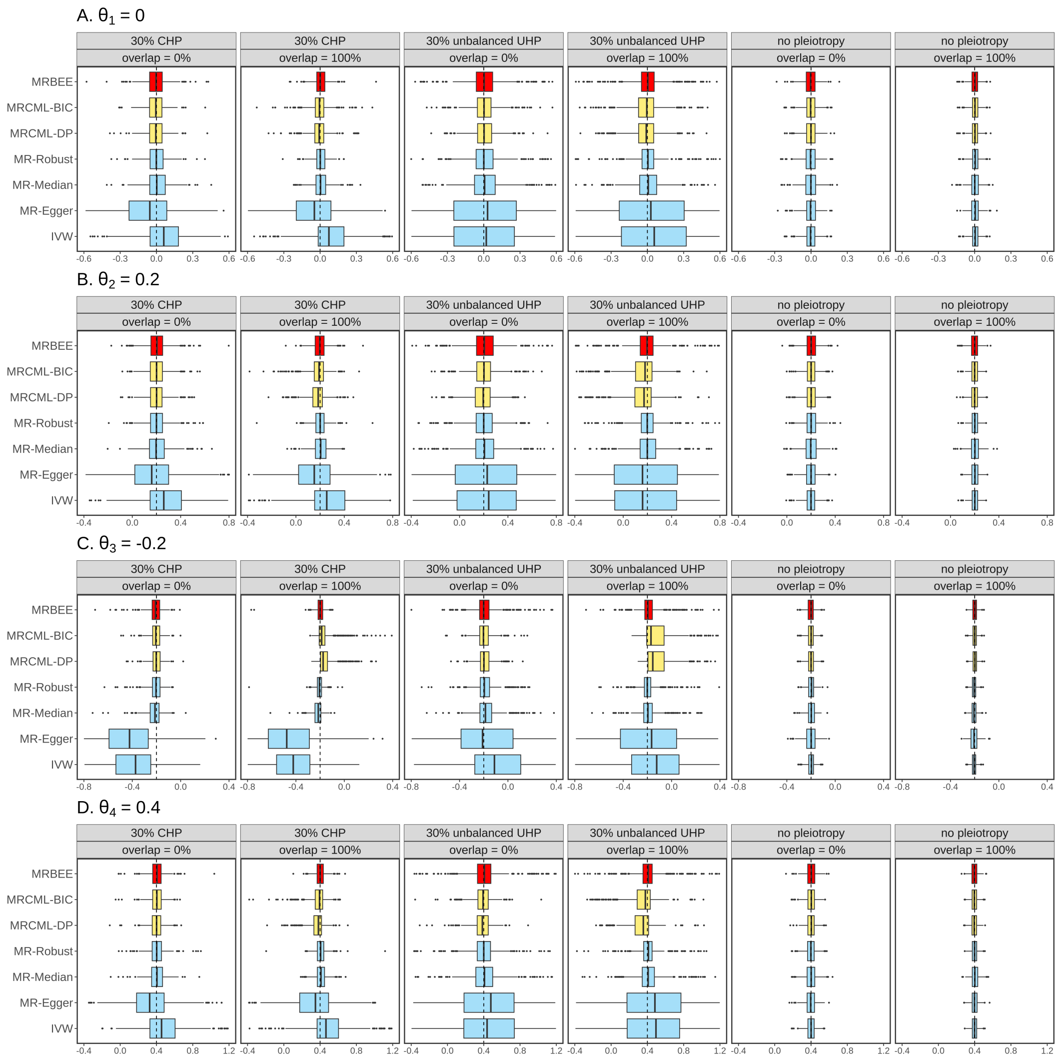


Figure S4: Estimation results of Lin et al. Panel A - L displays the boxplots of causal effect estimates from seven methods in the MVMR simulation. The four rows represent the four causal effects θ_j, $j = 1, 2, 3, 4$. Each column corresponds to one of the three scenarios. The x-axis indicates the value of the causal effect estimate, while the y-axis lists the seven methods. The true values of causal effects are denoted by dashed lines.

1.7 Replication of Wu et al

Table S25. Replication of the Table 1 in Wu et al. MRBEE represents the estimates with effect size standardization. MRBEE.Understandarized represents the estimates without effect size standardization. (see highlighted in yellow, the results were based on 500 replications).

		beta01=-0.5				beta02=-0.7				beta03=-0.3			
lambda/sqrt p	Estimator	Est	SD	SE	CP	EST	SD	SE	CP	EST	SD	SE	CP
35.6	IVW	-0.4011	0.0257	0.0265	0.0460	-0.6529	0.0200	0.0216	0.4020	0.3821	0.0274	0.0295	0.2200
	Egger	-0.4266	0.0353	0.0370	0.4920	-0.6526	0.0200	0.0208	0.3560	0.3847	0.0278	0.0305	0.2140
	Median	-0.4049	0.0350	0.0418	0.3600	-0.6474	0.0277	0.0358	0.7560	0.3797	0.0393	0.0446	0.6060
	GRAPPLE	-0.4957	0.0337	0.0346	0.9520	-0.6974	0.0243	0.0253	0.9600	0.3042	0.0367	0.0391	0.9680
	MRBEE	-0.5074	0.0393	0.0408	0.9740	-0.7024	0.0263	0.0272	0.9520	0.2941	0.0411	0.0447	0.9680
	MRBEE Unstandarized	-0.5581	0.0620	7.6143	1.0000	-0.7126	0.0424	3.0688	1.0000	0.2544	0.0629	7.0391	1.0000
	dIVW	-0.5049	0.0391	0.0390	0.9700	-0.7018	0.0261	0.0271	0.9640	0.2963	0.0411	0.0426	0.9720
	adIVW	-0.5049	0.0391	0.0390	0.9700	-0.7018	0.0261	0.0271	0.9640	0.2963	0.0411	0.0426	0.9720
11	IVW	-0.2890	0.0350	0.0359	0.0000	-0.5850	0.0290	0.0317	0.0520	0.4528	0.0401	0.0400	0.0380
	Egger	-0.3230	0.0529	0.0558	0.1200	-0.5861	0.0294	0.0326	0.0600	0.4544	0.0404	0.0442	0.0480
	Median	-0.2933	0.0504	0.0570	0.0440	-0.5733	0.0423	0.0569	0.3580	0.4474	0.0575	0.0596	0.3300
	GRAPPLE	-0.4828	0.0708	0.0694	0.9180	-0.6936	0.0437	0.0464	0.9560	0.3173	0.0793	0.0793	0.9480
	MRBEE	-0.5217	0.0996	0.1096	0.9560	-0.7090	0.0535	0.0597	0.9640	0.2829	0.1053	0.1161	0.9720
	MRBEE Unstandarized	-0.9042	2.6129	34.9001	1.0000	-0.8177	0.7338	10.6841	1.0000	-0.0506	2.5668	32.0195	1.0000
	dIVW	-0.5131	0.0973	0.0974	0.9320	-0.7055	0.0521	0.0560	0.9540	0.2903	0.1023	0.1036	0.9520
	adIVW	-0.5016	0.0798	0.0900	0.9320	-0.7012	0.0483	0.0545	0.9540	0.2974	0.0869	0.0949	0.9520
7.4	IVW	-0.2507	0.0396	0.0389	0.0000	-0.5502	0.0334	0.0359	0.0220	0.4571	0.0450	0.0436	0.0700
	Egger	-0.2806	0.0590	0.0623	0.0660	-0.5510	0.0334	0.0374	0.0300	0.4582	0.0451	0.0489	0.0840
	Median	-0.2616	0.0566	0.0597	0.0200	-0.5268	0.0515	0.0615	0.1640	0.4463	0.0645	0.0631	0.3640
	GRAPPLE	-0.4808	0.0927	0.0902	0.9380	-0.6889	0.0571	0.0582	0.9360	0.3128	0.1049	0.1036	0.9460
	MRBEE	-0.5566	0.1843	0.2033	0.9780	-0.7208	0.0837	0.0945	0.9600	0.2420	0.1965	0.2132	0.9760
	MRBEE Unstandarized	2.1489	80.3074	5037.7991	1.0000	0.2806	24.8002	1530.8156	1.0000	2.2589	64.7862	4079.5503	1.0000
	dIVW	-0.5442	0.1782	0.1613	0.9640	-0.7167	0.0810	0.0832	0.9580	0.2536	0.1899	0.1702	0.9520
	adIVW	-0.4895	0.0899	0.1136	0.9600	-0.6940	0.0586	0.0707	0.9560	0.2875	0.1075	0.1164	0.9440

1.8 Bias-correction terms: Correlation matrix estimation from insignificant GWAS statistics

We investigate the estimation error of $\hat{\mathbf{R}}_{W_\beta \times w_\alpha}$, i.e., the correlation version of covariance matrix $\hat{\mathbf{\Sigma}}_{W_\beta \times w_\alpha}$. We first examine if increasing M results in a decreasing estimation error. Besides, we consider studying the Frobenius norm rather than the ℓ_2 norm, as $||\mathbf{A}||_2 \leq ||\mathbf{A}||_\mathrm{F}$ and the calculation of the Frobenius norm is much less costly than the ℓ_2 norm. In comparison, we also consider the correlation matrix estimate directly yielded by the individual data, whose convergence rate is roughly $O(\min(\sqrt{n_1}, \sqrt{n_0}))$. The number of replications is 1000.

For this purpose, we set $M = 250, 500, \ldots, 2000$, $n_1 = n_0 = 2000, 20000$, and $n_o/n_0 = 0.5$. Figure S5 shows the investigation, from which we witness: (1), as M increases, the Frobenius norm of $\hat{\mathbf{R}}_{W_\beta \times w_\alpha}$ is reduced; (2) directly estimating $\hat{\mathbf{R}}_{W_\beta \times w_\alpha}$ from the individual data is always more precise than indirectly estimating it from insignificant GWAS statistics. In addition, although the estimation error of $\hat{\mathbf{R}}_{W_\beta \times w_\alpha}$ only depends on M, low sample sizes will introduce finite-sample bias into the estimation.

We then study if increasing n_1 and n_0 will influence the estimation error of $\hat{\mathbf{R}}_{W_\beta \times w_\alpha}$. For this purpose, we set $M = 250, 500, 1000$ and let n_1 and n_0 increase from 5000 to 40000 with a lag 5000. The number of replications is 1000. Figure S6 exhibits the results, from which we observe: increasing n_1 and n_0 cannot reduce estimation error of $\hat{\mathbf{R}}_{W_\beta \times w_\alpha}$. These results confirm our theory: the estimation error of $\hat{\mathbf{R}}_{W_\beta \times w_\alpha}$ only depends on M.

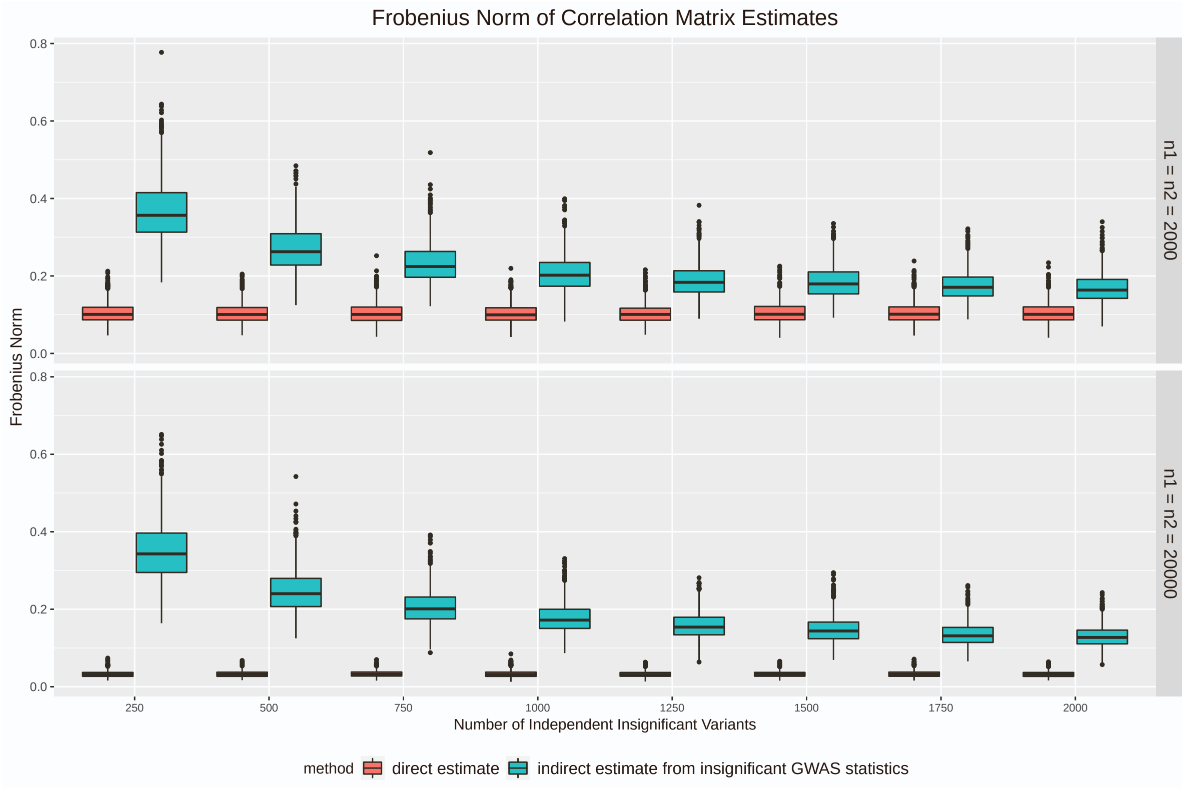


Figure S5: The Frobenius norms of correlation matrix esimates when M increases.

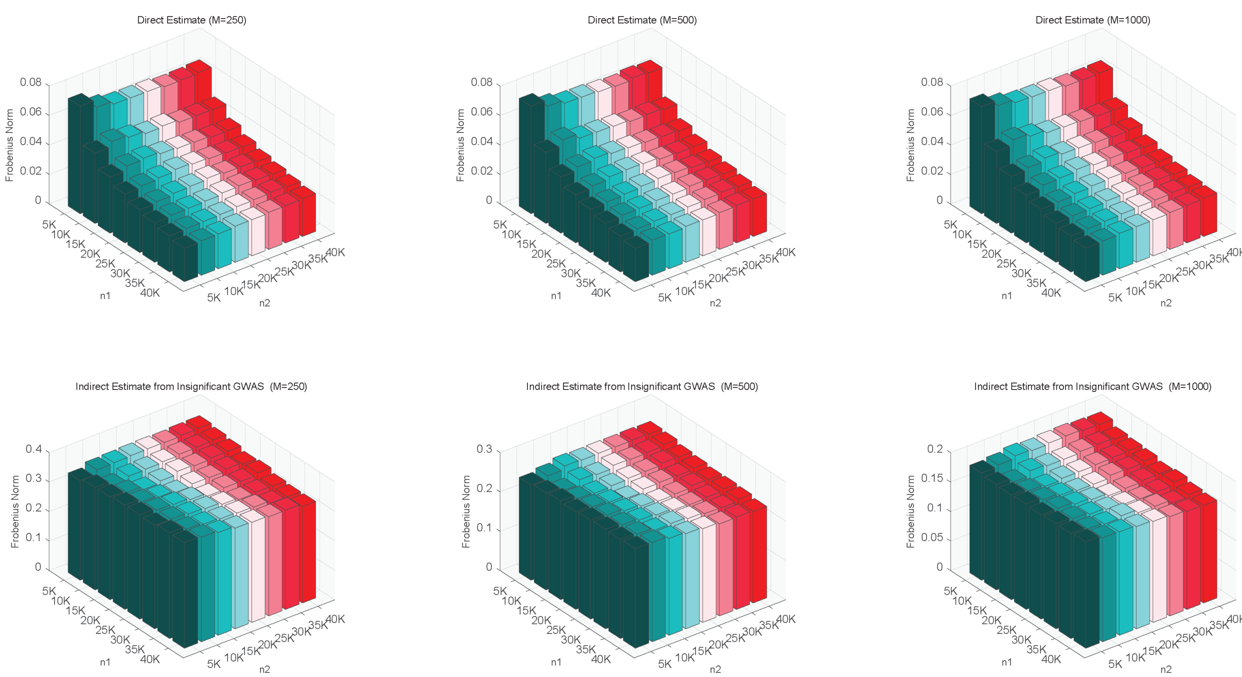


Figure S6: The Frobenius norms of correlation matrix estimates when n_0 and n_1 increase.

2 Supplemental Univariable Simulations

2.1 Overlapping Fraction

We briefly introduce the simulation settings for UVMR. First, we generate a binomial variable from $\mathrm{Binom}(2, b_j)$ where $b_j \sim \mathrm{Unif}(0.05, 0.5)$ and standardize it as g_{ij}, the direct effect β_j from $\mathcal{N}(0, 1/m)$, and u_i, v_i from a normal distribution with correlation coefficient 0.5. The variances of u_i and v_i are chosen such that the IV-heritabilities are $\sigma_{\beta\beta}/\sigma_{xx} = 0.3$ and $\theta^2 \times (\sigma_{\beta\beta}/\sigma_{yy}) = 0.15$, respectively. We specify the causal effect $\theta = 0.3/\sqrt{2}$. We compare MRBEE with IVW, DIVW, MR-RAPS, MR-Egger, MR-Lasso, MR-Median, IMRP, MR-Conmix, and MR-MiX, where most are implemented by using the R package `MendelianRandomization`. We fix $n_0 = n_1 = 20000$, specify n_{01} according to the overlapping fraction, and assume no UHP or CHP. The so-called overlapping fraction is n_{01}/n_0, where the special fraction such that $\mathrm{E}(S_{\mathrm{IVW}}(\theta)) = 0$ is $n_{01}/n_0 \approx 0.77$. The number of independent replications is 1000.

First, we study the influences of overlapping fraction n_{01}/n_0 and the number of IVs m, with the results displayed in Fig.S7. It is easy to see that only MRBEE is able to yield an unbiased estimate of θ in all cases. For a special overlapping fraction $n_{01}/n_0 \approx 0.77$, all approaches become unbiased except MR-RAPS and DIVW. These two methods perform badly because are based on no sample overlap assumption, which in turn add extra biases to the estimates as long as sample overlap exists. The SEs of causal effect estimates for all methods increases as the overlapping fraction decreases but remains unchanged by the increase of m, confirming that the convergence rates of causal estimates are mainly determined by $n_{\min}$.

As for the SE estimation, we display the boxplot of $\hat{\mathrm{se}}(\hat{\theta}) - \mathrm{se}(\hat{\theta})$ where $\mathrm{se}(\hat{\theta})$ is approximated by the empirical SE calculated from the independent replications. It is evident that the SE estimates produced by all approaches have reduced variances as m grows. However, only MRBEE and DIVW can provide consistent SE estimates, confirming the accuracy of their SE formulas. MR-ConMix is extremely likely to underestimate the standard error, while MR-Egger, MR-Lasso, MR-Median, and MR-Mix constantly overestimate it. IVW underestimates the SE when the fraction is large and overestimates it when the fraction is small. In contrast, MR-RAPS seems to overestimate the SE unless the overlapping fraction is 0%.

The coverage frequency refers to the frequency that the confidence interval covers the true causal effect among simulations. Here, this confidence interval is constructed by doubling $\hat{\mathrm{se}}(\hat{\theta})$, meaning that the coverage frequency corresponding to neither an inflated type-I error nor an inflated type-II error should be 0.95. We observed that only MRBEE enjoys a coverage frequency around 0.95. When $m = 250$, MR-Egger, MR-Lasso, and MR-Median suffer from inflated type-II errors, likely because these methods cannot estimate the SE properly. These approaches also result in inflated type-I errors caused by weak instrument bias as m increases. Additionally, because MR-Mix overestimates the SE, it consistently exhibits a substantially inflated type-II error. Furthermore, IMRP and MR-ConMix consistently have inflated type I errors because they frequently underestimate the SE.

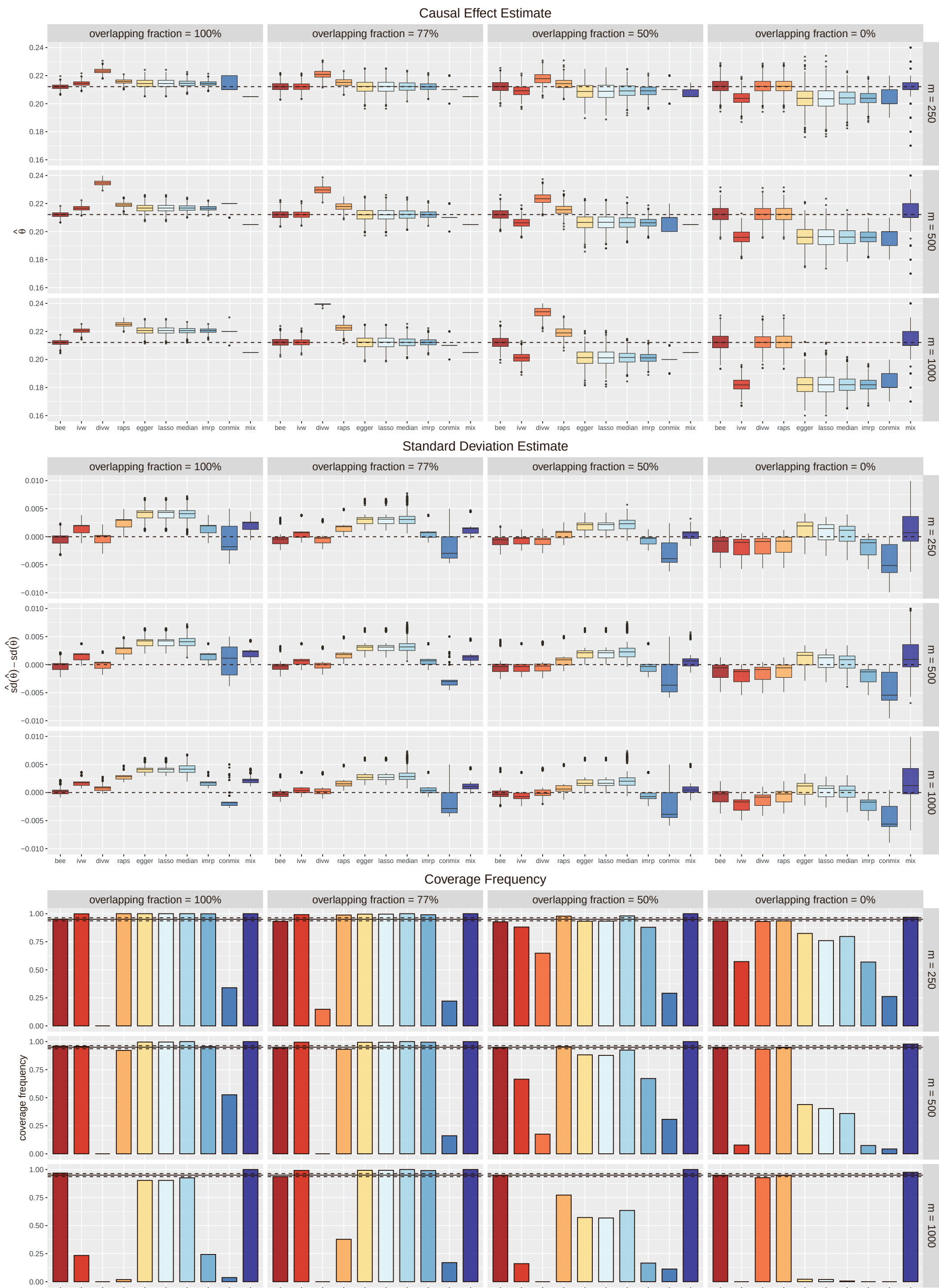


Figure S7: Investigation of UVMR approaches for univariable MR with sample sizes $n_0 = n_1 = 20000$, in terms of overlapping fraction and number of instrumental variants.

2.2 Sample size

In this section, we examine the influence of sample sizes. Here, we fix the number of variants $m = 500$ and consider $n_0 = n_1 = 20K$; $n_0 = 40K, n_1 = 20K$; $n_0 = 20K, n_1 = 40K$; and $n_0 = n_1 = 40K$ four cases. Recall that the overlapping fraction is defined as n_{01}/n_0, and 100%, 77%,50%, and 0% four cases will be studied. Other setting remains the same as the one shown in section 4.1 in the main paper.

Figure S8 displays the results of this examination. Preliminary, it illustrates neither increasing n_0 nor increasing n_1 along is able to make the causal effect estimate more accuracy. Besides, increasing the sample sizes of the exposure GWAS and the outcome GWAS has different impacts: the former decreases the measurement error bias, while the latter reduces the variance of all causal effect estimates. The reason is that the estimation error of $\hat{\alpha}_j$ will not cause estimation bias, in contrast, it is indeed the random error term of the multivariate MR model. Furthermore, only the MR-BEE is able to produce unbiased causal effect estimate and reliable SE estimate in all cases.

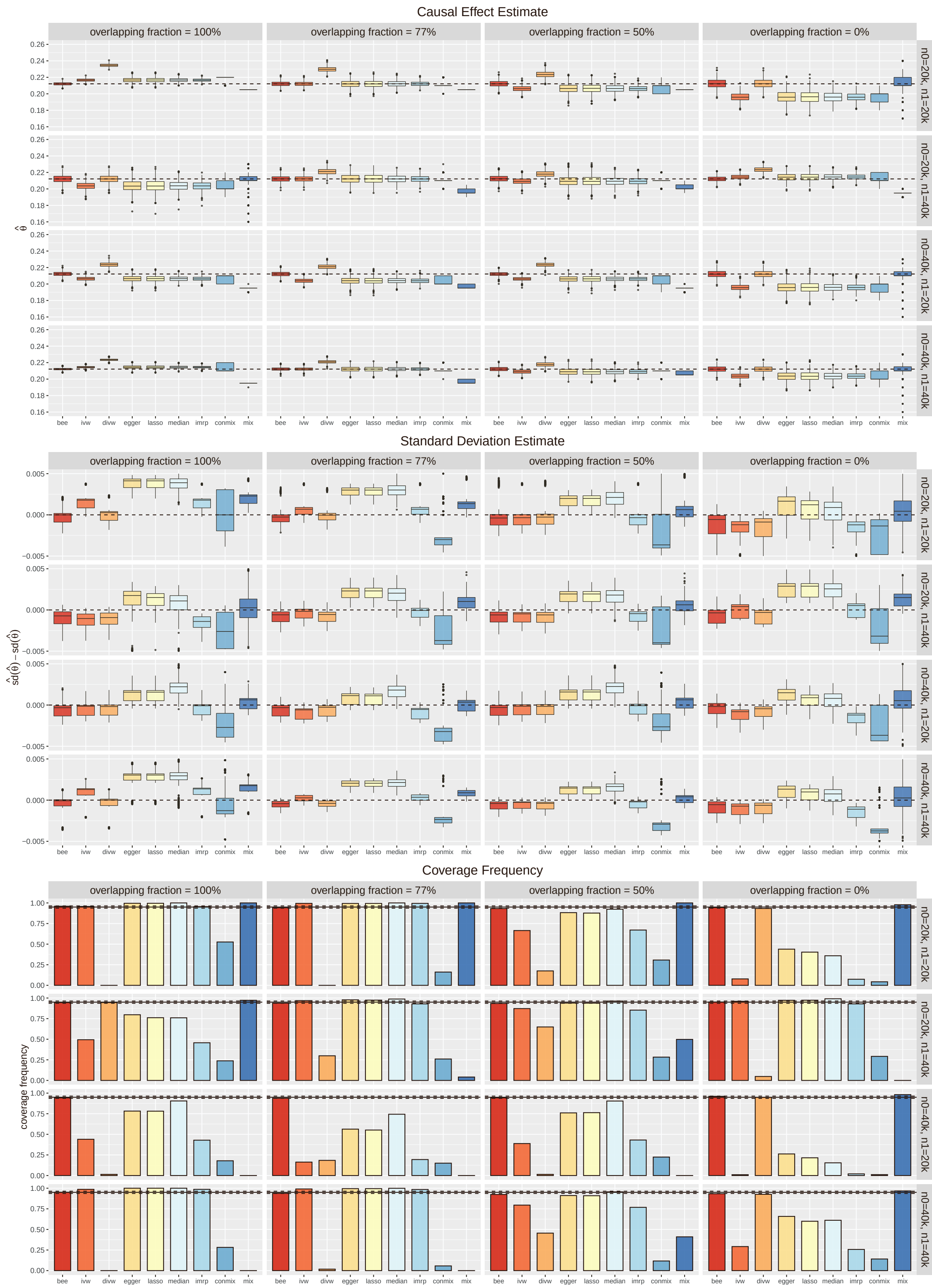


Figure S8: The investigation of univariate MR in terms of sample sizes.

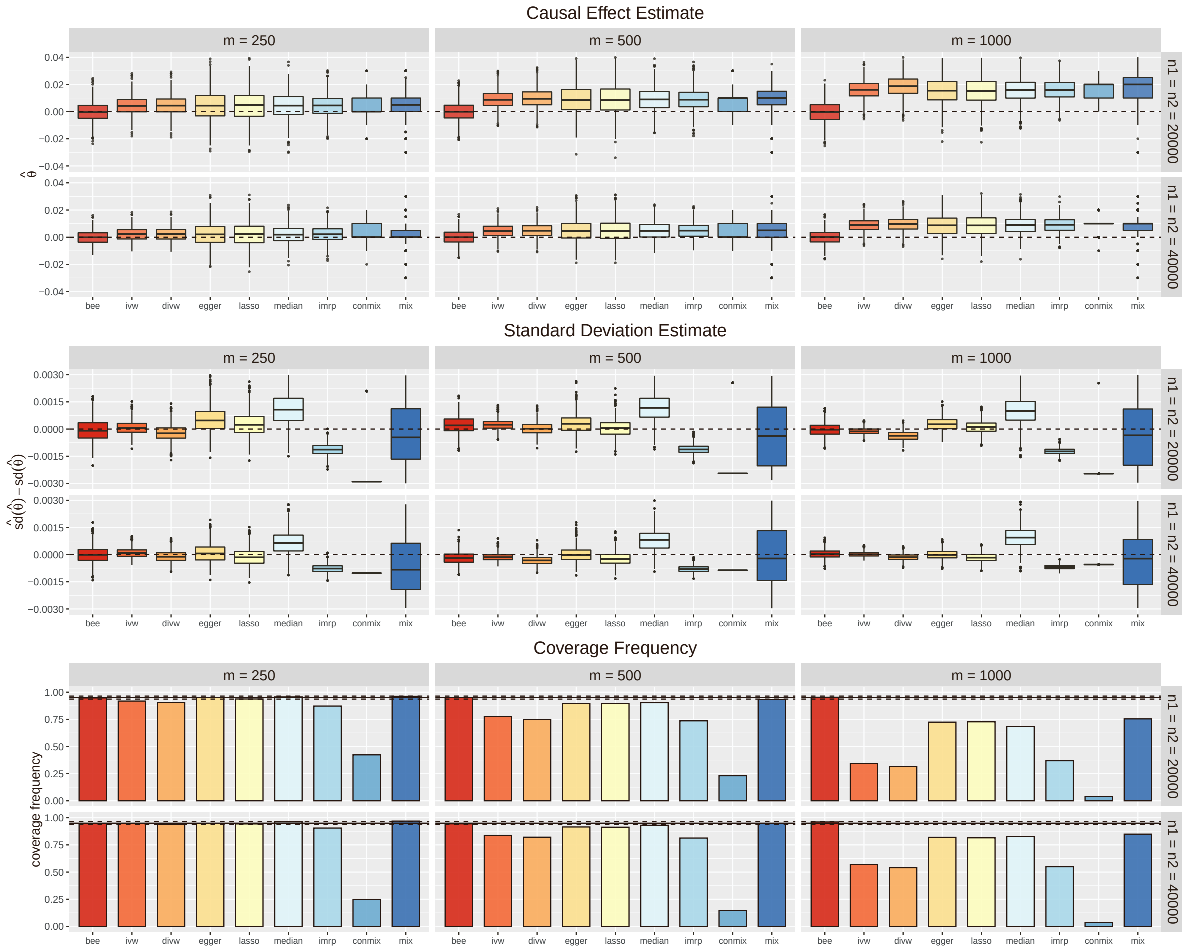


Figure S9: The investigation of univariate MR in terms of type-I error.

2.3 Type-I error

Now we turn to examine whether MRBEE and existing approaches produce inflated type-I error rates when UHP is present. Note that the UHP γ_{u_j} must exist otherwise the IV-heritability of outcome will be zero when $\theta = 0$. We independently generate γ_{u_j} from the same distribution as β_j. The simulation settings are: IV-heritability of exposure = 0.3, IV-heritability of outcome = 0.15, $n_{01}/n_0 = 0.5$, and the number of replications is 1000.

Figure S9 exhibits the results, from which some phenomena are consistently observed; e.g., increasing n_1 and n_0 simultaneously reduces the variances of all causal estimates, while increasing m increases weak instrument bias. Since $\theta = 0$ implies that only the confounder bias ($n_{01}/n_0\sigma_{uv}$) exists, all the weak instrument biases are upward. (The correlation coefficient between u_i and v_i is 0.5.) In addition, all the existing approaches incur inflated type-I errors as m rises. The result suggests that the weak instrument bias is likely to explain some significant causal relationships observed in the literature. However, using MRBEE can produce reliable causal inferences..

2.4 Winner's curse

In this section, we examine the impact winner's curse. We use the exactly same setting as the one in section 4.1. To simulate the winner's curse, we only use the variants with absolute t-statistics (i.e., $|\hat{\beta}/\text{se}(\hat{\beta})|$) larger than 1 or 2.

Figure S10 displays the results of this examination. It shows the winner's curse will not introduce a significant bias into MR-BEE as long as the overlapping fraction is not zero. As for other MR approaches that suffer from biases, we observed that the winner's curse will indeed slightly reduce the biases but inflate the variances. We believe that only selecting the significant variants will reduce the weak instrument bias somehow, because the weak instrument bias is determined by the ratio of signal-by-noise, i.e.,

$$\frac{\psi_{\beta\beta}}{m} \text{ v.s. } \sigma_{W_\beta W_\beta}, \tag{10}$$

where $\psi_{\beta\beta} = \sum_{j=1}^m \text{var}(\beta_j)$. If $\psi_{\beta\beta}/m$ is significantly larger than $\sigma_{W_\beta W_\beta}$, the bias of the IVW estimate should disappear due to the structure of "weak instrument bias x estimation error bias".

In addition, as the overlapping fraction decreases, the MR-BEE also encounters small bias especially when this fraction is zero. The reason for this problem is

$$\frac{1}{m}\sum_{j=1}^m \beta_j \omega_{\beta_j} \to 0, \quad \frac{1}{|\mathcal{W}|}\sum_{j\in\mathcal{W}} \beta_j \omega_{\beta_j} \not\to 0, \tag{11}$$

where $\mathcal{W}$ is the set of all "winners". In this case, extra selection bias arises but MR-BEE fails to account for it. Fortunately, such a bias is usually modest and it seems only existing when the overlapping fraction is 0. Increasing the sample size to identify more causal variants is one of the practical ways to resolve the winner's curse in this case.

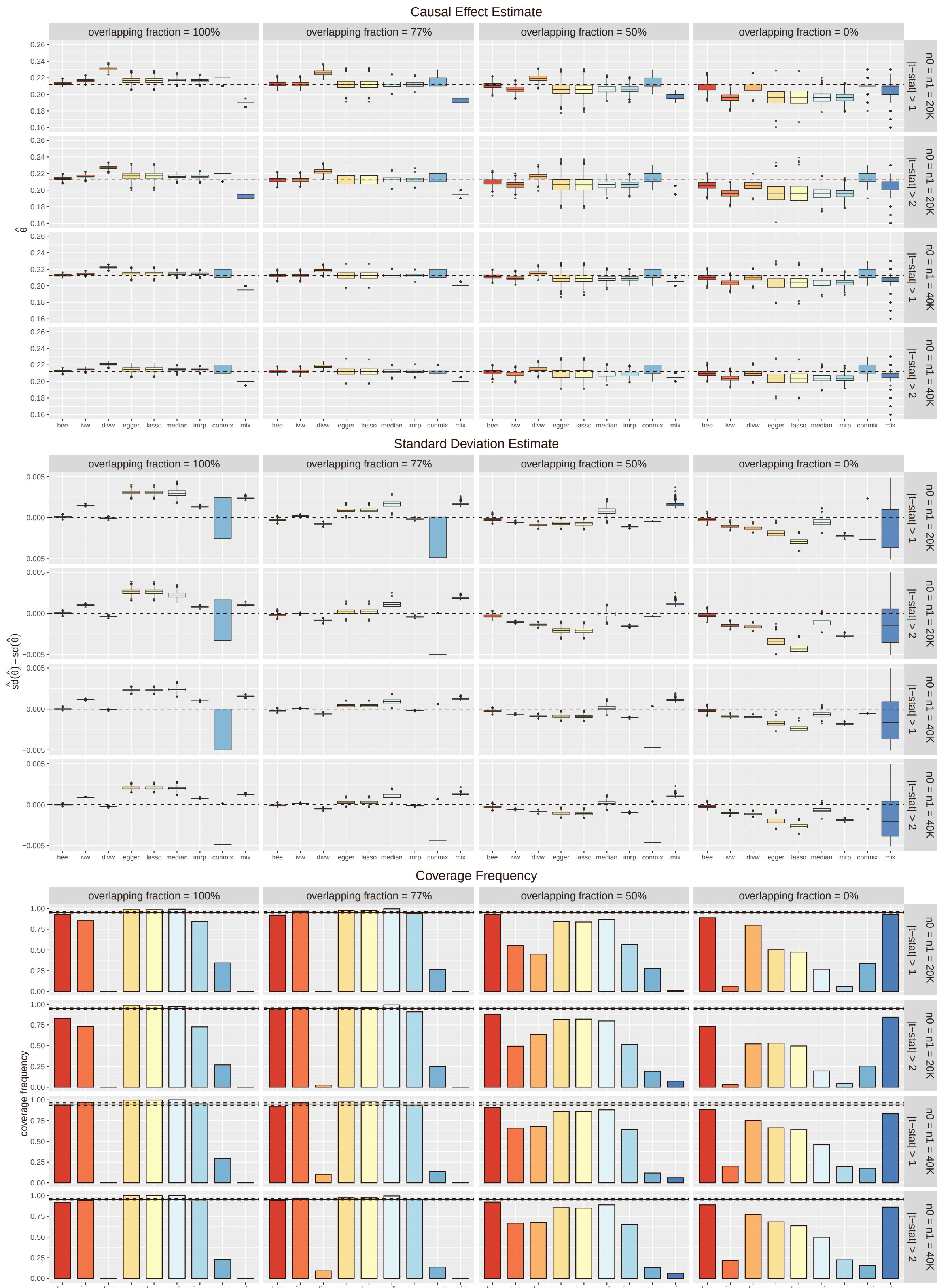


Figure S10: The investigation of univariate MR in terms of winner's curse.

2.5 Outlier test

In this section, we investigate if the MR-BEE with the IMRP pleiotropy test is able to remove the pleiotropy as resembling the outlier detection. The methods for comparsion include IMRP, MR-Lasso and MR-ConMix. Detailed setting of outliers can be found in the subsection of outlier detection setting. Here, we consider three criteria: estimation error of causal effect, true negative (TN) and true positive (TP). Here, the TN refers to the proportion of removing all outliers, while the TP refers to the proportion of not removing any valid IV. For IMRP and MR-BEE, we need to specify the threshold κ. For IMRP, we consider two thresholds: $\kappa = 0.05$ and $\kappa = 0.05/s$ where s is the number of real outliers. Regarding MR-BEE with IMRP, we not only consider this two thresholds but also consider two FDR control methods "BH" and "Sidak", where the thresholds in these two methods are 0.05. Details of the FDR control methods can be found in R package `FDRestimation`.

Figure S11 displays the results of outlier detection. As for estimation error, MR-BEE with threshold $\kappa = 0.05$ suffers from a small selection bias, because this estimator is supposed to remove many valid IVs because of false discovery. As for MR-BEE with other thresholds, they do not suffer from bias. As for other methods, they incur large bias introduced by the weak instrument bias and estimation error bias.

As for TN, the results show all methods are able to remove the true outliers. As for TP, however, only the MR-Lasso is able to keep all valid IVs. MR-BEE and IMRP with the oracle threshold (i.e., $\kappa = 0.05/s$) have large probabilities to keep every valid IV with the increasing of outlier fractions, but this probability is not 1. Other methods cannot keep valid IVs at all, although the causal effect estimates may not have biases. These results show that there exists a theoretical threshold $\kappa \asymp F_{\chi^2}(\log m)$ to distinguish the outliers and the valid IVs, but this threshold may be difficult to specify in practice. In contrast, the MR-Lasso seems to enjoy the oracle property thanks to the consistency of lasso-type regularizer (Fan, 2001).

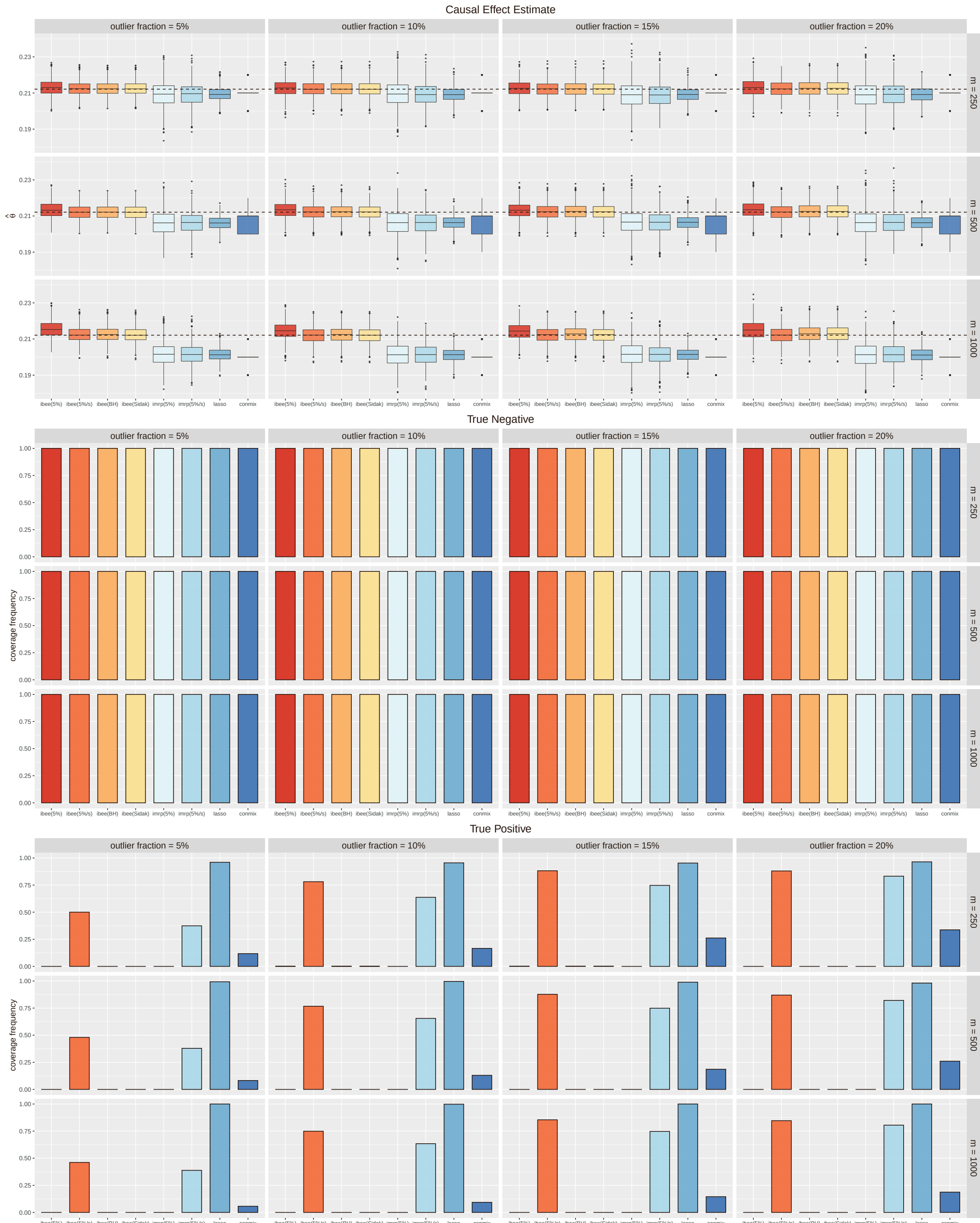


Figure S11: The investigation of univariate MR in terms of outlier detection.

2.6 Verification of Asymptotic Theroy

We next verify if the asymptotic normal distributions in Theorem 1.2 and Theorem 1.2 are correct. For a general estimate $\hat{\theta}$, the asymptotic bias and SE are $\sqrt{s}_n(\hat{\theta}-\theta)$ and $\sqrt{s}_n\text{se}(\hat{\theta})$, respectively, where $\sqrt{s}_n$ is the convergence rate of $\hat{\theta}$. If this estimate is strongly asymptotically unbiased, the asymptotic bias $s_n(\hat{\theta}-\theta)$ should also be 0. Besides, if two estimates have equal asymptotic SEs, they are equally powerful in terms of statistical efficiency. We select MR-BEE, IVW, MR-Median, and MR-Lasso to compare, only consider two overlapping fractions: 100% and 0%, set $n_0 = n_1 = n_{\min}$, and fix the causal effect $\theta = 0.5$. As for m and $n_{\min}$, we focus on the following four cases:

(1) $m = 2500, 5000, \ldots, 50000$ and $m^{0.9}/n = c_0 = 0.1$ and 0.2; we examine the direct bias: $\hat{\theta}-\theta$, asymptotic SE: $\sqrt{n_{\min}^2/m}\ \text{se}(\hat{\theta})$, and coverage frequency;

(2) $m = 250, 500, \ldots, 5000$ and $m/n = c_0 = 0.1$ and 0.2; we examine the direct bias: $\hat{\theta}-\theta$, asymptotic SE: $\sqrt{n}_{\min}\ \text{se}(\hat{\theta})$, and coverage frequency;

(3) $m = 250, 500, \ldots, 5000$ and $m^2/n = c_0 = 5$ and 10; we examine the asymptotic bias: $\sqrt{n}_{\min}(\hat{\theta}-\theta)$, asymptotic SE: $\sqrt{n}_{\min}\ \text{se}(\hat{\theta})$, and coverage frequency;

(4) $m = 250, 500, \ldots, 5000$ and $m^3/n = c_0 = 5$ and 10; we examine the asymptotic bias: $\sqrt{n}_{\min}(\hat{\theta}-\theta)$, asymptotic SE: $\sqrt{n}_{\min}\ \text{se}(\hat{\theta})$, and coverage frequency.

Note that we directly generate the estimation errors $\mathbf{W}_\beta$ and $\boldsymbol{w}_\alpha$ according to Theorem 1 because $n_{\min}$ in cases (3) and (4) can be larger than one million. %The calculations involving individual-data are extremely time-consuming in these cases.

Fig. S12 demonstrates the simulation results. In case (1), $\hat{\theta}_{\text{BEE}}$ is unbiased while the other three estimates suffer from non-removable biases. For the asymptotic SE, $\sqrt{n_{\min}^2/m}\ \text{se}(\hat{\theta}_{\text{BEE}})$ remains unchanged when $n_{\min}$ and m are sufficiently large (e.g., the bars colored in blue), verifying conclusion (*iii*) in Theorem 1.3. However, the coverage frequency of MR-BEE is a little larger than 0.95, meaning that the SE of $\hat{\theta}_{\text{BEE}}$ is overestimated in this extreme case. This phenomenon is reasonable because Theorem 1.4 points out that the convergence rate of the sandwich formula is $\min(\sqrt{n}_{\min}, n_{\min}/\sqrt{m}, \sqrt{m/\log m})$, which slows down as m increases. In case (2), the direct bias of $\hat{\theta}_{\text{IVW}}$ is unchanged as $n_{\min}$ tends to infinity, confirming conclusion (*iii*) in Theorem 1.2. As for $\hat{\theta}_{\text{BEE}}$, its asymptotic SE is a little larger than $\hat{\theta}_{\text{IVW}}$, verifying item (*ii*) in Theorem 1.3.

In case (3), the asymptotic bias of $\hat{\theta}_{\text{IVW}}$ is constant as $n_{\min}$ goes to infinity, illustrating that $\hat{\theta}_{\text{IVW}}$ is not strongly asymptotically unbiased. As a result, the coverage frequencies of $\hat{\theta}_{\text{IVW}}$ are significantly smaller than 0.95, confirming our claim that any inference made based on $\hat{\theta}_{\text{IVW}}$ is invalid. Besides, the asymptotic SEs of $\hat{\theta}_{\text{BEE}}$ and $\hat{\theta}_{\text{IVW}}$ are essentially the same, indicating that $\hat{\theta}_{\text{BEE}}$ and $\hat{\theta}_{\text{IVW}}$ are equally efficient as long as $m/n_{\min} \to 0$. In case (4), the asymptotic bias of IVW, MR-Median, and MR-Lasso vanish as $n_{\min}$ increases and their coverage frequencies are around 0.95, which is consistent with conclusion (*i*) in Theorem 1.2. The equal asymptotic SEs also indicate that $\hat{\theta}_{\text{BEE}}$ and $\hat{\theta}_{\text{IVW}}$ are equally efficient in this scenario. In addition, IVW, MR-Median, and MR-Lasso suffer from the same degree of bias when there is no pleiotropy, while MR-Median not only suffers from a large asymptotic SE but also is likely to overestimate it.

Figure S12: Investigations of MR-BEE and IVW in terms of asymptotic bias and covariance matrix.

2.7 Larger numbers of IVs

Here, we directly generate the GWAS summary data from the normal distribution using the following model:

$$\hat{\boldsymbol{\beta}}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\beta\beta} + \boldsymbol{\Sigma}_{W_\beta W_\beta}), \quad \hat{\alpha}_j \sim \mathcal{N}(\boldsymbol{\beta}_j^\top \boldsymbol{\theta}, \boldsymbol{\theta}^\top \boldsymbol{\Sigma}_{W_\beta W_\beta} \boldsymbol{\theta} + \sigma_{w_\alpha w_\alpha} + 2\boldsymbol{\theta}^\top \boldsymbol{\sigma}_{W_\beta w_\alpha}.)$$

This helps us to evaluate the performances of the existing methods in the cases of larger numbers of IVs.

For larger numbers of IVs, the degrees of the weak instruments are higher, MRBEE and MR.CUE are two methods consistently performing well in the no pleiotropy cases. This confirms our conjecture that the key to removing weak instrument bias is accounting for the covariance matrix of estimation errors—however, MR.CUE suffers from bias in the presence of pleiotropy. We believe this is due to the fact that MR.CUE only considers the UHP satisfying the InSide condition, which cannot address the unbalanced UHP. In addition, the univariable version of MRCML is generalized bias because it does not require the user to provide the correlation between exposure and outcome GWAS, which implies it does not account for the correlation between exposure and outcome GWAS estimation errors. In contrast, the multivariable version of MRCML requires us to provide it, and hence it is unbiased.

2.8 Additional pleiotropy simulation

We performed a univariable MR simulation to compare the performance of horizontal pleiotropy identification methods used by MRBEE and MRCML-BIC and their subsequent effects on their causal estimates. The simulation models and R code used to generate the simulated data are presented in Figure 15. In these simulations, we fixed the number of causal exposure SNPs at 100, the exposure heritability at 0.15, the true causal effect at 0.2, and the exposure and outcome GWAS sample sizes at 30k and non-overlapping and varied the mean of UHP from a value of 0 to a value of 0.1. For each UHP mean, we drew UHP effects for each SNP from a normal distribution with variance that was one fourth of the variance of the true SNP-outcome associations. We then estimated causal effects using MRBEE and MRCML-BIC. We then recorded the number of horizontally pleiotropic IVs that were identified by each method and the corresponding causal effect estimates after excluding them. These results indicate that the results of which isare presented below, which suggestreveals that MRBEE correctlyonsistently unbiasedly estimateds the causal effects and identified a stableconstant proportion of UHP IVs regardless of the UHP mean, whereas the BIC method of MVMR-cML MRCML-BIC identifieds UHP IVs at different rates as the UHP mean changeds, thus affecting the its subsequent causal effect estimate. In this simulation, the causal estimate was based on observed values of $\hat{\beta}_X$ and $\hat{\beta}_Y$, the observed SNP-exposure and SNP-outcome associations, respectively, and both methods were adjusted for GWAS estimation error.

3 Real Data Analysis

3.1 Myopia data: heritability, genetic correlation matrix, and estimation error correlation matrix

3.2 SCZ data: heritability, genetic correlation matrix, and estimation error correlation matrix

3.3 CAD data: heritability, genetic correlation matrix, and estimation error correlation matrix

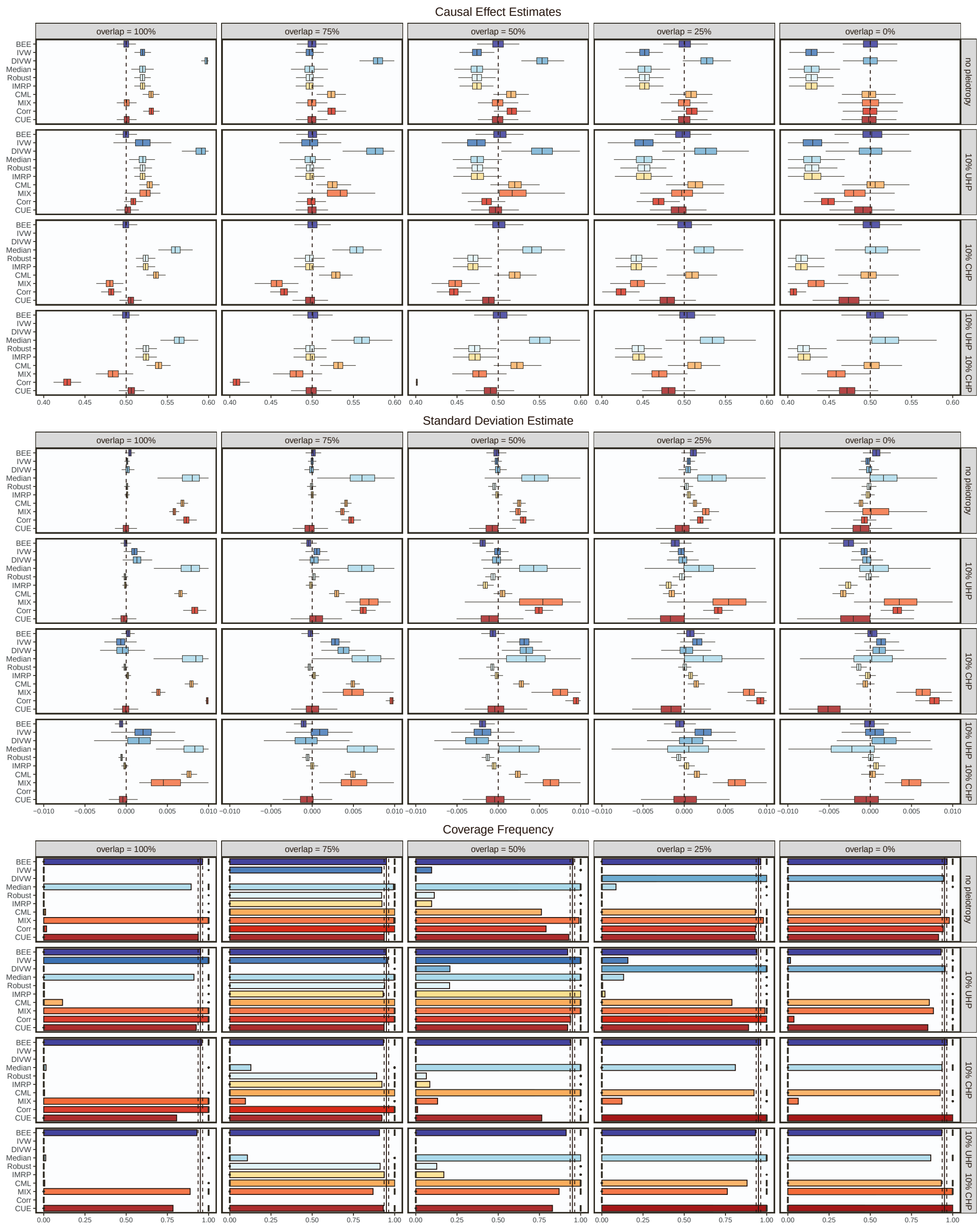


Figure S13: Investigation of UVMR approaches for UVMR model with sample sizes $n_0 = \cdots = n_6 = 20000$ and number of IVs $m = 1000$.

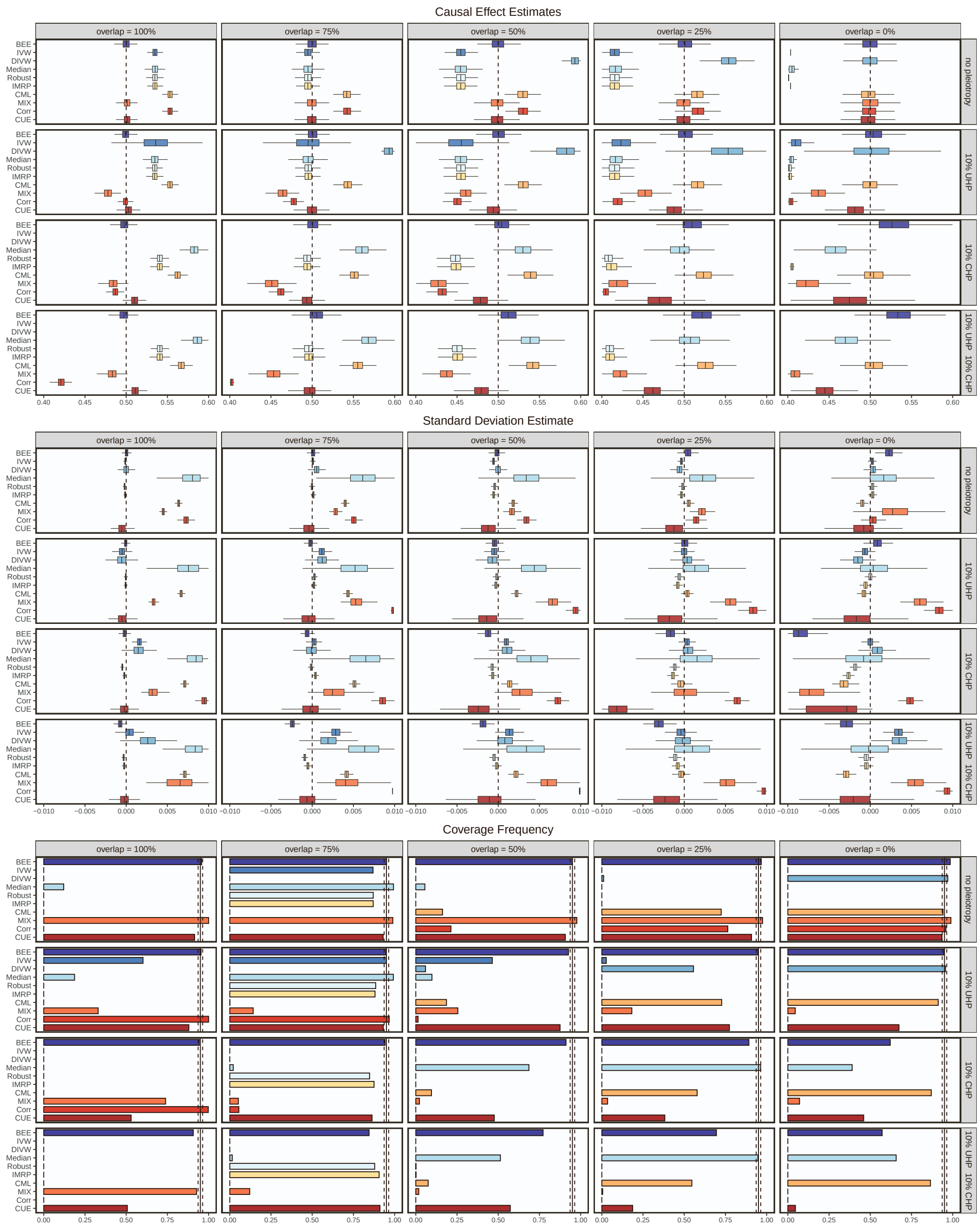
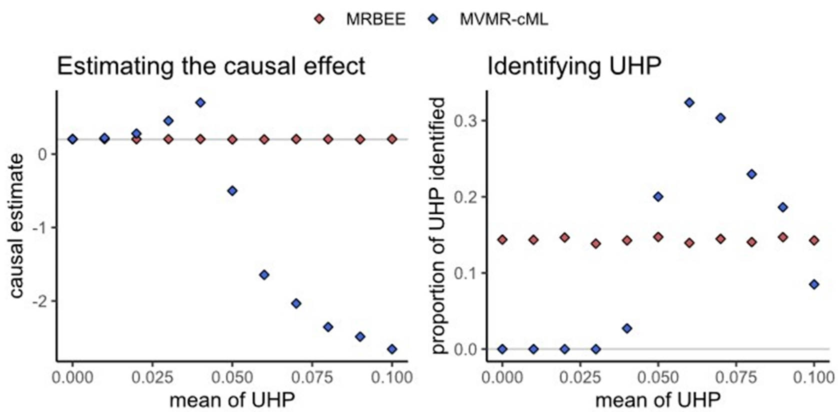


Figure S14: Investigation of UVMR approaches for UVMR model with sample sizes $n_0 = \cdots = n_6 = 20000$ and number of IVs $m = 2000$.



$$\beta_X \sim N\left(0, \frac{0.15}{100}\right)$$

$$\beta_Y = \beta_X \times 0.2 + \gamma$$

$$\gamma \sim N\left(\eta, \frac{0.15/100}{2} 0.2^2\right)$$

$$\eta \in \{0, 0.01, 0.02, \ldots, 0.10\}$$

$$\hat{\beta}_X = \beta_X + w_X$$

$$\hat{\beta}_Y = \beta_Y + w_Y$$

$$w_X, w_Y \overset{iid}{\sim} N\left(0, \frac{1}{30{,}000}\right)$$

```
m=100 # number of IVs
nx=ny=30000;h20.15;theta=0.2
niter=100
pleio_mean=seq(0,0.1,0.01)
byse=bxse=rep(1/sqrt(ny),m)
N=diag(sqrt(1/c(ny,nx)))
Sig_inv_l=lapply(1:m,function(h) N^2)
dims=c(niter,length(pleio_mean),2)
EST=NPLEIO=array(dim=dims)
for(p in 1:length(pleio_mean)) {
 for(iter in 1:niter) {
  bx=rnorm(m,0,sqrt(h2x/m))
  by=bx*theta
  adj=sqrt(var(by)/2)
  gamma=rnorm(m,pleio_mean[p],adj)
  by=by+gamma
  bxhat=bx+rnorm(m,0,sqrt(1/nx))
  byhat=by+rnorm(m,0,sqrt(1/ny))
  # estimate causal effect below
 }
}
```

Figure S15: These are the results of simulations described above comparing the performance of MRBEE and MRCML-BIC in identifying horizontal pleitropy and estimating the causal effect as the UHP mean changes.

References

Lin, Z., Xue, H. and Pan, W. (2023). Robust multivariable Mendelian randomization based on constrained maximum likelihood. The American Journal of Human Genetics, **110**(4), pp.592-605.

Fan. J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, **96**(456), pp.1348-1360.

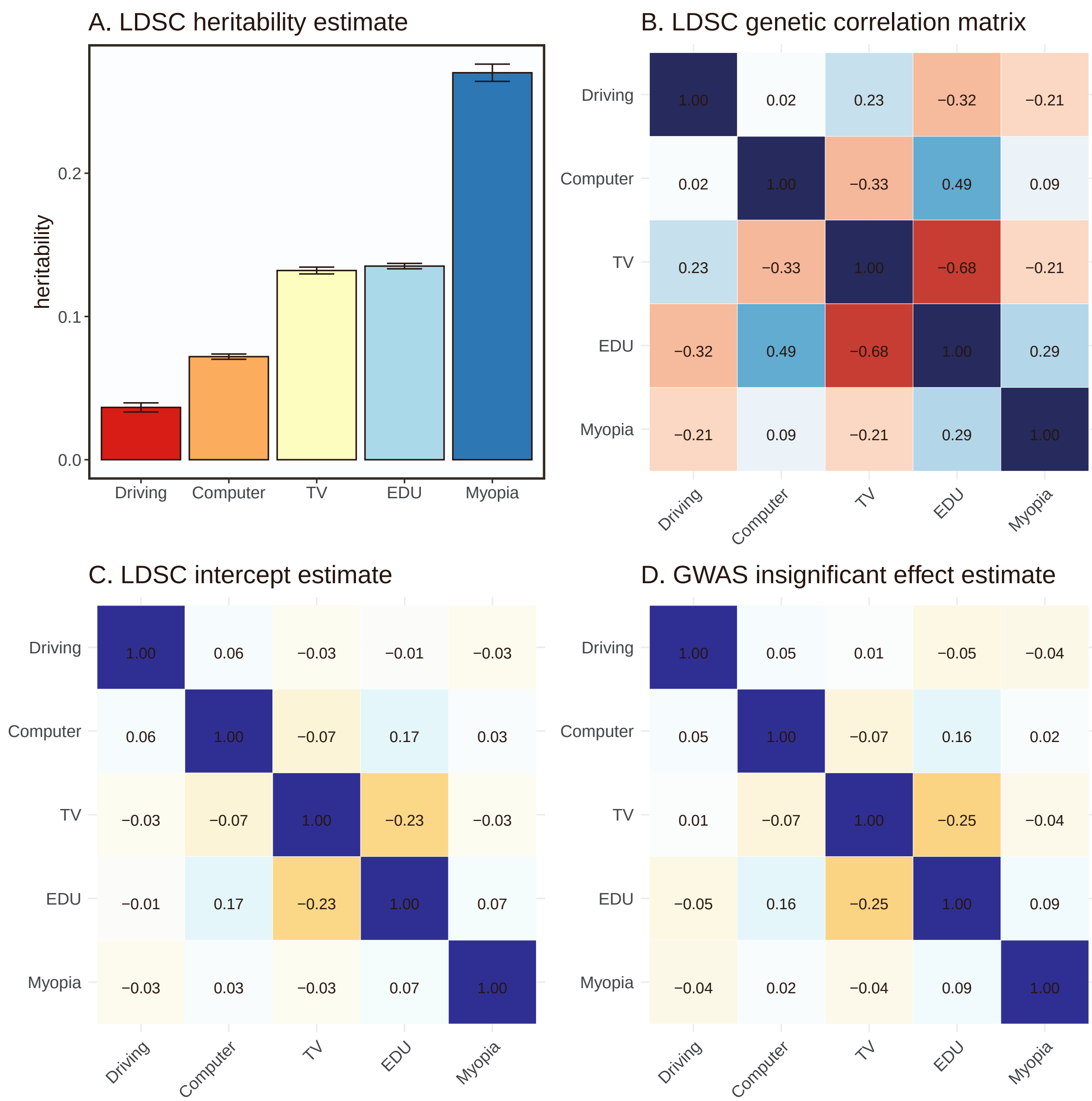


Figure S16: Myopia data. A. Heritability estimated by LDSC and the corresponding confidence intervals (radius is double SE). B. Genetic correlation matrix estimated by LDSC. C. Correlation matrix of estimation error constructed using the intercept from LDSC estimation. D. Correlation matrix of estimation error constructed using GWAS insignificant statistics.

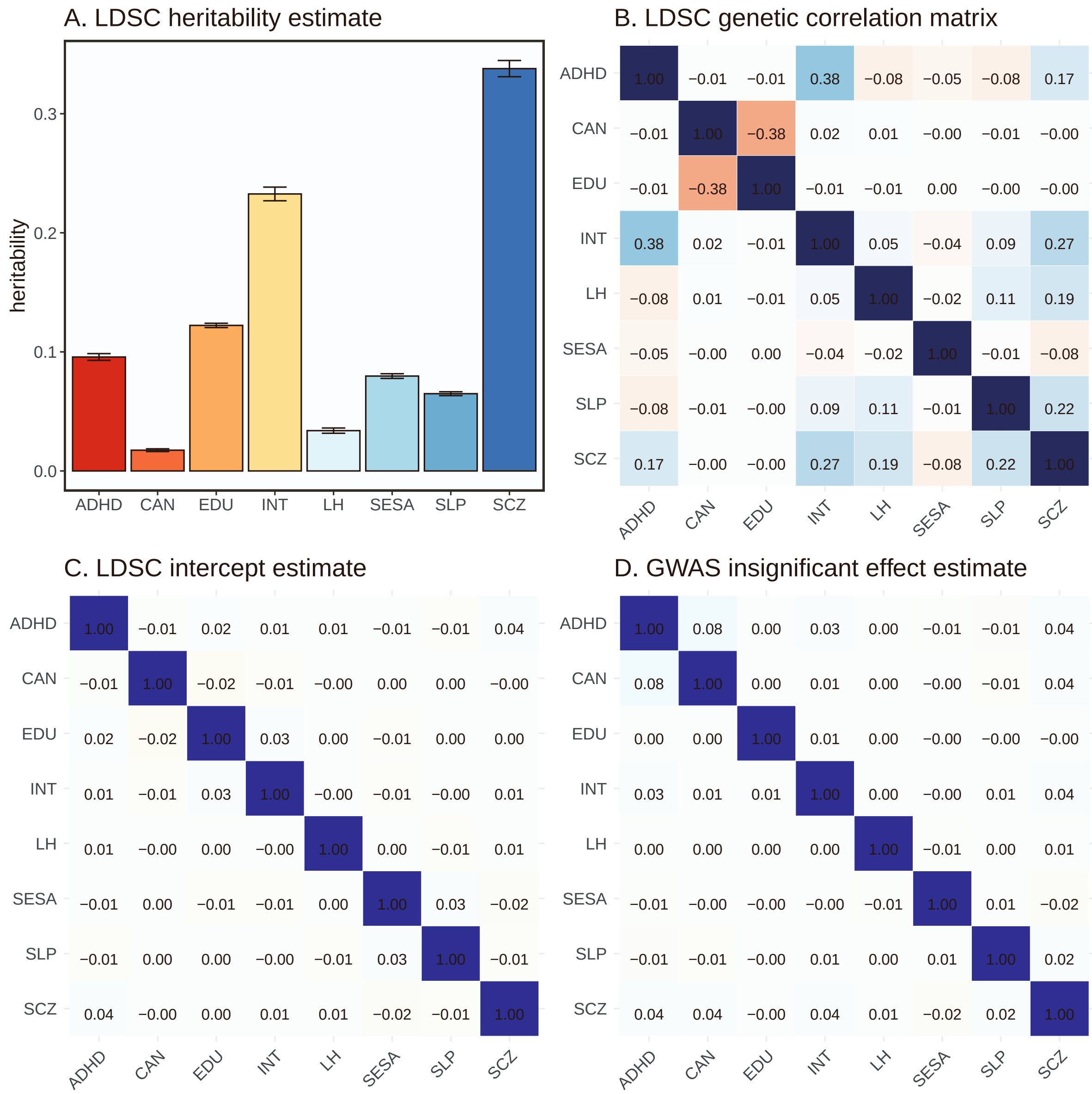


Figure S17: SCZ data. A. Heritability estimated by LDSC and the corresponding confidence intervals (radius is double SE). B. Genetic correlation matrix estimated by LDSC. C. Correlation matrix of estimation error constructed using the intercept from LDSC estimation. D. Correlation matrix of estimation error constructed using GWAS insignificant statistics.

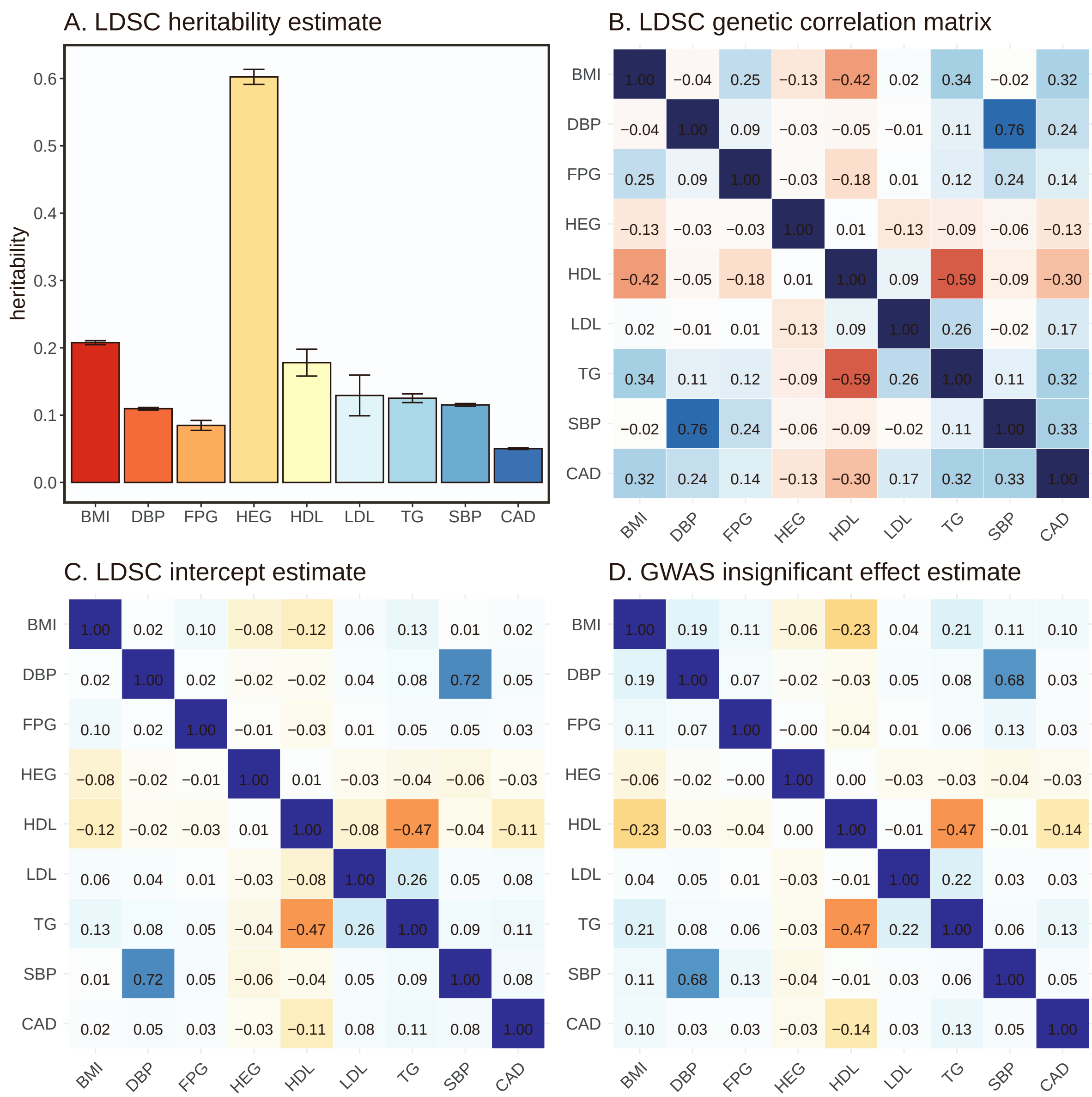


Figure S18: CAD data. A. Heritability estimated by LDSC and the corresponding confidence intervals (radius is double SE). B. Genetic correlation matrix estimated by LDSC. C. Correlation matrix of estimation error constructed using the intercept from LDSC estimation. D. Correlation matrix of estimation error constructed using GWAS insignificant statistics.

Supplementary material 2 of MRBEE: asymptotic results

Noah Lorincz-Comi, Yihe Yang, Gen Li, Xiaofeng Zhu

Contents

1 Asymptotic Results

1.1 Regular conditions

we investigate the asymptotic behavior of the multivariable IVW estimate as the number of IVs m and the minimum sample size $n_{\min}$ go to infinity. To facilitate the theoretical derivation, we specify three definitions and four regularity conditions.

Definition 1.1 (Sub-Gaussian variable)**.** *A random variable x is sub-Gaussian distributed with sub-Gaussian parameter $\tau_x > 0$ if for all $t > 0$, $\Pr(|x - E(x)| \geq t) \leq 2e^{-t^2/\tau_x^2}$.*

Definition 1.2 (Well-conditioned covariance matrix)**.** *A covariance matrix $\boldsymbol{\Sigma}$ is well-conditioned if there is a positive constant d_0 such that $0 < d_0^{-1} \leq \lambda_{min}(\boldsymbol{\Sigma}) \leq \lambda_{max}(\boldsymbol{\Sigma}) \leq d_0 < \infty$.*

Definition 1.3 (Strongly asymptotically unbiased estimate)**.** *Let $\hat{\boldsymbol{\theta}}$ be a consistent estimate of $\boldsymbol{\theta}$ with an asymptotic normal distribution $\sqrt{s_n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$, where $\boldsymbol{\mu}_\theta$ is a vector with a bounded ℓ_2-norm, $\boldsymbol{\Sigma}_\theta$ is a well-conditioned covariance matrix, and s_n is a sequence of n. Then $\hat{\boldsymbol{\theta}}$ is called a strongly asymptotically unbiased estimate of $\boldsymbol{\theta}$ if $\boldsymbol{\mu}_\theta = \mathbf{0}$.*

Sub-Gaussianity and well-conditioned covariance matrix are two of the basic concepts in modern statistics (Vershynin, 2018). In addition, we define the strongly asymptotic unbiasedness to distinguish the consistent estimate whose squared bias vanishes with an equal and a smaller rate than its variance, respectively. If an estimate is consistent but its squared bias and variance vanish at the same rate, the classic confidence interval cannot cover the true parameter with a probability of 0.95, thus leading to invalid statistical inference (Jankova, 2018).

Condition 1.1 (Regularity conditions for multivariable MR)**.**

- *(C1) For $\boldsymbol{g}_i = (g_{i1}, \ldots, g_{im})^\top$, each entry g_{ij} is a bounded sub-Gaussian with $\mathbb{E}(g_{ij})$=0, $\boldsymbol{var}(g_{ij})$=1, and sub-Gaussian parameter $\tau_g \in (0, \infty)$. For all $(i,j) \neq (t,s)$, g_{ij} is independent of g_{ts}.*
- *(C2) For $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ip})^\top$, each entry u_{ij} is a sub-Gaussian with $\mathbb{E}(u_{ij}) = 0$, $\boldsymbol{var}(u_{is}) \in (0, \infty)$, and sub-Gaussian parameter $\tau_u \in (0, \infty)$; v_i is a sub-Gaussian with $\mathbb{E}(v_i) = 0$, $\boldsymbol{var}(v_i) \in (0, \infty)$, and sub-Gaussian parameter $\tau_v \in (0, \infty)$. Besides, $(\boldsymbol{u}_i^\top, v_i)^\top$ is independent of $(\boldsymbol{u}_t^\top, v_t)^\top$ for all $i \neq t$. Furthermore, $\boldsymbol{\Sigma}_{u\times v}$ is a well-conditioned covariance matrix of $(\boldsymbol{u}_i^\top, v_i)^\top$.*
- *(C3) For $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp})^\top$, $\sqrt{m}\beta_{js}$ is sub-Gaussian with $\mathbb{E}(\sqrt{m}\beta_{js}) = 0$, $\boldsymbol{var}(\sqrt{m}\beta_{js}) \in (0, \infty)$, and sub-Gaussian parameter $\tau_\beta \in (0, \infty)$. For all $j \neq t$, $\boldsymbol{\beta}_j$ is independent of $\boldsymbol{\beta}_t$ and $\boldsymbol{\Psi}_{\beta\beta}$ is a well-conditioned covariance matrix of $\sqrt{m}\boldsymbol{\beta}_j$.*
- *(C4) The genetic variant g_{ij}, the genetic effect $\boldsymbol{\beta}_j$, the noise terms $\boldsymbol{u}_i$ and v_i, are three mutually independent groups.*

Conditions (C1)-(C4) restrict that all variables involved in this paper are sub-Gaussian distributed. In practice, g_{ij} is standardized from a binomial variable with status 0, 1, and 2. Hence, it is supposedly a bounded sub-Gaussian variable as long as its MAF is not rare. Besides, we assume $\sqrt{m}\boldsymbol{\beta}_j$ to be sub-Gaussian with a well-conditioned covariance matrix $\boldsymbol{\Psi}_{\beta\beta}$, because the cumulative covariance explained by the m IVs $\boldsymbol{\Psi}_{\beta\beta}$ should be fixed while the covariance explained by each IV $\boldsymbol{\Sigma}_{\beta\beta} \to 0$ as $m \to \infty$. This is because we adopt the infinitesimal random effect model in which $\text{cov}(\boldsymbol{\beta}_j) = h_m^2/m$ (Bulik-Sullivan et al., 2015; Fisher, 1919), where h_m^2 is the additive SNP heritability explained by the m IVs. In MR analysis, the number of IVs can increase as the sample size increases because of increasing statistical power. Our theoretical work assumes that the heritability of IVs always keeps a constant. This is a reasonable assumption because the effect sizes because smaller and smaller under the infinitesimal model as the number of causal SNPs grows. In additional, the sub-Gaussian distribution is more general than the normal distribution, allowing for the possibility of partial elements in $\boldsymbol{\beta}_j$ to be a product of a continuous variable and a binary variable. This flexibility aligns with the scenario in multivariable MR analysis where the IVs from multiple exposures are combined, inevitably leading to the inclusion of numerous weak or null IVs for some exposures.

1.2 Asymptotic Results for Multivariable IVW

Theorem 1.1. *Denote* $w_{\alpha_j} = \hat{\alpha}_j - \alpha_j$ *and* $\omega_{js} = \hat{\beta}_{js} - \beta_{js}$, $s = 1, \ldots, p$. *If conditions (C1)-(C4) are satisfied, then for all* j,

$$\begin{pmatrix} \sqrt{n_0} w_{\alpha_j} \\ \sqrt{n_1} w_{\beta_{1j}} \\ \vdots \\ \sqrt{n_p} w_{\beta_{1p}} \end{pmatrix} \xrightarrow{D} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & \frac{n_{01}}{\sqrt{(n_0 n_1)}}\sigma_{yx_1} & \cdots & \frac{n_{01}}{\sqrt{(n_0 n_p)}}\sigma_{yx_p} \\ \frac{n_{01}}{\sqrt{(n_0 n_1)}}\sigma_{yx_1} & \sigma_{x_1 x_1} & \cdots & \frac{n_{1p}}{\sqrt{(n_1 n_p)}}\sigma_{x_1 x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n_{0p}}{\sqrt{(n_0 n_p)}}\sigma_{yx_p} & \frac{n_{1p}}{\sqrt{(n_1 n_p)}}\sigma_{x_1 x_p} & \cdots & \sigma_{x_p x_p} \end{pmatrix} \right),$$

if $n_0, \ldots, n_p$ *and* $m \to \infty$.

Theorem 1.1 demonstrates the asymptotic normal distribution of the estimation errors, from which we are able to obtain

$$\boldsymbol{\Sigma}_{W_\beta W_\beta} = \boldsymbol{\Delta}_{xx} \odot \boldsymbol{\Sigma}_{xx}, \quad \boldsymbol{\sigma}_{W_\beta w_\alpha} = \boldsymbol{\delta}_{xy} \odot \boldsymbol{\sigma}_{xy}, \quad \sigma_{w_\alpha w_\alpha} = \sigma_{yy}/n_0, \tag{1}$$

where the (j,s)th element of $\boldsymbol{\Delta}_{xx}$ is $n_{js}/(n_j n_s)$, the jth element of $\boldsymbol{\delta}_{xy}$ is $n_{j0}/(n_0 n_j)$, and the operator $\odot$ is the Hadamard product of two matrices. Our work is the first to rigorously prove this theorem under regularity conditions (C1)-(C4) and highlight the role of sample overlap.

Based on this theorem, the expectations of $\boldsymbol{S}_{\texttt{IVW}}(\boldsymbol{\theta})$ and $\mathbf{H}_{\texttt{IVW}}$ are given by

$$\texttt{E}(\boldsymbol{S}_{\texttt{IVW}}(\boldsymbol{\theta})) = (\boldsymbol{\Delta}_{xx} \odot \boldsymbol{\Sigma}_{xx})\boldsymbol{\theta} - \boldsymbol{\delta}_{xy} \odot \boldsymbol{\sigma}_{xy}, \quad \texttt{E}(\mathbf{H}_{\texttt{IVW}}) = \boldsymbol{\Sigma}_{\beta\beta} + \boldsymbol{\Delta}_{xx} \odot \boldsymbol{\Sigma}_{xx}. \tag{2}$$

By expressing $\boldsymbol{\sigma}_{xy} = \boldsymbol{\Sigma}_{xx}\boldsymbol{\theta} + \boldsymbol{\sigma}_{uv}$, an alternative expectation of $\boldsymbol{S}_{\texttt{IVW}}(\boldsymbol{\theta})$) is obtained:

$$\underbrace{\texttt{E}(\boldsymbol{S}_{\texttt{IVW}}(\boldsymbol{\theta}))}_{\text{measurement error bias}} = \underbrace{\{(\boldsymbol{\Delta}_{xx} - \boldsymbol{\delta}_{xy}\mathbf{1}^\top) \odot \boldsymbol{\Sigma}_{xx}\}\boldsymbol{\theta}}_{\text{null bias}} - \underbrace{\boldsymbol{\delta}_{xy} \odot \boldsymbol{\sigma}_{uv}}_{\text{confounder bias}}. \tag{3}$$

From this expectation, it is clear that there are two sources of measurement error bias: $\{(\boldsymbol{\Delta}_{xx} - \boldsymbol{\delta}_{xy}\mathbf{1}^\top) \odot \boldsymbol{\Sigma}_{xx}\}\boldsymbol{\theta}$ comes from the measurement error, while $\{\boldsymbol{\delta}_{xy} \odot \boldsymbol{\sigma}_{uv}\}$ is caused by the confounder. Here, we call $\{(\boldsymbol{\Delta}_{xx} - \boldsymbol{\delta}_{xy}\mathbf{1}^\top) \odot \boldsymbol{\Sigma}_{xx}\}\boldsymbol{\theta}$ null bias because it always shrinks the coefficient estimate toward zero. In contrast, we term $\{\boldsymbol{\delta}_{xy} \odot \boldsymbol{\sigma}_{uv}\}$ confounder bias because $\boldsymbol{\sigma}_{uv} \neq \mathbf{0}$ implies that there are underlying confounders simultaneously affecting both $\boldsymbol{x}_i$ and y_i. Moreover, the overlapping fractions $\boldsymbol{\delta}_{xy}$ linearly trade off these two sources of biases. Generally, null bias is dominant when the elements of $\boldsymbol{\delta}_{xy}$ are small, while confounder bias dominates when the elements of $\boldsymbol{\delta}_{xy}$ are large. And there may exist a special sample overlap such that $\boldsymbol{\delta}_{xy} \odot \boldsymbol{\sigma}_{uv} = \{(\boldsymbol{\Delta}_{xx} - \boldsymbol{\delta}_{xy}\mathbf{1}^\top) \odot \boldsymbol{\Sigma}_{xx}\}\boldsymbol{\theta}$. In univariable MR, this special fraction is $n_{01}/n_0 = \sigma_{xx}\theta/\sigma_{xy}$, which guarantees that $\texttt{E}(S_{\texttt{IVW}}(\theta)) = 0$ and $\texttt{E}(\hat{\theta}_{\texttt{IVW}}) = \theta$. This theoretical result explains why in the empirical studies, $\hat{\theta}_{\texttt{IVW}}$ has a negative bias when n_{01}/n_0 is small, has a positive bias when n_{01}/n_0 is large, and is unbiased at this specific point.

Theorem 1.2. *Suppose conditions (C1)-(C4) hold and* m, $n_{\min} \to \infty$. *Then*

(i) *if* $m/\sqrt{n_{\min}} \to 0$, $\sqrt{n_{\min}}(\hat{\boldsymbol{\theta}}_{\texttt{IVW}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \boldsymbol{\Psi}_{\beta\beta}^{-1})$;

(ii) *if* $m/\sqrt{n_{\min}} \to c_0$, $\sqrt{n_{\min}}(\hat{\boldsymbol{\theta}}_{\texttt{IVW}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(-c_0 \boldsymbol{\Psi}_{\beta\beta}^{-1}(\boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta} - \boldsymbol{\psi}_{W_\beta w_\alpha}), \psi_\theta \boldsymbol{\Psi}_{\beta\beta}^{-1})$;

(iii) *if* $m/n_{\min} \to 0$, $||\hat{\boldsymbol{\theta}}_{\texttt{IVW}} - \boldsymbol{\theta}||_2 = O_P(m/n_{\min})$;

(iv) *if* $m/n_{\min} \to c_0 \in (0, \infty)$, $\hat{\boldsymbol{\theta}}_{\texttt{IVW}} - \boldsymbol{\theta} \xrightarrow{P} -c_0(\boldsymbol{\Psi}_{\beta\beta} + c_0 \boldsymbol{\Psi}_{W_\beta W_\beta})^{-1}(\boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta} - \boldsymbol{\psi}_{W_\beta w_\alpha})$;

(v) *if* $m/n_{\min} \to \infty$, $\hat{\boldsymbol{\theta}}_{\texttt{IVW}} \xrightarrow{P} \boldsymbol{\Psi}_{W_\beta W_\beta}^{+} \boldsymbol{\psi}_{W_\beta w_\alpha}$;

where

$$\mathbf{\Psi}_{W_\beta \times w_\alpha} = \begin{pmatrix} \mathbf{\Psi}_{W_\beta W_\beta} & \boldsymbol{\psi}_{W_\beta w_\alpha} \\ \boldsymbol{\psi}_{W_\beta w_\alpha}^\top & \psi_{w_\alpha w_\alpha} \end{pmatrix} = \lim_{n_{\min}\to\infty} \begin{pmatrix} n_{\min}\mathbf{\Sigma}_{W_\beta W_\beta} & n_{\min}\boldsymbol{\sigma}_{W_\beta w_\alpha} \\ n_{\min}\boldsymbol{\sigma}_{W_\beta w_\alpha}^\top & n_{\min}\sigma_{w_\alpha w_\alpha} \end{pmatrix},$$

and $\psi_{\theta} = \psi_{w_\alpha w_\alpha} + \boldsymbol{\theta}^\top \mathbf{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta} - 2\boldsymbol{\theta}^\top \boldsymbol{\psi}_{W_\beta w_\alpha}$.

Theorem 1.2 is one of two main theorems in this paper and points out five scenarios. First, if m goes to infinity with a lower rate than $\sqrt{n_{\min}}$, then $\hat{\boldsymbol{\theta}}_{\texttt{IVW}}$ is strongly asymptotically unbiased. In other words, $\hat{\boldsymbol{\theta}}_{\texttt{IVW}}$ is able to reliably infer causality only when the sample size of GWAS data is quadratically larger than the number of IVs. On the other hand, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{\texttt{IVW}}$ is the inverse of the cumulative covariance matrix $\mathbf{\Psi}_{\beta\beta} = \sum_{j=1}^m \texttt{cov}(\boldsymbol{\beta}_j)$, therefore, it is optimal to include as many associated variants as possible in order to have $\mathbf{\Psi}_{\beta\beta}$ large enough. In contrast, using a few top significant variants to perform MR analysis is not recommended.

Second, if m tends to infinity with the same rate as $\sqrt{n_{\min}}$, $\sqrt{n_{\min}}(\hat{\boldsymbol{\theta}}_{\texttt{IVW}} - \boldsymbol{\theta})$ converges to an asymptotic normal distribution with a non-zero asymptotic bias $\{c_0\mathbf{\Psi}_{\beta\beta}^{-1}(\boldsymbol{\psi}_{W_\beta w_\alpha} - \mathbf{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta})\}$. In this asymptotic bias, $\{c_0(\boldsymbol{\psi}_{W_\beta w_\alpha} - \mathbf{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta})\}$ is caused by $\boldsymbol{S}_{\texttt{IVW}}(\boldsymbol{\theta})$ and $\mathbf{\Psi}_{\beta\beta}^{-1}$ is caused by $\mathbf{H}_{\texttt{IVW}}^{-1}$. Since the asymptotic bias and asymptotic covariance matrix are of the same order in this scenario, the inference made is invalid although the bias of $\hat{\boldsymbol{\theta}}_{\texttt{IVW}}$ is infinitesimal. When $m/n_{\min} \to 0$, $\hat{\boldsymbol{\theta}}_{\texttt{IVW}}$ still converges to $\boldsymbol{\theta}$ with a rate $O(m/n_{\min})$, but it no longer has an asymptotic normal distribution. Scenario (iv) is more serious than (iii) because the bias of $\hat{\boldsymbol{\theta}}_{\texttt{IVW}}$ will not vanish even when $\sqrt{n_{\min}}$ goes to infinity. In the fifth scenario, $\hat{\boldsymbol{\theta}}_{\texttt{IVW}}$ converges to a term irrelevant to $\boldsymbol{\theta}$.

1.3 Asymptotic Results for MRBEE

Theorem 1.3. *Suppose conditions (C1)-(C4) hold and m, $n_{\min} \to \infty$. Then*

(i) if $m/n_{\min} \to 0$, $\sqrt{n_{\min}}(\hat{\boldsymbol{\theta}}_{BEE} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \boldsymbol{\Psi}_{\beta\beta}^{-1})$;

(ii) if $m/n_{\min} \to c_0 \in (0, \infty)$, $\sqrt{n_{\min}}(\hat{\boldsymbol{\theta}}_{BEE} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \boldsymbol{\Psi}_{\beta\beta}^{-1} + c_0 \boldsymbol{\Psi}_{\beta\beta}^{-1} \boldsymbol{\Psi}_{\mathrm{BC}} \boldsymbol{\Psi}_{\beta\beta}^{-1})$;

(iii) if $m/n_{\min} \to \infty$ and $m/n_{\min}^2 \to 0$, $\sqrt{(n_{\min}^2/m)}(\hat{\boldsymbol{\theta}}_{BEE} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_{\beta\beta}^{-1} \boldsymbol{\Psi}_{\mathrm{BC}} \boldsymbol{\Psi}_{\beta\beta}^{-1})$;

where ψ_θ is defined in Theorem 1.2 and $\boldsymbol{\Psi}_{\mathrm{BC}}$ is a semi-positive symmetric matrix whose expression is shown in equation (64) in supplementary materials.

Theorem 1.3 indicates three scenarios. First, if $m/n \to 0, \sqrt{n_{\min}}(\hat{\boldsymbol{\theta}}_{\mathtt{BEE}} - \boldsymbol{\theta})$ converges to a normal distribution with a zero mean and the covariance matrix being exactly the same as $\hat{\boldsymbol{\theta}}_{\mathtt{IVW}}$. In other words, $\hat{\boldsymbol{\theta}}_{\mathtt{BEE}}$ is not only strongly asymptotically unbiased but also loses no efficiency in comparison to $\hat{\boldsymbol{\theta}}_{\mathtt{IVW}}$. Second, if $m/n_{\min} \to c_0 \in (0, \infty)$, there is an additional covariance matrix $c_0 \boldsymbol{\Psi}_{\beta\beta}^{-1} \boldsymbol{\Psi}_{\mathrm{BC}} \boldsymbol{\Psi}_{\beta\beta}^{-1}$ in the asymptotic normal distribution, where $\boldsymbol{\Psi}_{\mathrm{BC}}$ is introduced by the bias-correction terms:

$$\boldsymbol{\Psi}_{\mathrm{BC}} = \lim_{n_{\min} \to \infty} \mathtt{var}\left[\frac{n_{\min}}{\sqrt{m}} \left((\mathbf{W}_\beta^\top \mathbf{W}_\beta - m \boldsymbol{\Sigma}_{W_\beta W_\beta}) \boldsymbol{\theta} - (\mathbf{W}_\beta^\top \boldsymbol{w}_\alpha - m \boldsymbol{\sigma}_{W_\beta w_\alpha}) \right)\right]. \tag{4}$$

In this scenario, $\hat{\boldsymbol{\theta}}_{\mathtt{BEE}}$ is again strongly asymptotically unbiased with a convergence rate $\sqrt{n_{\min}}$, while $\hat{\boldsymbol{\theta}}_{\mathtt{IVW}}$ incurs a bias not vanishing asymptotically. In the third scenario, $\hat{\boldsymbol{\theta}}_{\mathtt{BEE}}$ is still strongly asymptotically unbiased with a convergence rate $\sqrt{(n_{\min}^2/m)}$, and the asymptotic distribution is dominated by the bias correction terms. Note that $\hat{\boldsymbol{\theta}}_{\mathtt{IVW}}$ is not consistent unless $m/n_{\min} \to 0$ and the inference made by $\hat{\boldsymbol{\theta}}_{\mathtt{IVW}}$ is unreliable unless $m/\sqrt{n_{\min}} \to 0$. In contrast, $\hat{\boldsymbol{\theta}}_{\mathtt{BEE}}$ is strongly asymptotically unbiased as long as $m/n_{\min}^2 \to 0$. Thus, MRBEE is superior to multivariable IVW in terms of both unbiasedness and asymptotic validity in all possible scenarios.

Theorem 1.4. *Suppose conditions (C1)-(C4) hold. Let $g_{ij}^{\{s\}}$ satisfy the condition (C1), $E(x_i^{[s]} | g_{ij}^{\{s\}}) = 0$ for all $1 \le s \le p$, and $E(y_i^{[0]} | g_{ij}^{\{0\}}) = 0$. Then*

$$||\boldsymbol{\Sigma}_{W_\beta \times w_\alpha}^{-\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}_{W_\beta \times w_\alpha} \boldsymbol{\Sigma}_{W_\beta \times w_\alpha}^{-\frac{1}{2}} - \mathbf{I}_{p+1}||_2 = O_P\left(\frac{1}{\sqrt{M}}\right),$$

if $n_{\min}$ and $M \to \infty$.

Theorem 1.4 shows that $\widehat{\boldsymbol{\Sigma}}_{W_\beta \times w_\alpha}$ has a $O(\sqrt{M})$ convergence rate after adjusting the scale of $\boldsymbol{\Sigma}_{W_\beta \times w_\alpha}$. As there may be more than 1 million independent variants in the whole genome, $\widehat{\boldsymbol{\Sigma}}_{W_\beta \times w_\alpha}$ has high precision. Besides, $n_0, n_1, ..., n_p \to \infty$ are required such that $\sqrt{n_0} \hat{\alpha}_j^*$ and $\sqrt{n_s} \hat{\beta}_{js}^*$ are asymptotically normally distributed.

Theorem 1.5. *Under the conditions of Theorem 1.4,*

$$||\boldsymbol{\Sigma}_{BEE}^{-\frac{1}{2}}(\boldsymbol{\theta}) \widehat{\boldsymbol{\Sigma}}_{BEE}(\hat{\boldsymbol{\theta}}_{BEE}) \boldsymbol{\Sigma}_{BEE}^{-\frac{1}{2}}(\boldsymbol{\theta}) - \mathbf{I}_p||_2 = O_P\left(\max\left\{\frac{1}{\sqrt{n_{\min}}}, \frac{\sqrt{m}}{n_{\min}}, \sqrt{\frac{\log m}{m}}\right\}\right)$$

if $n_{\min}, m$ and $M \to \infty$ and $m/n_{\min}^2 \to 0$.

Theorem 1.5 shows that $\widehat{\boldsymbol{\Sigma}}_{\mathtt{BEE}}(\boldsymbol{\theta})$ has a $\min(\sqrt{n_{\min}}, \sqrt{(n_{\min}^2/m)}, \sqrt{(m/\log m)})$ convergence rate when $m/n_{\min}^2 \to 0$. The first two convergence rates are brought by $||\widehat{\mathbf{F}}_{\mathtt{BEE}} - \mathbf{F}_{\mathtt{BEE}}||_2$, while the third convergence rate is yielded by $||\widehat{\mathbf{V}}_{\mathtt{BEE}}(\hat{\boldsymbol{\theta}}_{\mathtt{BEE}}) - \mathbf{V}_{\mathtt{BEE}}(\boldsymbol{\theta})||_2$. Note that the SE estimation should be of the same importance as the causal effect estimation. Although the inference is made based on an unbiased estimate, it could still be invalid if the SE estimate is not reliable. As the dependability of the sandwich formula has been extensively investigated empirically, it is a reliable technique to obtain the SE estimate for MRBEE.

Theorem 1.6. *Assume that* $|\mathcal{O}|$ *is fixed and bounded and* $\gamma_1^*, \dots, \gamma_m^*$ *are a series of non-random numbers. Then under the conditions of Theorem 1.5, there exists a threshold* $\kappa = F_{\chi_1^2}(C_0 \log m)$ *such that* $\Pr(\mathcal{O} = \hat{\mathcal{O}}) \to 1$*, where* $\hat{\mathcal{O}} = \{j : F_{\chi_1^2}(t_{\gamma_j}) > \kappa\}$ *and* C_0 *is a sufficiently large constant.*

Theorem 1.6 indicates that there is a theoretical threshold $\kappa = F_{\chi_1^2}(C_0 \log m)$ to consistently identify all horizontal pleiotropy. This threshold increases with a rate $O(\log m)$ to reduce the false discovery rate (FDR) and its concrete value can be chosen by a FDR control method (Benjamini, 1995). In practice, MRBEE iteratively applies the hypothesis test to remove the outliers and uses the remaining IVs to estimate $\boldsymbol{\theta}$. The stable estimate is regarded as $\hat{\boldsymbol{\theta}}_{\texttt{BEE}}$.

1.4 Preliminary lemmas

In this subsection, we specify some lemmas that can facilitate the proofs, most of which can be found in the existing papers. We first discuss the equivalent characterizations of sub-Gaussian and sub-exponential variables.

Lemma 1.1 (Equivalent characterizations of sub-Guassian variables). *Given any random variable X, the following properties are equivalent:*

(I) there is a constant $K_1 \geq 0$ such that

$$\Pr(|X| \geq t) \leq 2\exp(-t^2/K_1^2), \quad \textit{for all } t \geq 0,$$

(II) the moments of X satisfy

$$||X||_{L_p} = (E(|X|^p))^{\frac{1}{p}} \leq K_2\sqrt{p}, \quad \textit{for all } p \geq 1,$$

(III) the moment generating function (MGF) of X^2 satisfies:

$$E\{\exp(\lambda^2 X^2)\} \leq \exp(K_3^2\lambda^2), \quad \textit{for all } \lambda \textit{ staisfying } |\lambda| \leq K_3^{-1},$$

(IV) the MGF of X^2 is bounded at some point, namely

$$E\{\exp(X^2/K_4^2)\} \leq 2,$$

(V) if $E(X) = 0$, the MGF of X satisfies

$$E\{\exp(\lambda X)\} \leq \exp(K_5^2\lambda^2), \quad \textit{for all } \lambda \in \mathbb{R},$$

where $K_1, \ldots, K_5$ are certain strictly positive constants.

This lemma summarizes some well-known properties of sub-Guassian and can be found in Vershynin (2018, Proposition 2.5.2).

Lemma 1.2 (Equivalent characterizations of sub-exponential variables). *Given any random variable X, the following properties are equivalent:*

(I) there is a constant $K_1 \geq 0$ such that

$$\Pr(|X| \geq t) \leq 2\exp(-t/K_1), \quad \textit{for all } t \geq 0,$$

(II) the moments of X satisfy

$$||X||_{L_p} = (E(|X|^p))^{\frac{1}{p}} \leq K_2 p, \quad \textit{for all } p \geq 1,$$

(III) the moment generating function (MGF) of $|X|$ satisfies:

$$E\{\exp(\lambda|X|)\} \leq \exp(K_3\lambda), \quad \textit{for all } \lambda \textit{ staisfying } 0 \leq \lambda \leq K_3^{-1},$$

(IV) the MGF of $|X|$ is bounded at some point, namely

$$E\{\exp(|X|/K_4)\} \leq 2,$$

(V) if $E(X) = 0$, the MGF of X satisfies

$$E\{\exp(\lambda X)\} \leq \exp(K_5^2\lambda^2), \quad \textit{for all } \lambda \leq K_5^{-1},$$

where $K_1, \ldots, K_5$ are certain strictly positive constants.

This lemma summarizes some well-known properties of sub-exponential and can be found in Vershynin (2018, Proposition 2.7.1).

Lemma 1.3 (Product of sub-Gaussian variable is sub-exponential)**.** *Suppose that X, Z are two sub-Gaussian variable, then $Y = XZ$ is a sub-exponential variable. Besides, if X is a bounded sub-Gaussian variable, then then $Y = XZ$ is a sub-Gaussian variable.*

The first claim of this lemma is provided by Vershynin (2018, Proposition 2.7.7). The second claim of this lemma is a direct inference of Fan et al. (2011, Lemma A.2).

Lemma 1.4 (ℓ_2-norm of matrices with sub-Gaussian entries)**.** *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n $(p \times 1)$ independent identically distributed random vector with entries $x_{i1}, \ldots, x_{ip}$ are sub-Gaussian with zero-mean. Besides, define the covariance matrix of $\boldsymbol{X}_i$ as*

$$\boldsymbol{\Sigma} = E(\boldsymbol{X}_i \boldsymbol{X}_i^\top)$$

and the related sample covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top.$$

Then for every positive integer n,

$$E(||\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_2) \leq C\left(\frac{p}{n} + \sqrt{\frac{p}{n}}\right)||\boldsymbol{\Sigma}||_2,$$

where C is certain positive constant.

This lemma is provided by Vershynin (2018, Theorem 4.7.1). It shows the convergence rate of sample covariance matrix is $\sqrt{(n/m)}$.

Lemma 1.5 (ℓ_2-norm of matrices with sub-exponential entries)**.** *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n $(p \times 1)$ independent identically distributed random vector with entries $x_{i1}, \ldots, x_{ip}$ are sub-exponential with zero-mean. Besides, define the covariance matrix of $\boldsymbol{X}_i$ as*

$$\boldsymbol{\Sigma} = E(\boldsymbol{X}_i \boldsymbol{X}_i^\top)$$

and the related sample covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top.$$

Then for ever $t \geq 0$, the following inequality holds with probability at least $1 - p\exp(-ct^2)$:

$$||\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_2 \leq \max(||\boldsymbol{\Sigma}||_2 \delta, \delta^2),$$

where c is certain positive constant and $\delta = t\sqrt{p/n}$.

This lemma is the direct inference of Vershynin (2010, Theorem 5.44). Besides, by letting $t = \sqrt{p \log n}$ we further obtain

$$\mathrm{E}(\|\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|\|\_2) = O\left(\sqrt{\frac{p \log n}{n}}\right)\|\|\boldsymbol{\Sigma}\|\|\_2,$$

if $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix of sub-exponential vector. Note that in our method, the dimension p is fixed and hence we cannot chose $t = \sqrt{p \log p}$ such that the estimation bound becomes $\sqrt{(p \log p)/n}||\boldsymbol{\Sigma}||_2$.

Lemma 1.6 (Asymptotic normal distribution of Wishart matrix)**.** *Suppose* $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ *are* n *IID relaxation of the* p*-dimensional variable* $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ *with a well-conditioned covariance matrix* $\boldsymbol{\Sigma}$*. Besides, define the sample covariance matrix of* $\boldsymbol{\Sigma}$ *as*

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^{\top}.$$

If p *is a fixed number, then as* $n \to \infty$*,*

$$\sqrt{n}(vec(\hat{\boldsymbol{\Sigma}}) - vec(\boldsymbol{\Sigma})) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, (\mathbf{I}_{p^2} + \mathbf{K}_{p^2})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\right),$$

where $\mathbf{K}_{p^2}$ *is the so-called commutation matrix, which is able to ensure* $\mathbf{K}_{p^2} vec(\mathbf{A}) = vec(\mathbf{A}')$ *for all* $(p \times p)$ *matrix.*

This lemma can be found in Muirhead (2009, equation (5), p90).

1.5 Specific Lemmas

In this subsection, we specify the following lemmas that are made based on the preliminary lemmas.

Lemma 1.7 (Asymptotic normal distribution of sub-Gaussian and sub-exponential variables)**.** *Suppose $X_1, \ldots, X_n$ are n independent sub-Gaussian or sub-exponential variables with mean-zero and variance $\sigma_1^2, \ldots, \sigma_n^2$. Then*

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{D} \mathcal{N}(0, \sigma_x^2),$$

where

$$\sigma_x^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2.$$

Proof of Lemma 1.7. It is easy to verify the Lyapunov's condition: for all fixed $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathrm{E}(|X_i|^{2+2\delta}) \le \frac{\sqrt{2K_2 + 2K_2\delta}^{2+2\delta}}{n^{\delta}} \to 0$$

by the (II) of Lemma 1.1, if $X_1, \ldots, X_n$ are sub-Gaussian variables;

$$\lim_{n\to\infty} \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathrm{E}(|X_i|^{2+2\delta}) \le \frac{(2K_2 + 2K_2\delta)^{2+2\delta}}{n^{\delta}} \to 0$$

by the (II) of Lemma 1.2, if $X_1, \ldots, X_n$ are sub-exponential variables. And hence the asymptotic normal distribution holds. ☐

Lemma 1.8 (Asymptotic normal distribution of estimation error)**.** *Let*

$$\xi_j^{[s]} = \frac{1}{\sqrt{n_s}} \sum_{i=1}^{n_s} g_{ij}^{[s]} x_{i,-j}^{[s]},$$

where

$$x_{i,-j}^{[s]} = x_i^{[s]} - \beta_{js} g_{i,j}^{[s]},$$

$s = 0, 1, \ldots, p$*,* $x_{i,-j}^{[0]}$ *represents* $y_{i,-j}^{[0]}$ *and* $\beta j0$ *represent* α_j*. Then*

$$\xi_j^{[s]} \xrightarrow{D} \mathcal{N}(0, \sigma_{x_s x_s} - \sigma_{\beta_s \beta_s}),$$

where $\sigma_{x_0 x_0}$ *represents* σ_{yy} *and* $\sigma_{\beta_0\beta_0}$ *represents* $\boldsymbol{\theta}^\top \boldsymbol{\Sigma}_{\beta\beta}\theta$*.*

Proof of Lemma 1.8. Note that both $g_{ij}^{[s]}$ and $x_{i,-j}^{[s]}$ are sub-Gaussian ($x_{i,-j}^{[s]}$ is the product of a sub-Gaussian variable and a bounded sub-Gaussian variable), and it holds $\mathrm{E}(g_{ij}^{[s]} x_{i,-j}^{[s]}) = 0$ and

$$\mathrm{var}(g_{ij}^{[s]} x_{i,-j}^{[s]}) = \mathrm{var}(g_{ij}^{[s]}) \times \mathrm{var}(x_{i,-j}^{[s]}) = \sigma_{x_s x_s} - \sigma_{\beta_s \beta_s}. \tag{5}$$

As a result,

$$\xi_j^{[s]} = \frac{1}{\sqrt{n_s}} \sum_{i=1}^{n_s} g_{ij}^{[s]} x_{i,-j}^{[s]} \xrightarrow{D} \mathcal{N}(0, \sigma_{x_s x_s} - \sigma_{\beta_s \beta_s}), \tag{6}$$

according Lemma 1.7. ☐

Lemma 1.9 (Asymptotic normality of bias-correction terms)**.** *Let*

$$\boldsymbol{\zeta}_j = \left(\frac{n_{\min}}{n_1}\xi_j^{[1]}, \frac{n_{\min}}{n_2}\xi_j^{[2]}, \ldots, \frac{n_{\min}}{n_p}\xi_j^{[p]}, \frac{n_{\min}}{n_0}\xi_j^{[0]}\right)^\top.$$

Under the conditions (C1)-(C4),

$$\lim_{m\to\infty}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}(vec(\boldsymbol{\zeta}_j\boldsymbol{\zeta}_j^\top) - vec(\boldsymbol{\Psi}_{W_\beta\times w_\alpha})) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0}, (\mathbf{I}_{(p+1)^2} + \mathbf{K}_{(p+1)^2})(\boldsymbol{\Psi}_{W_\beta\times w_\alpha}\otimes\boldsymbol{\Psi}_{W_\beta\times w_\alpha})\Big).$$

as $n_{\min}, m \to \infty$.

Proof of Lemma 1.9. By using Lemma 1.7, $\boldsymbol{\zeta}_j$ follows $\mathcal{N}(0, \boldsymbol{\Psi}_{W_\beta\times w_\alpha})$ as $n_{\min}\to\infty$. Then by using Lemma 1.6, this lemma holds. □

Lemma 1.10 (Asymptotic normality of residual term)**.** *Under the conditions (C1)-(C4),*

$$\lim_{m\to\infty}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\sqrt{m}\boldsymbol{\beta}_j\xi_j^{[s]} \xrightarrow{D} \mathcal{N}(0, \sigma_{x_sx_s}\boldsymbol{\Sigma}_{\beta\beta}),$$

and

$$\lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}\sqrt{m}\boldsymbol{\beta}_j\sqrt{m}\boldsymbol{\beta}_j^\top\xi_j^{[s]}\xi_j^{[k]} \xrightarrow{P} \frac{n_{sk}}{\sqrt{n_sn_k}}\sigma_{x_sx_k}\boldsymbol{\Sigma}_{\beta\beta},$$

for $s = 0, \ldots, p$*, where* $\sigma_{x_0x_k}$ *represents* $\sigma_{yx_k} = \sum_{l=1}^{p}\theta_l\sigma_{x_lx_k}$.

Proof of Lemma 1.10. By condition (C4), $\sqrt{m}\boldsymbol{\beta}_j$ is independent of $\xi_j^{[s]}$. By Lemma 1.3, $\sqrt{m}\boldsymbol{\beta}_j\xi_j^{[s]}$ is sub-exponential with mean $\mathbf{0}$ and covariance matrix

$$\begin{aligned}\text{cov}(\sqrt{m}\boldsymbol{\beta}_j\xi_j^{[s]}) &= \text{cov}(\sqrt{m}\boldsymbol{\beta}_j)\times\text{var}(\xi_j^{[s]})\\ &= (\sigma_{x_sx_s} - \sigma_{\beta_s\beta_s})\boldsymbol{\Sigma}_{\beta\beta}.\end{aligned}\tag{7}$$

Hence, by Lemma 1.6,

$$\lim_{m\to\infty}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\sqrt{m}\boldsymbol{\beta}_j\xi_j^{[s]} \xrightarrow{D} \mathcal{N}(0, \sigma_{x_sx_s}\boldsymbol{\Sigma}_{\beta\beta}).$$

On the other hand, $\boldsymbol{\beta}_j\xi_j^{[s]}$ is sub-exponential variable according to Lemma 1.3, and

$$\begin{aligned}\text{cov}(\sqrt{m}\boldsymbol{\beta}_j\xi_j^{[s]}, \sqrt{m}\boldsymbol{\beta}_j\xi_j^{[k]}) &= \text{cov}(\xi_j^{[s]}, \xi_j^{[k]})\times\boldsymbol{\Sigma}_{\beta\beta}\\ &= \frac{n_{sk}}{\sqrt{n_sn_k}}(\sigma_{x_sx_k} - \sigma_{\beta_s\beta_k})\boldsymbol{\Sigma}_{\beta\beta}.\end{aligned}\tag{8}$$

Hence, by using Lemma 1.5

$$\lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}\sqrt{m}\boldsymbol{\beta}_j\sqrt{m}\boldsymbol{\beta}_j^\top\xi_j^{[s]}\xi_j^{[k]} \xrightarrow{P} \frac{n_{sk}}{\sqrt{n_sn_k}}\sigma_{x_sx_k}\boldsymbol{\Sigma}_{\beta\beta}.$$

□

1.6 Proofs of Theorems for IVW

Proof of Theorem 1.1. As for the estimation error $\boldsymbol{\omega}_\alpha$, we have

$$w_{\alpha_j} = \frac{\boldsymbol{g}_j^{[0]\top}\boldsymbol{y}^{[0]}}{n_0} - \alpha_j = \frac{\boldsymbol{g}_j^{[0]\top}\boldsymbol{y}_{-j}^{[0]}}{n_0}, \tag{9}$$

where

$$\boldsymbol{y}_{-j}^{[0]} = \boldsymbol{y}^{[0]} - \alpha_j\boldsymbol{g}_j^{[0]} = \sum_{s\neq j}^{m}\alpha_t\boldsymbol{g}_t^{[0]} + \mathbf{U}^{[0]}\boldsymbol{\theta} + \boldsymbol{v}^{[0]}, \tag{10}$$

and $\mathbf{U}^{[0]}$ and $\boldsymbol{v}^{[0]}$ are the corresponding noise terms in the outcome GWAS cohort. According to Lemma 1.8,

$$\xi_j^{[0]} = \frac{1}{\sqrt{n_0}}\sum_{i=1}^{n_0} g_{ij}^{[0]}y_{i,-j}^{[0]} \xrightarrow{D} \mathcal{N}(0, \sigma_{yy} - \boldsymbol{\theta}^\top\boldsymbol{\Sigma}_{\beta\beta}\boldsymbol{\theta}). \tag{11}$$

As for the estimation error $w_{\beta_{js}}$, we have

$$w_{\beta_{js}} = \frac{\boldsymbol{g}_j^{[s]\top}\boldsymbol{x}^{[s]}}{n_s} - \beta_{js} = \frac{\boldsymbol{g}_j^{[s]\top}\boldsymbol{x}_{-j}^{[s]}}{n_s}, \tag{12}$$

where

$$\boldsymbol{x}_{-j}^{[s]} = \boldsymbol{x}^{[s]} - \boldsymbol{g}_j^{[s]}\beta_{js} = \sum_{t\neq j}\beta_{ts}\boldsymbol{g}_t^{[s]} + \boldsymbol{u}^{[s]}. \tag{13}$$

Let

$$\xi_j^{[s]} = \frac{\boldsymbol{g}_j^{[s]\top}\boldsymbol{x}_{-j}^{[s]}}{\sqrt{n_s}} = \frac{1}{\sqrt{n}_s}\sum_{i=1}^{n_s} g_{ij}^{[s]}x_{i,-j}^{[s]}, \tag{14}$$

where $x_{i,-j}^{[s]}$ is the ith element in vector $\boldsymbol{x}_{-j}^{[s]}$. According to Lemma 1.8,

$$\xi_j^{[s]} = \frac{1}{\sqrt{n}_s}\sum_{i=1}^{n_s} g_{ij}^{[s]}x_{i,-j}^{[s]} \xrightarrow{D} \mathcal{N}(0, \sigma_{x_sx_s} - \sigma_{\beta_s\beta_s}). \tag{15}$$

Now we show the covariance between $\xi_j^{[s]}$ and $\xi_j^{[k]}$:

$$\mathrm{cov}(\xi_j^{[s]}, \xi_j^{[k]}) = \mathrm{E}\left(\frac{\boldsymbol{x}_{-j}^{[s]\top}\boldsymbol{g}_j^{[s]}\boldsymbol{g}_j^{[k]\top}\boldsymbol{x}_{-j}^{[k]}}{\sqrt{n_sn_k}}\right), \tag{16}$$

where $\boldsymbol{x}_{-j}^{[0]}$ represents $\boldsymbol{y}_{-j}^{[0]}$ for simplicity. Denote $\mathbf{Q}^{[sk]} = (Q_{it}^{[sk]})$ being a $(n_s \times n_k)$ matrix whose (i,t)th element is

$$Q_{it}^{[sk]} = \mathrm{E}(g_{ij}^{[s]}g_{tj}^{[k]}) = \begin{cases} 1, & (i,t) \in \mathcal{Q}^{[sk]}, \\ 0, & (i,t) \notin \mathcal{Q}^{[sk]}, \end{cases} \tag{17}$$

where

$$\mathcal{Q}^{[sk]} = \{(i,t) : \ g_{ij}^{[s]} \text{ and } g_{tj}^{[k]} \text{ come from the same individual}\}. \tag{18}$$

As a result,

$$\begin{aligned}\operatorname{cov}(\xi_j^{[s]}, \xi_j^{[k]}) &= \mathrm{E}\left(\frac{\boldsymbol{x}_{-j}^{[s]\top}\mathbf{Q}^{[sk]}\boldsymbol{x}_{-j}^{[k]}}{\sqrt{n_s n_k}}\right) = \frac{1}{\sqrt{n_s n_k}}\sum_{(i,t)\in\mathcal{Q}^{[sk]}}\mathrm{E}(x_{i,-j}^{[s]}x_{t,-j}^{[k]}) \\ &= \frac{n_{sk}}{\sqrt{n_s n_k}}\left(\sigma_{x_s x_k} - \sigma_{\beta_s\beta_k}\right), \qquad (19)\end{aligned}$$

where $\sigma_{x_0 x_k}$ represents $\sigma_{y x_k}$ for simplicity, and $\sigma_{\beta_0\beta_k}$ represents

$$\sigma_{\beta_0\beta_k} = \operatorname{cov}(\sqrt{m}\boldsymbol{\beta}_j^\top\boldsymbol{\theta}, \sqrt{m}\beta_{jk}) = \sum_{l=1}^{p}\theta_l\sigma_{\beta_l\beta_k}. \qquad (20)$$

Finally, we show $\xi_j^{[s]}$ is uncorrelated with $\xi_t^{[s]}$ for all $t \neq j$ and $s = 0, \ldots, p$. Specifically,

$$\operatorname{cov}(\xi_j^{[s]}, \xi_t^{[s]}) = \mathrm{E}\left(\frac{\boldsymbol{x}_{-j}^{[s]\top}\boldsymbol{g}_j^{[s]}\boldsymbol{g}_t^{[s]\top}\boldsymbol{x}_{-j}^{[s]}}{n_s}\right). \qquad (21)$$

According the model setting, $\boldsymbol{g}_j^{[s]}$ is independent of $\boldsymbol{g}_t^{[s]}$ for all $t \neq s$. Therefore, $\operatorname{cov}(\xi_j^{[s]}, \xi_t^{[s]}) = 0$. Note that if $m \to \infty$, $\boldsymbol{\Sigma}_{\beta\beta} = \frac{1}{m}\boldsymbol{\Psi}_{\beta\beta}$ vanishes. And so Theorem 1.1 is proved. □

Proof of Theorem 1.2. Before showing the proof, we first recall the following definitions: m is the number of IVs, $n_{\min}$ is the minimum sample size,

$$\boldsymbol{\Sigma}_{\beta\beta} = \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} \boldsymbol{\beta}_j\boldsymbol{\beta}_j^\top, \quad \boldsymbol{\Psi}_{\beta\beta} = m\boldsymbol{\Sigma}_{\beta\beta},$$

$$\boldsymbol{\Sigma}_{W_\beta W_\beta} = \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} \boldsymbol{w}_{\beta_j}\boldsymbol{w}_{\beta_j}^\top, \quad \boldsymbol{\Psi}_{W_\beta W_\beta} = n_{\min}\boldsymbol{\Sigma}_{W_\beta W_\beta},$$

$$\boldsymbol{\sigma}_{W_\beta w_\alpha} = \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} w_{\alpha_j}\boldsymbol{w}_{\beta_j}, \quad \boldsymbol{\psi}_{W_\beta w_\alpha} = n_{\min}\boldsymbol{\sigma}_{W_\beta w_\alpha},$$

$$\sigma_{w_\alpha w_\alpha} = \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} w_{\alpha_j}^2, \quad \psi_{w_\alpha w_\alpha} = n_{\min}\sigma_{w_\alpha w_\alpha}.$$

The score function of IVW is

$$-\frac{1}{m}\hat{\mathbf{B}}^\top(\hat{\boldsymbol{a}} - \hat{\mathbf{B}}\hat{\boldsymbol{\theta}}_{\text{IVW}}) = -\frac{1}{m}\hat{\mathbf{B}}^\top(\hat{\boldsymbol{a}} - \hat{\mathbf{B}}\boldsymbol{\theta}) + \frac{1}{m}\hat{\mathbf{B}}^\top\hat{\mathbf{B}}(\hat{\boldsymbol{\theta}}_{\text{IVW}} - \boldsymbol{\theta}) \tag{22}$$

which leads to

$$\mathbf{H}_{\text{IVW}}(\hat{\boldsymbol{\theta}}_{\text{IVW}} - \boldsymbol{\theta}) = -\boldsymbol{S}_{\text{IVW}}(\boldsymbol{\theta}), \tag{23}$$

where

$$\mathbf{H}_{\text{IVW}} = \frac{1}{m}\hat{\mathbf{B}}^\top\hat{\mathbf{B}}, \quad \boldsymbol{S}_{\text{IVW}}(\boldsymbol{\theta}) = -\frac{1}{m}\hat{\mathbf{B}}^\top(\hat{\boldsymbol{a}} - \hat{\mathbf{B}}\boldsymbol{\theta}). \tag{24}$$

We first work with the Hessian matrix $\mathbf{H}_{\text{IVW}}$:

$$\begin{aligned} m\mathbf{H}_{\text{IVW}} &= \hat{\mathbf{B}}^\top\hat{\mathbf{B}} = \mathbf{B}^\top\mathbf{B} + \mathbf{B}^\top\mathbf{W}_\beta + \mathbf{W}_\beta^\top\mathbf{B} + \mathbf{W}_\beta^\top\mathbf{W}_\beta \\ &= \mathbf{J}_1 + \mathbf{J}_2 + \mathbf{J}_3 + \mathbf{J}_4. \end{aligned} \tag{25}$$

As for $\mathbf{J}_1$,

$$\mathbf{J}_1 = \sum_{j=1}^{m} \boldsymbol{\beta}_j\boldsymbol{\beta}_j^\top \xrightarrow{P} \boldsymbol{\Psi}_{\beta\beta}. \tag{26}$$

As for $\mathbf{J}_2$,

$$\begin{aligned} \|\sqrt{n}_{\min}\mathbf{J}_2\|_2 &= \left\| \frac{1}{\sqrt{m}}\sum_{j=1}^{m}(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})(\sqrt{m}\boldsymbol{\beta}_j)^\top \right\|_2 \\ &\le \sqrt{\left\| \frac{1}{m}\sum_{j=1}^{m}(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})^\top \right\|_2} \times \sqrt{\left\| \frac{1}{m}\sum_{j=1}^{m}(\sqrt{m}\boldsymbol{\beta}_j)(\sqrt{m}\boldsymbol{\beta}_j)^\top \right\|_2} \\ &\le \lambda_{\max}^{\frac{1}{2}}(\boldsymbol{\Psi}_{W_\beta W_\beta}) \times \lambda_{\max}^{\frac{1}{2}}(\boldsymbol{\Psi}_{\beta\beta}), \end{aligned} \tag{27}$$

which means

$$\|\mathbf{J}_2\|_2 = O_P(1/\sqrt{n}_{\min}). \tag{28}$$

As for $\mathbf{J}_3$, it has the same order as $\mathbf{J}_2$. As for $\mathbf{J}_4$,

$$\frac{n_{\min}}{m}\mathbf{J}_4 = \frac{1}{m}\sum_{j=1}^{m}(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})^\top \xrightarrow{P} \boldsymbol{\Psi}_{W_\beta W_\beta} \tag{29}$$

Hence:

(1) If $m/n_{\min} \to 0$,

$$\|\mathbf{J}_4\|_2 \le \lambda_{\max}(\mathbf{\Psi}_{W_\beta W_\beta}) \times \frac{m}{n_{\min}} \to 0. \tag{30}$$

Therefore,

$$m\mathbf{H}_{\text{IVW}} \xrightarrow{P} \mathbf{\Psi}_{\beta\beta}. \tag{31}$$

(2) If $m/n_{\min} \to c_0 \in (0, \infty)$, then

$$\mathbf{J}_4 = \frac{m}{n_{\min}} \times \frac{1}{m}\sum_{j=1}^{m}(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})^\top \xrightarrow{P} c_0\mathbf{\Psi}_{W_\beta W_\beta}. \tag{32}$$

Therefore,

$$m\mathbf{H}_{\text{IVW}} \xrightarrow{P} \mathbf{\Psi}_{\beta\beta} + c_0\mathbf{\Psi}_{W_\beta W_\beta}. \tag{33}$$

(3) If $m/n_{\min} \to \infty$ and $m/n_{\min}^{1+\tau} \to c_0 \in (0, +\infty)$ with certain constant $\tau > 0$, then

$$\frac{1}{n_{\min}^{\tau}}\mathbf{J}_4 = \frac{m}{n_{\min}^{1+\tau}} \times \frac{1}{m}\sum_{j=1}^{m}(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j})^\top \xrightarrow{P} c_0\mathbf{\Psi}_{W_\beta W_\beta}. \tag{34}$$

Therefore,

$$\frac{m}{n_{\min}^{\tau}}\mathbf{H}_{\text{IVW}} = c_0 n_{\min}\mathbf{H}_{\text{IVW}} \xrightarrow{P} c_0\mathbf{\Psi}_{W_\beta W_\beta}. \tag{35}$$

We then work with $\boldsymbol{S}_{\text{IVW}}(\theta)$:

$$\begin{aligned} m\boldsymbol{S}_{\text{IVW}}(\theta) &= -\mathbf{B}^\top\boldsymbol{w}_\alpha - \mathbf{W}_\beta^\top\boldsymbol{w}_\alpha + \mathbf{B}^\top\mathbf{W}_\beta\boldsymbol{\theta} + \mathbf{W}_\beta^\top\mathbf{W}_\beta\boldsymbol{\theta} \\ &= \boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3 + \boldsymbol{K}_4. \end{aligned} \tag{36}$$

As for $\boldsymbol{K}_1 + \boldsymbol{K}_3$,

$$\sqrt{n}_{\min}(\boldsymbol{K}_1 + \boldsymbol{K}_3) = \frac{1}{\sqrt{m}}\sum_{j=1}^{m}(-\sqrt{n}_{\min}w_{\alpha_j} + \sqrt{n}_{\min}\boldsymbol{w}_{\beta_j}^\top\boldsymbol{\theta})(\sqrt{m}\boldsymbol{\beta}_j) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta\mathbf{\Psi}_{\beta\beta}), \tag{37}$$

where

$$\psi_\theta = \psi_{w_\alpha w_\alpha} + \boldsymbol{\theta}^\top\mathbf{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta} - 2\boldsymbol{\theta}^\top\boldsymbol{\psi}_{W_\beta w_\alpha}. \tag{38}$$

As for $\boldsymbol{K}_2$,

$$\frac{n_{\min}}{m}\boldsymbol{K}_2 = -\frac{1}{m}\sum_{j=1}^{m}(\sqrt{n}_{\min}w_{\alpha_j})(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j}) \xrightarrow{P} -\boldsymbol{\psi}_{W_\beta w_\alpha}. \tag{39}$$

As for $\boldsymbol{K}_4$,

$$\frac{n_{\min}}{m}\boldsymbol{K}_4 = \left(\frac{1}{m}\sum_{j=1}^{m}(\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j}\sqrt{n}_{\min}\boldsymbol{w}_{\beta_j}\right)\boldsymbol{\theta} \xrightarrow{P} \mathbf{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta}, \tag{40}$$

Jointing these results, we summary the asymptotic behavior of $\hat{\boldsymbol{\theta}}_{\text{IVW}}$:

(1) If $m/\sqrt{n}_{\min} \to 0$, then

$$\sqrt{n}_{\min}||\boldsymbol{K}_2 + \boldsymbol{K}_4|| = O_P\left(\frac{m}{\sqrt{n}_{\min}}\right) = o_P(1). \tag{41}$$

Therefore,

$$\sqrt{n}_{\min} \times m\boldsymbol{S}_{\text{IVW}}(\boldsymbol{\theta}) = \sqrt{n}_{\min}(\boldsymbol{K}_1 + \boldsymbol{K}_3) + o_P(1) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \boldsymbol{\Psi}_{\beta\beta}). \tag{42}$$

Note that when $m/n_{\min} \to 0$, $m\mathbf{H}_{\text{IVW}} \xrightarrow{P} \boldsymbol{\Psi}_{\beta\beta}$. Therefore,

$$\sqrt{n}_{\min}(\hat{\boldsymbol{\theta}}_{\text{IVW}} - \boldsymbol{\theta}) = -\sqrt{n}_{\min}(m\mathbf{H}_{\text{IVW}})^{-1}(m\boldsymbol{S}_{\text{IVW}}(\theta)) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \boldsymbol{\Psi}_{\beta\beta}^{-1}), \tag{43}$$

(2) If $m/\sqrt{n}_{\min} \to c_0$, then

$$\sqrt{n}_{\min}(\boldsymbol{K}_2 + \boldsymbol{K}_4) \to -c_0 \boldsymbol{\psi}_{W_\beta w_\alpha} + c_0 \boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta}, \tag{44}$$

and hence

$$\sqrt{n}_{\min} \times m\boldsymbol{S}_{\text{IVW}}(\boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(-c_0(\boldsymbol{\psi}_{W_\beta w_\alpha} + \boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta}), \psi_\theta \boldsymbol{\Psi}_{\beta\beta}). \tag{45}$$

Note that when $m/n_{\min} \to 0$, $m\mathbf{H}_{\text{IVW}} \xrightarrow{P} \boldsymbol{\Psi}_{\beta\beta}$. Therefore,

$$\begin{aligned}\sqrt{n}_{\min}(\hat{\boldsymbol{\theta}}_{\text{IVW}} - \boldsymbol{\theta}) &= -\sqrt{n}_{\min}(m\mathbf{H}_{\text{IVW}})^{-1}(m\boldsymbol{S}_{\text{IVW}}(\theta)) \\ &\xrightarrow{D} \mathcal{N}(c_0 \boldsymbol{\Psi}_{\beta\beta}^{-1}(\boldsymbol{\psi}_{W_\beta w_\alpha} - \boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta}), \psi_\theta \boldsymbol{\Psi}_{\beta\beta}^{-1}).\end{aligned} \tag{46}$$

(3) If $m/\sqrt{n}_{\min} \to \infty$ and $m/n_{\min} \to c_0$, then $||\boldsymbol{K}_1 + \boldsymbol{K}_3||_2 = O_P(1/\sqrt{n}_{\min})$,

$$\boldsymbol{K}_2 + \boldsymbol{K}_4 \xrightarrow{P} -c_0 \boldsymbol{\psi}_{W_\beta w_\alpha} + c_0 \boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta}, \tag{47}$$

and

$$m\mathbf{H}_{\text{IVW}} \xrightarrow{P} \boldsymbol{\Psi}_{\beta\beta} + c_0 \boldsymbol{\Psi}_{W_\beta W_\beta}. \tag{48}$$

Hence,

$$\hat{\boldsymbol{\theta}}_{\text{IVW}} - \boldsymbol{\theta} \xrightarrow{P} c_0(\boldsymbol{\Psi}_{\beta\beta} + c_0 \boldsymbol{\Psi}_{W_\beta W_\beta})^{-1}(\boldsymbol{\psi}_{W_\beta w_\alpha} - \boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta}). \tag{49}$$

Note that if $c_0 = 0$, then (iii) in Theorem 1.2 holds.

(4) If $m/n_{\min} \to \infty$ and $m/n_{\min}^{1+\tau} \to c_0$, then

$$\frac{1}{n_{\min}^\tau}(\boldsymbol{K}_2 + \boldsymbol{K}_4) \xrightarrow{P} -c_0 \boldsymbol{\psi}_{W_\beta w_\alpha} + c_0 \boldsymbol{\Psi}_{W_\beta W_\beta}\boldsymbol{\theta} \tag{50}$$

and

$$\frac{m}{n_{\min}^\tau}\mathbf{H}_{\text{IVW}} \xrightarrow{P} c_0 \boldsymbol{\Psi}_{W_\beta W_\beta}. \tag{51}$$

Therefore,

$$\hat{\boldsymbol{\theta}}_{\text{IVW}} \xrightarrow{P} \boldsymbol{\Psi}_{W_\beta W_\beta}^{-1} \boldsymbol{\psi}_{W_\beta w_\alpha}. \tag{52}$$

Now Theorem 1.2 is proved. □

1.7 Proofs of Theorems for MRBEE

Proofs of Theorem 1.3. Note that

$$\mathbf{0} = \boldsymbol{S}_{\text{BEE}}(\hat{\boldsymbol{\theta}}_{\text{BEE}}) = \boldsymbol{S}_{\text{BEE}}(\boldsymbol{\theta}) + \mathbf{H}_{\text{BEE}}(\hat{\boldsymbol{\theta}}_{\text{BEE}} - \boldsymbol{\theta}), \tag{53}$$

where

$$\boldsymbol{S}_{\text{BEE}}(\boldsymbol{\theta}) = -\frac{1}{m}\hat{\mathbf{B}}^\top(\hat{\boldsymbol{\alpha}} - \hat{\mathbf{B}}\boldsymbol{\theta}) - \boldsymbol{\Sigma}_{W_\beta W_\beta}\boldsymbol{\theta} + \boldsymbol{\sigma}_{W_\beta w_\alpha}, \tag{54}$$

and

$$\mathbf{H}_{\text{BEE}} = \frac{1}{m}\hat{\mathbf{B}}^\top\hat{\mathbf{B}} - \boldsymbol{\Sigma}_{W_\beta W_\beta}. \tag{55}$$

As for $\boldsymbol{S}_{\text{BEE}}(\boldsymbol{\theta})$,

$$\begin{aligned} m\boldsymbol{S}_{\text{BEE}}(\boldsymbol{\theta}) &= -(\mathbf{B} + \mathbf{W}_\beta)^\top(\boldsymbol{\alpha} + \boldsymbol{w}_\alpha - \mathbf{B}\boldsymbol{\theta} - \mathbf{W}_\beta\boldsymbol{\theta}) - m\boldsymbol{\Sigma}_{W_\beta W_\beta} + m\boldsymbol{\sigma}_{W_\beta w_\alpha} \\ &= -\left\{\mathbf{B}^\top(\boldsymbol{w}_\alpha - \mathbf{W}_\beta\boldsymbol{\theta})\right\} + \left\{\left(\mathbf{W}_\beta^\top\mathbf{W}_\beta - m\boldsymbol{\Sigma}_{W_\beta W_\beta}\right)\boldsymbol{\theta}\right\} - \left\{\mathbf{W}_\beta^\top\boldsymbol{w}_\alpha - m\boldsymbol{\sigma}_{W_\beta w_\alpha}\right\} \\ &= \boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3. \end{aligned} \tag{56}$$

Here, we define a new vector $\boldsymbol{\vartheta} = (\boldsymbol{\theta}^\top, 1)^\top$, an alternative vector

$$\boldsymbol{\zeta}_j = \left(\frac{n_{\min}}{n_1}\xi_j^{[1]}, \frac{n_{\min}}{n_2}\xi_j^{[2]}, \ldots, \frac{n_{\min}}{n_p}\xi_j^{[p]}, \frac{n_{\min}}{n_0}\xi_j^{[0]}\right)^\top,$$

where

$$\xi_j^{[s]} = \frac{1}{\sqrt{n_s}}\sum_{i=1}^{n_s} g_{ij}^{[s]}x_{is}^{[s]}, \quad s = 0, 1, \ldots, p,$$

and a new covariance matrix

$$\text{cov}(\boldsymbol{\zeta}_j) = \boldsymbol{\Psi}_{W_\beta \times w_\alpha} = \begin{pmatrix} \boldsymbol{\Psi}_{W_\beta W_\beta} & \boldsymbol{\psi}_{W_\beta w_\alpha} \\ \boldsymbol{\psi}_{W_\beta w_\alpha}^\top & \psi_{w_\alpha w_\alpha} \end{pmatrix}. \tag{57}$$

As for $\boldsymbol{K}_1$, it can be rewritten as

$$\begin{aligned} \sqrt{n}_{\min}\boldsymbol{K}_1 = -\sum_{j=1}^m \sqrt{n}_{\min}(w_{\alpha_j} - \boldsymbol{w}_{\beta_j}^\top\boldsymbol{\theta})\boldsymbol{\beta}_j = \frac{1}{\sqrt{m}}\sum_{j=1}^m(\sqrt{n}_{\min}\boldsymbol{\zeta}_j^\top\boldsymbol{\vartheta})(\sqrt{m}\boldsymbol{\beta}_j) \\ \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta\boldsymbol{\Psi}_{\beta\beta}), \end{aligned} \tag{58}$$

where ψ_θ defined in (38) can be rewritten as

$$\psi_\theta = \boldsymbol{\vartheta}^\top\boldsymbol{\Psi}_{W_\beta \times w_\alpha}\boldsymbol{\vartheta}. \tag{59}$$

As for $\boldsymbol{K}_2 + \boldsymbol{K}_3$, it can be rewritten as

$$\begin{aligned} \boldsymbol{K}_2 + \boldsymbol{K}_3 &= \mathbf{I}_{p+1}^{1:p}\begin{pmatrix} \mathbf{W}_\beta^\top\mathbf{W}_\beta - m\boldsymbol{\Sigma}_{W_\beta W_\beta} & \mathbf{W}_\beta^\top\boldsymbol{w}_\alpha - m\boldsymbol{\sigma}_{W_\beta w_\alpha} \\ \boldsymbol{w}_\alpha^\top\mathbf{W}_\beta - m\boldsymbol{\sigma}_{W_\beta w_\alpha}^\top & \boldsymbol{w}_\alpha^\top\boldsymbol{w}_\alpha - m\sigma_{w_\alpha w_\alpha} \end{pmatrix}\begin{pmatrix} \boldsymbol{\theta} \\ -1 \end{pmatrix} \\ &= \frac{\sqrt{m}}{n_{\min}}\mathbf{I}_{p+1}^{1:p}\left(\frac{1}{\sqrt{m}}\sum_{j=1}^m \boldsymbol{\zeta}_j\boldsymbol{\zeta}_j^\top - \boldsymbol{\Psi}_{W_\beta \times w_\alpha}\right)\boldsymbol{\vartheta} \\ &= \frac{\sqrt{m}}{n_{\min}}\mathbf{I}_{p+1}^{1:p}\mathbf{K}_4\boldsymbol{\vartheta}, \end{aligned} \tag{60}$$

where $\mathbf{I}_{p+1}^{1:p}$ is a $(p \times (p+1))$ matrix consisting of the first p row of $\mathbf{I}_{p+1}$ and

$$\mathbf{K}_4 = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \boldsymbol{\zeta}_j \boldsymbol{\zeta}_j^\top - \boldsymbol{\Psi}_{W_\beta \times w_\alpha}. \tag{61}$$

According to Lemma 1.6,

$$\text{vec}(\mathbf{K}_4) \xrightarrow{D} \mathcal{N}\bigg(\mathbf{0}, (\mathbf{I}_{(p+1)^2} + \mathbf{K}_{(p+1)^2})(\boldsymbol{\Psi}_{W_\beta \times w_\alpha} \otimes \boldsymbol{\Psi}_{W_\beta \times w_\alpha})\bigg). \tag{62}$$

As a result,

$$\frac{n_{\min}}{\sqrt{m}}(\boldsymbol{K}_2 + \boldsymbol{K}_3) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{BC}}) \tag{63}$$

where

$$\boldsymbol{\Sigma}_{\text{BC}} = \underbrace{\left[\boldsymbol{\vartheta}^\top \otimes \mathbf{I}_{p+1}^{1:p}\right]}_{p\times(p+1)^2} \underbrace{\left[(\mathbf{I}_{(p+1)^2} + \mathbf{K}_{(p+1)^2})(\boldsymbol{\Psi}_{W_\beta \times w_\alpha} \otimes \boldsymbol{\Psi}_{W_\beta \times w_\alpha})\right]}_{(p+1)^2\times(p+1)^2} \underbrace{\left[\boldsymbol{\vartheta}^\top \otimes \mathbf{I}_{p+1}^{1:p}\right]^\top}_{(p+1)^2\times p}. \tag{64}$$

Now we show $\boldsymbol{K}_1$ and $\boldsymbol{K}_2 + \boldsymbol{K}_3$ are uncorrelated. Note that $\boldsymbol{\beta}_j$ is independent of $\boldsymbol{w}_{\beta_j}$ and w_{α_j}, and hence $\boldsymbol{K}_1$ and $\boldsymbol{K}_2 + \boldsymbol{K}_3$ are uncorrelated. So far, we can obtain:

(1) If $m/n_{\min} \to 0$,

$$\sqrt{n}_{\min} \times m\boldsymbol{S}_{\text{BEE}}(\boldsymbol{\theta}) = \sqrt{n}_{\min}\boldsymbol{K}_1 + o_P(1) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \boldsymbol{\Psi}_{\beta\beta}). \tag{65}$$

(2) If $m/n_{\min} \to c_0$,

$$\sqrt{n}_{\min} \times m\boldsymbol{S}_{\text{BEE}}(\boldsymbol{\theta}) = \sqrt{n}_{\min}\boldsymbol{K}_1 + \sqrt{n}_{\min}(\boldsymbol{K}_2 + \boldsymbol{K}_3) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \boldsymbol{\Psi}_{\beta\beta} + c_0\boldsymbol{\Sigma}_{\text{BC}}). \tag{66}$$

(3) If $m/n_{\min} \to \infty$ and $\sqrt{m}/n_{\min} \to 0$,

$$\frac{n_{\min}}{\sqrt{m}} \times m\boldsymbol{S}_{\text{BEE}}(\boldsymbol{\theta}) = \frac{n_{\min}}{\sqrt{m}}(\boldsymbol{K}_2 + \boldsymbol{K}_3) + \frac{n_{\min}}{\sqrt{m}}\boldsymbol{K}_1 \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{BC}}), \tag{67}$$

where

$$\frac{n_{\min}}{\sqrt{m}}\boldsymbol{K}_1 = \sqrt{\frac{n_{\min}}{m}} \times \sqrt{n}_{\min}\boldsymbol{K}_1 = O_P\left(\sqrt{\frac{n_{\min}}{m}}\right) = o_P(1). \tag{68}$$

Now we move to $\mathbf{H}_{\text{BEE}}$:

$$\begin{aligned} m\mathbf{H}_{\text{BEE}} &= \mathbf{B}^\top\mathbf{B} + \left(\mathbf{W}_\beta^\top\mathbf{W}_\beta - m\boldsymbol{\Sigma}_{W_\beta W_\beta}\right) + \mathbf{B}^\top\mathbf{W}_\beta + \mathbf{W}_\beta^\top\mathbf{B} \\ &= \mathbf{J}_1 + \mathbf{J}_2 + \mathbf{J}_3 + \mathbf{J}_4. \end{aligned} \tag{69}$$

As for $\mathbf{J}_1 = \mathbf{B}^\top\mathbf{B}$, we have

$$\begin{aligned} ||\mathbf{J}_1 - \boldsymbol{\Psi}_{\beta\beta}||_2 &= \left\|\frac{1}{m}\sum_{j=1}^{m} \sqrt{m}\boldsymbol{\beta}_j\sqrt{m}\boldsymbol{\beta}_j^\top - \boldsymbol{\Psi}_{\beta\beta}\right\|_2 \\ &= O_P\left(\frac{1}{\sqrt{m}}\right). \end{aligned} \tag{70}$$

As for $\mathbf{J}_2 = \mathbf{W}_\beta^\top \mathbf{W}_\beta - m\mathbf{\Sigma}_{W_\beta W_\beta}$, we have

$$\mathbf{J}_2 = \sum_{j=1}^{m} \left(\boldsymbol{w}_{\beta_j} \boldsymbol{w}_{\beta_j}^\top - \mathbf{\Sigma}_{W_\beta W_\beta} \right) = \frac{\sqrt{m}}{n_{\min}} \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \left(\boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top - \mathbf{\Psi}_{W_\beta W_\beta} \right). \tag{71}$$

As a result,

$$\frac{n_{\min}}{\sqrt{m}} \text{vec}(\mathbf{J}_2) \xrightarrow{D} \mathcal{N}(\mathbf{0}, (\mathbf{I}_{p^2} + \mathbf{K}_{p^2})(\mathbf{\Psi}_{W_\beta W_\beta} \otimes \mathbf{\Psi}_{W_\beta W_\beta})), \tag{72}$$

which means $||\mathbf{J}_2|| = O_P(\sqrt{m}/n_{\min})$. As for $\mathbf{J}_3 = \mathbf{B}^\top \mathbf{W}_\beta$,

$$\begin{aligned}
\sqrt{n}_{\min} ||\mathbf{J}_3||_2 &= \left\| \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \sqrt{m} \boldsymbol{\beta}_j \sqrt{n}_{\min} \boldsymbol{\omega}_{\beta_j}^\top \right\|_2 \\
&\leq \sqrt{\left\| \frac{1}{m} \sum_{j=1}^{m} \sqrt{m} \boldsymbol{\beta}_j \sqrt{m} \boldsymbol{\beta}_j^\top \right\|_2} \sqrt{\left\| \frac{1}{m} \sum_{j=1}^{m} \sqrt{n}_{\min} \boldsymbol{\omega}_{\beta_j} \sqrt{n}_{\min} \boldsymbol{\omega}_{\beta_j}^\top \right\|_2} \\
&\leq \lambda_{\max}^{\frac{1}{2}}(\mathbf{\Psi}_{\beta\beta}) \times \lambda_{\max}^{\frac{1}{2}}(\mathbf{\Psi}_{W_\beta W_\beta}),
\end{aligned} \tag{73}$$

which means

$$||\mathbf{J}_3||_2 = O_P\left(\frac{1}{\sqrt{n}_{\min}} \right) \tag{74}$$

As for $\mathbf{J}_4$, it is easy to see $||\mathbf{J}_4||_2^2 = ||\mathbf{J}_3||_2^2$. Hence, for all three scenarios in Theorem 1.3,

$$||m\mathbf{H}_{\text{BEE}} - \mathbf{\Psi}_{\beta\beta}||_2 = O_P\left\{ \max\left(\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{n}_{\min}}, \frac{\sqrt{m}}{n_{\min}} \right) \right\}. \tag{75}$$

And hence, according to the Slutsky's theorem,

(1) If $m/n_{\min} \to 0$,

$$\sqrt{n}_{\min}(\hat{\boldsymbol{\theta}}_{\text{BEE}} - \boldsymbol{\theta}) = -\sqrt{n}_{\min} \mathbf{\Psi}_{\beta\beta}^{-1} \boldsymbol{K}_1 \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \mathbf{\Psi}_{\beta\beta}^{-1}). \tag{76}$$

(2) If $m/n_{\min} \to c_0$,

$$\sqrt{n}_{\min}(\hat{\boldsymbol{\theta}}_{\text{BEE}} - \boldsymbol{\theta}) = -\sqrt{n}_{\min} \mathbf{\Psi}_{\beta\beta}^{-1} (\boldsymbol{K}_1 + \boldsymbol{K}_2 + \boldsymbol{K}_3) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \psi_\theta \mathbf{\Psi}_{\beta\beta}^{-1} + c_0 \mathbf{\Psi}_{\beta\beta}^{-1} \mathbf{\Psi}_{\text{BC}} \mathbf{\Psi}_{\beta\beta}^{-1}). \tag{77}$$

(2) If $m/n_{\min} \to \infty$ and $m/n_{\min}^2 \to 0$,

$$\sqrt{n_{\min}^2/m}(\hat{\boldsymbol{\theta}}_{\text{BEE}} - \boldsymbol{\theta}) = -\frac{n_{\min}}{\sqrt{m}} \mathbf{\Psi}_{\beta\beta}^{-1} (\boldsymbol{K}_2 + \boldsymbol{K}_3) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{\Psi}_{\beta\beta}^{-1} \mathbf{\Psi}_{\text{BC}} \mathbf{\Psi}_{\beta\beta}^{-1}). \tag{78}$$

Thus, Theorem 1.3 is proved. □

Proof of Theorem 1.4. Similar to $\xi_j^{[s]}$, we define $\eta_j^{\{s\}}$ as

$$\eta_j^{\{s\}} = \frac{\boldsymbol{g}_j^{\{s\}\top}\boldsymbol{x}^{[s]}}{\sqrt{n_s}} = \frac{1}{\sqrt{n_s}}\sum_{i=1}^{n_s} g_{ij}^{\{s\}} x_i^{[s]}. \tag{79}$$

By using similar deduction as which in the proof of Theorem 1,

$$\eta_j^{\{s\}} \xrightarrow{D} \mathcal{N}(0, \sigma_{x_s x_s}) \tag{80}$$

and

$$\mathrm{cov}(\eta_j^{\{s\}}, \eta_j^{\{k\}}) = \frac{n_{sk}}{\sqrt{n_s n_k}}\sigma_{x_s x_k}. \tag{81}$$

Denote $\boldsymbol{\eta}_j = (\eta_j^{\{1\}}, \ldots, \eta_j^{\{p\}}, \eta_j^{\{0\}})$ where $\eta_j^{\{0\}}$ represents $\frac{1}{\sqrt{n_0}}\boldsymbol{g}_j^{\{s\}\top}\boldsymbol{y}^{[0]}$. Then we have

$$\mathrm{cov}(\boldsymbol{\eta}_j) = \mathbf{D}_\eta^{-1}\boldsymbol{\Sigma}_{W_\beta \times w_\alpha}\mathbf{D}_\eta^{-1}, \tag{82}$$

where

$$\mathbf{D}_\eta = \mathrm{diag}\left(\frac{1}{\sqrt{n_1}}, \ldots, \frac{1}{\sqrt{n_p}}, \frac{1}{\sqrt{n_0}}\right). \tag{83}$$

By using Lemma 1.4,

$$\left\|\frac{1}{M}\sum_{j=1}^{M}\boldsymbol{\eta}_j\boldsymbol{\eta}_j^\top - \mathrm{cov}(\boldsymbol{\eta}_j)\right\|_2 = O_P\left(\frac{1}{\sqrt{M}}\right), \tag{84}$$

and hence

$$\begin{aligned}\|\boldsymbol{\Sigma}_{W_\beta \times w_\alpha}^{-\frac{1}{2}}\hat{\boldsymbol{\Sigma}}_{W_\beta \times w_\alpha}\boldsymbol{\Sigma}_{W_\beta \times w_\alpha}^{\frac{1}{2}} - \mathbf{I}_{p+1}\|_2 &\le \lambda_{\min}^{-1}(\mathrm{cov}(\boldsymbol{\eta}_j))\left\|\frac{1}{M}\sum_{j=1}^{M}\boldsymbol{\eta}_j\boldsymbol{\eta}_j^\top - \mathrm{cov}(\boldsymbol{\eta}_j)\right\|_2 \\ &= O_P\left(\frac{1}{\sqrt{M}}\right). \end{aligned} \tag{85}$$

Thus, Theorem 1.4 is proved. □

Proof of Theorem 1.5. Note that

$$\begin{aligned}\boldsymbol{S}_j(\boldsymbol{\theta}) &= -(\hat{\alpha}_j - \boldsymbol{\theta}^\top \hat{\boldsymbol{\beta}}_j)\hat{\boldsymbol{\beta}}_j - \boldsymbol{\Sigma}_{W_\beta W_\beta}\boldsymbol{\theta} + \boldsymbol{\sigma}_{W_\beta w_\alpha} \\ &= (w_{\alpha_j} - \boldsymbol{\theta}^\top \boldsymbol{w}_{\beta_j})\boldsymbol{\beta}_j + \left\{(w_{\alpha_j} - \boldsymbol{\theta}^\top \boldsymbol{w}_{\beta_j})\boldsymbol{w}_{\beta_j} - \boldsymbol{\Sigma}_{W_\beta W_\beta}\boldsymbol{\theta} + \boldsymbol{\sigma}_{W_\beta w_\alpha}\right\} \\ &= \boldsymbol{J}_{1j} + \boldsymbol{J}_{2j}. \end{aligned} \tag{86}$$

Note that both $\boldsymbol{J}_{1j}$ and $\boldsymbol{J}_{2j}$ are sub-exponential variables with zero mean and covariance matrix

$$\mathrm{cov}(\boldsymbol{J}_{1j}) = \frac{1}{mn_{\min}}\psi_\theta \boldsymbol{\Psi}_{\beta\beta}, \quad \mathrm{cov}(\boldsymbol{J}_{2j}) = \frac{1}{n_{\min}^2}\boldsymbol{\Sigma}_{\mathrm{BC}}. \tag{87}$$

Therefore, we obtain

$$\mathrm{cov}(\boldsymbol{S}_j(\boldsymbol{\theta})) = \boldsymbol{\Sigma}_S = \begin{cases} \frac{1}{mn_{\min}}\psi_\theta \boldsymbol{\Psi}_{\beta\beta}, & \text{if } m/n_{\min} \to 0, \\ \frac{1}{mn_{\min}}\psi_\theta \boldsymbol{\Psi}_{\beta\beta} + \frac{c_0}{mn_{\min}}\boldsymbol{\Sigma}_{\mathrm{BC}}, & \text{if } m/n_{\min} \to c_0, \\ \frac{1}{n_{\min}^2}\boldsymbol{\Sigma}_{\mathrm{BC}},, & \text{if } m/n_{\min} \to \infty \text{ and } \sqrt{m}/n_{\min} \to 0. \end{cases} \tag{88}$$

Then by using Lemma 1.5,

$$\left\| \frac{1}{m}\sum_{j=1}^m \boldsymbol{S}_j(\boldsymbol{\theta})\boldsymbol{S}_j(\boldsymbol{\theta})^\top - \boldsymbol{\Sigma}_S \right\|_2 = O_P\left(\sqrt{\frac{\log m}{m}}\right)||\boldsymbol{\Sigma}_S||_2. \tag{89}$$

By using the Slutsky's theorem,

$$\left\| \frac{1}{m}\sum_{j=1}^m \hat{\boldsymbol{S}}_j(\hat{\boldsymbol{\theta}}_{\mathrm{BEE}})\hat{\boldsymbol{S}}_j(\hat{\boldsymbol{\theta}}_{\mathrm{BEE}})^\top - \boldsymbol{\Sigma}_S \right\|_2 = O_P\left(\sqrt{\frac{\log m}{m}}\right)||\boldsymbol{\Sigma}_S||_2. \tag{90}$$

where

$$\hat{\boldsymbol{S}}_j(\hat{\boldsymbol{\theta}}_{\mathrm{BEE}}) = -(\hat{\boldsymbol{\theta}}_{\mathrm{BEE}}^\top \hat{\boldsymbol{\beta}}_j - \hat{\alpha}_j)\hat{\boldsymbol{\beta}}_j + \hat{\boldsymbol{\Sigma}}_{W_\beta W_\beta}\hat{\boldsymbol{\theta}}_{\mathrm{BEE}} - \hat{\boldsymbol{\sigma}}_{W_\beta w_\alpha} \tag{91}$$

On the other hand, according to the proof of Theorem 1.3,

$$\|m\hat{\mathbf{F}}_{\mathrm{BEE}} - \boldsymbol{\Psi}_{\beta\beta}\|_2 = O_P\left\{\max\left(\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{n_{\min}}}, \frac{\sqrt{m}}{n_{\min}}\right)\right\}. \tag{92}$$

Note that Bickel and Levina (2008, A22(p223)) illustrates

$$\|\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3 - \mathbf{B}_1\mathbf{B}_2\mathbf{B}_3\|_2 = O_P\left\{\max\left(||\mathbf{A}_1 - \mathbf{B}_1||_2, ||\mathbf{A}_2 - \mathbf{B}_2||_2, ||\mathbf{A}_3 - \mathbf{B}_3||_2\right)\right\}, \tag{93}$$

where $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$ are six matrices with non-diverging maximum singular values. Hence,

$$\begin{aligned}||\hat{\boldsymbol{\Sigma}}_{\mathrm{BEE}}(\hat{\boldsymbol{\theta}}_{\mathrm{BEE}}) - \boldsymbol{\Sigma}_{\mathrm{BEE}}(\boldsymbol{\theta})||_2 &= \left\|(m\hat{\mathbf{F}}_{\mathrm{BEE}})^{-1}\left(\sum_{j=1}^m \hat{\boldsymbol{S}}_j(\hat{\boldsymbol{\theta}}_{\mathrm{BEE}})\hat{\boldsymbol{S}}_j(\hat{\boldsymbol{\theta}}_{\mathrm{BEE}})^\top\right)(m\hat{\mathbf{F}}_{\mathrm{BEE}})^{-1} - m\boldsymbol{\Psi}_{\beta\beta}^{-1}\boldsymbol{\Sigma}_S\boldsymbol{\Psi}_{\beta\beta}^{-1}\right\|_2 \\ &= O_P\left\{\max\left(\sqrt{\frac{\log m}{m}}, \frac{1}{\sqrt{n_{\min}}}, \frac{\sqrt{m}}{n_{\min}}\right)\right\}||m\boldsymbol{\Sigma}_S||_2, \end{aligned} \tag{94}$$

and consequently

$$||\boldsymbol{\Sigma}_{\mathrm{BEE}}^{-\frac{1}{2}}(\boldsymbol{\theta})\hat{\boldsymbol{\Sigma}}_{\mathrm{BEE}}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\mathrm{BEE}}^{-\frac{1}{2}}(\boldsymbol{\theta}) - \mathbf{I}_p||_2 = O_P\left\{\max\left(\sqrt{\frac{\log m}{m}}, \frac{1}{\sqrt{n_{\min}}}, \frac{\sqrt{m}}{n_{\min}}\right)\right\}. \tag{95}$$

Thus, Theorem 1.5 is proved. □

Proof of Theorem 1.6. Note that $||\hat{\boldsymbol{\theta}}_{\text{BEE}} - \boldsymbol{\theta}||_2 = O_P(n_{\min}^{-\frac{1}{2}})$ and hence $\hat{\alpha}_j - \hat{\boldsymbol{\beta}}_j^\top \hat{\boldsymbol{\theta}}_{\text{BEE}}$ and $\hat{\alpha}_j - \hat{\boldsymbol{\beta}}_j^\top \boldsymbol{\theta}$ have the same distribution. For $j \in \mathcal{O}^c$,

$$\begin{aligned}\hat{\gamma}_j = \varepsilon_j = \hat{\alpha}_j - \hat{\boldsymbol{\beta}}_j^\top \hat{\boldsymbol{\theta}}_{\text{BEE}} &= w_{\alpha_j} - \boldsymbol{w}_{\beta_j}^\top \boldsymbol{\theta} + \boldsymbol{w}_{\beta_j}^\top (\hat{\boldsymbol{\theta}}_{\text{BEE}} - \boldsymbol{\theta}) \\ &\sim \mathcal{N}(0, \sigma_{\varepsilon\varepsilon}),\end{aligned} \tag{96}$$

where

$$\sigma_{\varepsilon\varepsilon} = \boldsymbol{\theta}^\top \boldsymbol{\Sigma}_{W_\beta w_\alpha} \boldsymbol{\theta} + \sigma_{\omega_\gamma \omega_\gamma} - 2\boldsymbol{\theta}^\top \boldsymbol{\sigma}_{W_\beta w_\alpha}. \tag{97}$$

As a result,

$$\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}} \sim \chi_1^2. \tag{98}$$

Denote $\kappa^* = F_{\chi_1^2}^{-1}(\kappa)$. Then by using Lemma A.1 of Huang et al. (2012),

$$\begin{aligned}\Pr\left(\max_{j\in\mathcal{O}^c} \frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}} \le \kappa^*\right) &= 1 - \Pr\left(\max_{j\in\mathcal{O}^c} \frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}} > \kappa^*\right) \\ &\ge 1 - (m - |\mathcal{O}|) \Pr\left(\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}} > \kappa^*\right) \\ &\ge 1 - m \Pr\left(\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}} > \kappa^*\right) \\ &\ge 1 - m \exp\left(-\frac{(\sqrt{2\kappa^* - 1} - 1)^2}{4}\right).\end{aligned} \tag{99}$$

By letting $\kappa^* = C_0 \log m$ with C_0 being a sufficiently large constant,

$$\begin{aligned}\Pr\left(\max_{j\in\mathcal{O}^c} \frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}} \le \kappa^*\right) &\ge 1 - \exp\left(\log m - \frac{2C_0 \log m - 2\sqrt{C_0 \log m - 1}}{4}\right) \\ &\ge 1 - \exp\left(-\frac{(2C_0 - 4)\log m - 2\sqrt{C_0 \log m - 1}}{4}\right) \to 1,\end{aligned} \tag{100}$$

if $m \to \infty$.

On the other hand, for $j \in \mathcal{O}$,

$$\hat{\gamma}_j = \gamma_j + \varepsilon_j, \tag{101}$$

and hence

$$\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}} \sim \chi_1^2\left(\frac{\gamma_j^2}{\sigma_{\varepsilon\varepsilon}}\right), \tag{102}$$

where $\chi_1^2(\lambda)$ refers to the noncentral chi-squared distribution with degree of freedom 1 and noncentrality parameter λ. Let $F_{\chi_1^2(\lambda)}(\cdot)$ be the CDF of this noncentral chi-squared distribution, which is indeed equal to

$$F_{\chi_1^2(\lambda)}(x) = 1 - \Big(Q(\sqrt{x} - \sqrt{\lambda}) + Q(\sqrt{x} + \sqrt{\lambda})\Big), \tag{103}$$

where $F_{\chi_1^2(\lambda)}(\cdot)$ be the CDF of $\chi_1^2(\lambda)$ and $Q(x)$ is the Gaussian Q-function, i.e., $Q(x) = 1 - \Phi(x)$ and $\Phi(x)$ is the CDF of standard normal distribution.

Note that there should exist a constant D_0 such that

$$\frac{\gamma_j^2}{\sigma_{\varepsilon\varepsilon}} \ge D_0 n_{\min} \tag{104}$$

where D_0 is a sufficient large constant. And

$$\begin{aligned}\Pr\left(\min_{j\in\mathcal{O}}\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}}\geq\kappa^*\right)&=1-\Pr\left(\min_{j\in\mathcal{O}^c}\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}}<\kappa^*\right)\\&\geq 1-\Pr\left(\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}}<\kappa^*\right),\quad j\text{ is arbitrary element in }\mathcal{O}.\end{aligned}\tag{105}$$

Hence,

$$\begin{aligned}\Pr\left(\min_{j\in\mathcal{O}}\frac{\hat{\gamma}_j^2}{\sigma_{\varepsilon\varepsilon}}\geq\kappa^*\right)&\geq Q(\sqrt{\kappa^*}-\sqrt{D_0 n_{\min}})+Q(\sqrt{\kappa^*}+\sqrt{D_0 n_{\min}})\\&\geq Q(\sqrt{C_0\log m}-\sqrt{D_0 n_{\min}})+Q(\sqrt{C_0\log m}+\sqrt{D_0 n_{\min}})\to 1\end{aligned}\tag{106}$$

if $m, n_{\min}\to\infty$. Thus, Theorem 1.6 is proved. □

References

Bickel, P.J. and Levina, E. (2008). Regularized estimation of large covariance matrices. The Annals of Statistics. **36**(1), pp. 199-227.

Fisher, R.A. (1919). The correlation between relatives on the supposition of Mendelian inheritance. Earth and Environmental Science Transactions of the Royal Society of Edinburgh. **52**(2), pp. 339-433.

Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L. and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics. **47**(3), pp. 291-295.

Fan, J. and Li R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association. **96**(456), pp. 1348-60.

Fan, J., Liao, Y. and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. The Annals of Statistics. **39**(6), pp. 3320–3356.

Huang, J., Breheny, P. and Ma, S. (2012). A selective review of group selection in high-dimensional models. Statistical Science, **27**(4), pp. 481-499.

Muirhead, R.J., (2009). Aspects of multivariate statistical theory. John Wiley & Sons.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. arXiv preprint. arXiv:1011.3027.

Vershynin, R. (2018). High-dimensional probability: An introduction with applications in data science. Cambridge university press.

Jankova, J and Van De Geer, S. (2018). Semiparametric efficiency bounds for high-dimensional models The Annals of Statistics, **46**(5), pp. 2336-2359.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), **57**(1), pp.289-300.

Lin, Z., Xue, H. and Pan, W. (2023). Robust multivariable Mendelian randomization based on constrained maximum likelihood. The American Journal of Human Genetics, **110**(4), pp.592-605.