Web Appendix for "Secondary analysis of case-control association studies: Insights on weighting-based inference motivate a new specification test"

Fan Li*, **Andrew S. Allen**

***fan.f.li@yale.edu**

WEB APPENDIX A: REGULARITY CONDITIONS

The following regularity assumptions are required for proving the proposition.

Assumption 1 (Retrospective sampling)**.** Suppose the disease prevalence, $\lambda = Pr(D = 1)$, is strictly between 0 and 1. The control subsample, $\boldsymbol{Z}_{1:n_0} = \{(Y_i, \boldsymbol{X}_i, D_i = 0)\}_{i=1}^{n_0}$, represents a random sample among the disease-free subpopulation, and the case subsample, $\boldsymbol{Z}_{(n_0+1):n} = \{(Y_i, \boldsymbol{X}_i, D_i = 1)\}_{i=n_0+1}^{n}$, is a random sample from the diseases subpopulation. Further, there exists positive constants ρ_0 and ρ_1 such that $\lim_{n\to\infty} n_0/n = \rho_0$, $\lim_{n\to\infty} n_1/n = \rho_1$ with $\rho_0 + \rho_1 = 1$ and $n_0 + n_1 = n$.

Assumption 2 (Disease model)**.** There exists a unique $q \times 1$ vector $\boldsymbol{\gamma}_0 \in \text{int}(\boldsymbol{\Theta}_{\boldsymbol{\gamma}})$ with $\boldsymbol{\Theta}_{\boldsymbol{\gamma}} \in \mathbb{R}^q$ and compact, and a positive constant ϵ such that $Pr(D = 1|Y, \boldsymbol{X}) = H(Y, \boldsymbol{X}, \boldsymbol{\gamma}_0)$; the function $H(Y, \boldsymbol{X}, \boldsymbol{\gamma}_0) > \epsilon$ for all $Y, \boldsymbol{X}$ such that $f(Y, \boldsymbol{X}|D = 1) > 0$ and $H(Y, \boldsymbol{X}, \boldsymbol{\gamma}_0) < 1 - \epsilon$ for all $Y, \boldsymbol{X}$ satisfying $f(Y, \boldsymbol{X}|D = 0) > 0$.

To present the subsequent conditions, we define a gradient matrix $\mathcal{J}$ and a covariance matrix (assumed positive definite) $\mathcal{V}$ as

$$\mathcal{J} = \sum_{d=0}^{1} \rho_d E\left[\frac{\partial \Upsilon(\boldsymbol{Z}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\Big| D = d\right] = \begin{pmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \mathbf{0} & \mathcal{J}_{22} \end{pmatrix},$$

$$\mathcal{V} = \sum_{d=0}^{1} \rho_d E\left[\Upsilon(\boldsymbol{Z}, \boldsymbol{\theta}_0)^{\otimes 2}|D = d\right] = \begin{pmatrix} \mathcal{V}_{11} & \mathcal{V}_{12} \\ \mathcal{V}_{21} & \mathcal{V}_{22} \end{pmatrix},$$

with the respective component matrices defined by,

$$\mathcal{J}_{11} = \sum_{d=0}^{1} \rho_d E\left[\frac{\partial \Psi(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\beta}^T}\Big| D = d\right] \qquad \mathcal{J}_{22} = \sum_{d=0}^{1} \rho_d E\left[\frac{\partial \Phi(\boldsymbol{Z}, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}^T}\Big| D = d\right]$$

$$\mathcal{J}_{12} = \sum_{d=0}^{1} \rho_d E\left[\frac{\partial \Psi(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}^T}\Big| D = d\right] \qquad \mathcal{V}_{11} = \sum_{d=0}^{1} \rho_d E\left[\Psi(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)^{\otimes 2}|D = d\right]$$

$$\mathcal{V}_{12} = \sum_{d=0}^{1} \rho_d E\left[\Psi(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\Phi(\boldsymbol{Z}, \boldsymbol{\gamma}_0)^T|D = d\right] = \mathcal{V}_{21}^T$$

$$\mathcal{V}_{22} = \sum_{d=0}^{1} \rho_d E\left[\Phi(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)^{\otimes 2}|D = d\right] - \sum_{d=0}^{1} \rho_d E\left[\Phi(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)|D = d\right]^{\otimes 2}.$$

Notice that here we use the short-hand notation $\boldsymbol{A}^{\otimes 2} = \boldsymbol{A}\boldsymbol{A}^T$.

Assumption 3 (Identifiability)**.** There exists a unique $p \times 1$ vector $\boldsymbol{\beta}_0 \in \text{int}(\boldsymbol{\Theta}_\beta)$ with $\boldsymbol{\Theta}_\beta \in \mathbb{R}^p$ and compact, such that $E[\Psi_d(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)|D = d] = \boldsymbol{0}$ for both $d = 0, 1$ and

$$\mathcal{V}_{11}^{-1}\left[\sum_{d=0}^{1} \rho_d E[\Psi(\boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}_0)|D = d]\right] \neq \boldsymbol{0}$$

for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$.

Assumption 4 (Additional regularity conditions)**.**

1. With probability 1, $\Upsilon(\boldsymbol{Z}, \boldsymbol{\theta})$ is continuous for all $\boldsymbol{\theta} \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_\beta \times \boldsymbol{\Theta}_\gamma$ and is once continuously differentiable in an ϵ-neighborhood of $\boldsymbol{\theta}_0$, $\mathcal{N}(\boldsymbol{\theta}_0, \epsilon)$.

2. Denote $\|\cdot\|$ as the Frobenius norm, for both $d = 0, 1$,

$$E\left[\sup_{\theta\in\boldsymbol{\Theta}} \|\Upsilon(\boldsymbol{Z}, \boldsymbol{\theta})\| \,\Big|\, D = d\right] < \infty, \quad E\left[\sup_{\theta\in\boldsymbol{\Theta}} \|\Upsilon(\boldsymbol{Z}, \boldsymbol{\theta})\|^2 \,\Big|\, D = d\right] < \infty,$$
$$E\left[\sup_{\theta\in N(\boldsymbol{\theta}_0,\epsilon)} \left\|\partial\Upsilon(\boldsymbol{Z}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}^T\right\| \,\Big|\, D = d\right] < \infty.$$

WEB APPENDIX B. PROOF OF PROPOSITION 1

Using standard theory of estimating equations [1], we can easily show that $\hat{\boldsymbol{\beta}}^0 \xrightarrow{p} \boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}^1 \xrightarrow{p} \boldsymbol{\beta}_0$, and so it follows that $\hat{\boldsymbol{\beta}}^{01} \xrightarrow{p} \boldsymbol{R}\boldsymbol{\beta}_0$, where the rectangular matrix $\boldsymbol{R} = (\boldsymbol{I}^{p\times p}, \boldsymbol{I}^{p\times p})^T$ is defined in the main manuscript. By assumption $\hat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$, so we conclude by continuous mapping theorem that $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ from the GLS formulation $\hat{\boldsymbol{\beta}} = (\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\beta}}^{01}$.

For asymptotic normality of $\hat{\boldsymbol{\beta}}$, we define the $2p \times 2p$ block diagonal matrix

$$\mathcal{Q} = \begin{pmatrix} \rho_0 E\left[\partial\Psi_0(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)/\partial\boldsymbol{\beta}^T \Big| D = 0\right] & \boldsymbol{0} \\ \boldsymbol{0} & \rho_1 E\left[\partial\Psi_1(\boldsymbol{Z}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)/\partial\boldsymbol{\beta}^T \Big| D = 1\right] \end{pmatrix}.$$

We assume that $\mathcal{Q}$ is non-singular so that $\mathcal{Q}^{-1}$ exists. Observe that $\mathcal{J}_{11} = \mathcal{Q}\boldsymbol{R}$. Assuming the regularity conditions hold, we expand (note that Ψ_0 and Ψ_1 are has zero covariance)

$$\boldsymbol{0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \Psi_0(\boldsymbol{Z}_i, \hat{\boldsymbol{\beta}}^0, \hat{\boldsymbol{\gamma}}) \\ \Psi_1(\boldsymbol{Z}_i, \hat{\boldsymbol{\beta}}^1, \hat{\boldsymbol{\gamma}}) \end{pmatrix}$$

around $\boldsymbol{\theta}_0$ and rearrange the terms to get

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{01} - \boldsymbol{R}\boldsymbol{\beta}_0) = -\mathcal{Q}^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi(\boldsymbol{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\right] - \mathcal{Q}^{-1}\mathcal{J}_{12}\left[\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\right] + o_p(1). \tag{1}$$

Meanwhile, we have the following stochastic expansion for the logistic disease model

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = -\mathcal{J}_{22}^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi(\boldsymbol{Z}_i, \boldsymbol{\gamma}_0)\right] + o_p(1). \tag{2}$$

Inserting (2) into (1), we obtain

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{01} - \boldsymbol{R}\boldsymbol{\beta}_0) = -\mathcal{Q}^{-1}\mathcal{L}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Upsilon(\boldsymbol{Z}_i, \boldsymbol{\theta}_0)\right] + o_p(1) \xrightarrow{d} N(\boldsymbol{0}, \mathcal{Q}^{-1}\mathcal{E}\mathcal{Q}^{-T}), \tag{3}$$

where $\mathcal{E} = \mathcal{L}\mathcal{V}\mathcal{L}^T$, $\mathcal{L} = (\boldsymbol{I}^{2p\times 2p}, -\mathcal{J}_{12}\mathcal{J}_{22}^{-1})$, and the last result follows from Lindeberg-Feller central limit theorem and it is immediate that $\boldsymbol{\Omega} = \mathcal{Q}^{-1}\mathcal{E}\mathcal{Q}^{-T}$. Now since $\hat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$, we have $\hat{\boldsymbol{\Omega}}^{-1} \xrightarrow{p} \boldsymbol{\Omega}^{-1}$ and $(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1} \xrightarrow{p} \left(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}$, a finite quantity, and so we can write

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= -\left(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\mathcal{Q}^{-1}\mathcal{L}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Upsilon(\boldsymbol{Z}_i, \boldsymbol{\theta}_0)\right] + o_p(1) \\ &\xrightarrow{d} N(\boldsymbol{0}, \left[\boldsymbol{R}^T(\mathcal{Q}^{-1}\mathcal{E}\mathcal{Q}^{-T})^{-1}\boldsymbol{R}\right]^{-1}) = N\left(\boldsymbol{0}, (\mathcal{J}_{11}^T\mathcal{E}^{-1}\mathcal{J}_{11})^{-1}\right).\end{aligned} \tag{4}$$

To consistently estimate the above asymptotic variance matrix in case-control samples, we use the sample counterpart to replace the retrospective expectations.

From Hansen[2] and Hall[3], the asymptotic variance for $\hat{\boldsymbol{\theta}}_{\text{GMM}}$ is given by $(\mathcal{J}^T\mathcal{V}^{-1}\mathcal{J})^{-1}$. We now re-express the asymptotic variance for $\hat{\boldsymbol{\beta}}_{\text{GMM}}$ by a sequence of block matrix inversion arguments and show that it is identical to the asymptotic variance of the GLS estimator even after accounting for the uncertainty in estimating the weights from the disease model. Since the covariance matrix $\mathcal{V}$ is assumed finite and positive definite, $\mathcal{V}^{-1}$ exists and by block matrix inversion

$$\mathcal{V}^{-1} = \begin{pmatrix} \mathcal{V}^{11} & \mathcal{V}^{12} \\ \mathcal{V}^{21} & \mathcal{V}^{22} \end{pmatrix} = \begin{pmatrix} (\mathcal{V}_{11} - \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\mathcal{V}_{21})^{-1} & -\mathcal{V}_{11}^{-1}\mathcal{V}_{12}(\mathcal{V}_{22} - \mathcal{V}_{21}\mathcal{V}_{11}^{-1}\mathcal{V}_{12})^{-1} \\ -\mathcal{V}_{22}^{-1}\mathcal{V}_{21}(\mathcal{V}_{11} - \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\mathcal{V}_{21})^{-1} & (\mathcal{V}_{22} - \mathcal{V}_{21}\mathcal{V}_{11}^{-1}\mathcal{V}_{12})^{-1} \end{pmatrix}.$$

By matrix multiplication, we have

$$\mathcal{J}^T\mathcal{V}^{-1}\mathcal{J} = \begin{pmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} \\ \mathcal{K}_{21} & \mathcal{K}_{22} \end{pmatrix},$$

where $\mathcal{K}_{11} = \mathcal{J}_{11}^T\mathcal{V}^{11}\mathcal{J}_{11}$, $\mathcal{K}_{12} = \mathcal{K}_{21}^T = \mathcal{J}_{11}^T(\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})$ and $\mathcal{K}_{22} = \mathcal{J}_{12}^T\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{J}_{22}^T\mathcal{V}^{21}\mathcal{J}_{12} + \mathcal{J}_{12}^T\mathcal{V}^{12}\mathcal{J}_{22} + \mathcal{J}_{22}^T\mathcal{V}^{22}\mathcal{J}_{22}$. The asymptotic variance of $\hat{\boldsymbol{\beta}}_{\text{GMM}}$ is the upper-left $p \times p$ sub-matrix of $(\mathcal{J}^T\mathcal{V}^{-1}\mathcal{J})^{-1}$, which by block matrix inversion is

$$\begin{aligned}\text{avar}(\hat{\boldsymbol{\beta}}_{\text{GMM}}) &= (\mathcal{J}_{11}^T\mathcal{V}^{11}\mathcal{J}_{11} - \mathcal{K}_{12}\mathcal{K}_{22}^{-1}\mathcal{K}_{21})^{-1} = \left[\mathcal{J}_{11}^T\left[\mathcal{V}^{11} - (\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})\mathcal{K}_{22}^{-1}(\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})\right]\mathcal{J}_{11}\right]^{-1} \\ &= \left[\mathcal{J}_{11}^T\mathcal{F}^{-1}\mathcal{J}_{11}\right]^{-1}\end{aligned}$$

Applying the Sherman-Morrison-Woodbury inversion formula[4],

$$\begin{aligned}\mathcal{F} &= \left[\mathcal{V}^{11} - (\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})\mathcal{K}_{22}^{-1}(\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})\right]^{-1} \\ &= \left(\mathcal{V}^{11}\right)^{-1} + \left[\mathcal{J}_{12} + \left(\mathcal{V}^{11}\right)^{-1}\mathcal{V}^{12}\mathcal{J}_{22}\right] \times \\ &\quad \left[\mathcal{K}_{22} - (\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})^T\left(\mathcal{V}^{11}\right)^{-1}(\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})\right]^{-1}\left[\mathcal{J}_{12} + \left(\mathcal{V}^{11}\right)^{-1}\mathcal{V}^{12}\mathcal{J}_{22}\right]^T\end{aligned}$$

Observe that

$$\begin{aligned}&\left[\mathcal{K}_{22} - (\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})^T\left(\mathcal{V}^{11}\right)^{-1}(\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{V}^{12}\mathcal{J}_{22})\right] \\ &= \mathcal{J}_{12}^T\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{J}_{22}^T\mathcal{V}^{21}\mathcal{J}_{12} + \mathcal{J}_{12}^T\mathcal{V}^{12}\mathcal{J}_{22} + \mathcal{J}_{22}^T\mathcal{V}^{22}\mathcal{J}_{22} - \left(\mathcal{J}_{22}^T\mathcal{V}^{21}\mathcal{J}_{12} + \mathcal{J}_{12}^T\mathcal{V}^{11}\mathcal{J}_{12} + \mathcal{J}_{22}^T\mathcal{V}^{21}\left(\mathcal{V}^{11}\right)^{-1}\mathcal{V}^{12}\mathcal{J}_{22} + \mathcal{J}_{12}^T\mathcal{V}^{12}\mathcal{J}_{22}\right) \\ &= \mathcal{J}_{22}^T\mathcal{V}^{22}\mathcal{J}_{22} - \mathcal{J}_{22}^T\mathcal{V}^{21}\left(\mathcal{V}^{11}\right)^{-1}\mathcal{V}^{12}\mathcal{J}_{22} = \mathcal{J}_{22}^T\mathcal{V}_{22}^{-1}\mathcal{J}_{22},\end{aligned}$$

where the last equality comes from the fact that $\mathcal{V}_{22} = \left(\mathcal{V}^{22} - \mathcal{V}^{21}\left(\mathcal{V}^{11}\right)^{-1}\mathcal{V}^{12}\right)^{-1}$ by inverting $\mathcal{V}^{-1}$. Next, observe the following,

1. $\left(\mathcal{J}_{22}^T\mathcal{V}_{22}^{-1}\mathcal{J}_{22}\right)^{-1} = \mathcal{J}_{22}^{-1}\mathcal{V}_{22}\mathcal{J}_{22}^{-T}$,

2. The inverse of $\mathcal{V}$ can also be written as

$$\mathcal{V}^{-1} = \begin{pmatrix}\mathcal{V}^{11} & \mathcal{V}^{12} \\ \mathcal{V}^{21} & \mathcal{V}^{22}\end{pmatrix} = \begin{pmatrix}(\mathcal{V}_{11} - \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\mathcal{V}_{21})^{-1} & -(\mathcal{V}_{11} - \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\mathcal{V}_{21})^{-1}\mathcal{V}_{12}\mathcal{V}_{22}^{-1} \\ (\mathcal{V}_{22} - \mathcal{V}_{21}\mathcal{V}_{11}^{-1}\mathcal{V}_{12})^{-1}\mathcal{V}_{21}\mathcal{V}_{11}^{-1} & (\mathcal{V}_{22} - \mathcal{V}_{21}\mathcal{V}_{11}^{-1}\mathcal{V}_{12})^{-1}\end{pmatrix},$$

3. Since the aforementioned two representations are equal, we can write

$$\mathcal{V}^{21} = -\mathcal{V}_{22}^{-1}\mathcal{V}_{21}(\mathcal{V}_{11} - \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\mathcal{V}_{21})^{-1} = -\mathcal{V}_{22}^{-1}\mathcal{V}_{21}\mathcal{V}^{11} \tag{5}$$

$$\mathcal{V}^{12} = -(\mathcal{V}_{11} - \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\mathcal{V}_{21})^{-1}\mathcal{V}_{12}\mathcal{V}_{22}^{-1} = -\mathcal{V}^{11}\mathcal{V}_{12}\mathcal{V}_{22}^{-1} \tag{6}$$

Note that (5) and (6) can be obtained from each other from the fact that $\mathcal{V}^{-1}$ is a symmetric (positive definite) matrix.

Using the above intermediate results, we have

$$\begin{aligned}\mathcal{F} &= \left(\mathcal{V}^{11}\right)^{-1} + \left[\mathcal{J}_{12}\mathcal{J}_{22}^{-1} + \left(\mathcal{V}^{11}\right)^{-1}\mathcal{V}^{12}\right]\mathcal{V}_{22}\left[\mathcal{V}^{21}\left(\mathcal{V}^{11}\right)^{-1} + \mathcal{J}_{22}^{-1}\mathcal{J}_{12}^T\right] \\ &= \left(\mathcal{V}^{11}\right)^{-1} + \left[\mathcal{J}_{12}\mathcal{J}_{22}^{-1} - \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\right]\mathcal{V}_{22}\left[-\mathcal{V}_{22}^{-1}\mathcal{V}_{21} + \mathcal{J}_{22}^{-1}\mathcal{J}_{12}^T\right] \\ &= \left(\mathcal{V}^{11}\right)^{-1} - \mathcal{J}_{12}\mathcal{J}_{22}^{-1}\mathcal{V}_{21} + \mathcal{V}_{12}\mathcal{V}_{22}^{-1}\mathcal{V}_{21} + \mathcal{J}_{12}\mathcal{J}_{22}^{-1}\mathcal{V}_{22}\mathcal{J}_{22}^{-1}\mathcal{J}_{12}^T - \mathcal{V}_{12}\mathcal{J}_{22}^{-1}\mathcal{J}_{12}^T \\ &= \mathcal{V}_{11} - \mathcal{J}_{12}\mathcal{J}_{22}^{-1}\mathcal{V}_{21} + \mathcal{J}_{12}\mathcal{J}_{22}^{-1}\mathcal{V}_{22}\mathcal{J}_{22}^{-1}\mathcal{J}_{12}^T - \mathcal{V}_{12}\mathcal{J}_{22}^{-1}\mathcal{J}_{12}^T = \mathcal{L}\mathcal{V}\mathcal{L}^T = \mathcal{E}.\end{aligned}$$

Therefore, $\hat{\boldsymbol{\beta}}$ is asymptotically as efficient as $\hat{\boldsymbol{\beta}}_{\text{GMM}}$, and also achieves the asymptotic efficiency bound imposed by the retrospective moment conditions.

WEB APPENDIX C. PROOF OF PROPOSITION 2

Observe that the likelihood-ratio statistic can be simplified to

$$S_{\mathrm{LR}} = 2n(\hat{\boldsymbol{\beta}}^{01} - \boldsymbol{R}\hat{\boldsymbol{\beta}})^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}^T[\boldsymbol{C}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}^T]^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{t}) + S_{\mathrm{Wald}} = S_{\mathrm{Wald}}$$

since the first-order condition for minimizing the distance function is

$$\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\beta}}^{01} - \boldsymbol{R}\hat{\boldsymbol{\beta}}) = \boldsymbol{0}. \tag{7}$$

Similarly we can verify $S_{\mathrm{score}} = S_{\mathrm{Wald}}$ by using condition (7), thus the three tests statistics are numerically equivalent.

Under the null, $\boldsymbol{C}\boldsymbol{\beta}_0 - \boldsymbol{t} = \boldsymbol{0}$. By expansion (4), we observe that

$$\begin{aligned}\sqrt{n}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{t}) &= \left[\sqrt{n}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{t})\right] - \left[\sqrt{n}(\boldsymbol{C}\boldsymbol{\beta}_0 - \boldsymbol{t})\right] = \boldsymbol{C}\left[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] \\ &= -\boldsymbol{C}\left(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\mathcal{Q}^{-1}\mathcal{L}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Upsilon(\boldsymbol{Z}_i, \boldsymbol{\theta}_0)\right] + o_p(1).\end{aligned} \tag{8}$$

Since $\hat{\boldsymbol{\Omega}}^{-1} \xrightarrow{p} \boldsymbol{\Omega}^{-1} = (\mathcal{Q}^T\mathcal{E}^{-1}\mathcal{Q})^{-1}$, the Wald statistic can be written as

$$S_{\mathrm{Wald}} = \boldsymbol{u}^T\mathcal{V}^{1/2}\mathcal{L}^T\mathcal{Q}^{-T}\boldsymbol{\Omega}^{-1}\boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}^T\left[\boldsymbol{C}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}^T\right]^{-1}\boldsymbol{C}\left(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\mathcal{Q}^{-1}\mathcal{L}\mathcal{V}^{1/2}\boldsymbol{u} + o_p(1), \tag{9}$$

where $\boldsymbol{u}$ is a $2p + q$ dimensional, mean-zero and unit-variance Gaussian random vector. Note that the matrix sandwiched between $\boldsymbol{u}^T$ and $\boldsymbol{u}$ is a projection matrix with rank k since we could write $\boldsymbol{C}(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}^T = \boldsymbol{\Lambda}^T\boldsymbol{\Lambda}$ with $\boldsymbol{\Lambda} = \boldsymbol{C}\left(\boldsymbol{R}^T\boldsymbol{\Omega}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\mathcal{Q}^{-1}\mathcal{L}\mathcal{V}^{1/2}$. Hence $S_{\mathrm{Wald}} \xrightarrow{d} \chi^2(k)$. Further, under the sequence of Pitman local alternatives $H_{1,n}$, we could write

$$\begin{aligned}\sqrt{n}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{t}) &= \left[\sqrt{n}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{t})\right] - \left[\sqrt{n}(\boldsymbol{C}\boldsymbol{\beta}_0 - \boldsymbol{t} - \boldsymbol{\delta}/\sqrt{n})\right] = \boldsymbol{C}\left[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] + \boldsymbol{\delta} \\ &= -\boldsymbol{C}\left(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\mathcal{Q}^{-1}\mathcal{L}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Upsilon(\boldsymbol{Z}_i, \boldsymbol{\theta}_0)\right] + \boldsymbol{\delta} + o_p(1),\end{aligned}$$

and the non-central Chi-squared distributional results in Proposition 2 follow readily.

WEB APPENDIX D. PROOF OF PROPOSITION 3

For equivalence between $nM_n(\hat{\boldsymbol{\beta}})$ and T_n, it suffices to check $\hat{\boldsymbol{\Omega}}^{-1} - \hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1} = \bar{\boldsymbol{R}}\left(\bar{\boldsymbol{R}}^T\hat{\boldsymbol{\Omega}}\bar{\boldsymbol{R}}\right)^{-1}\bar{\boldsymbol{R}}^T$, which is equivalent to showing

$$\boldsymbol{I} - \hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1/2} = \hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}}\left(\bar{\boldsymbol{R}}^T\hat{\boldsymbol{\Omega}}\bar{\boldsymbol{R}}\right)^{-1}\bar{\boldsymbol{R}}^T\hat{\boldsymbol{\Omega}}^{1/2}. \tag{10}$$

It is straightforward to see that $\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1/2}$ is the orthogonal projection matrix to the column space of $\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R}$, denoted by $\mathbb{C}(\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R})$, while $\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}}\left(\bar{\boldsymbol{R}}^T\hat{\boldsymbol{\Omega}}\bar{\boldsymbol{R}}\right)^{-1}\bar{\boldsymbol{R}}^T\hat{\boldsymbol{\Omega}}^{1/2}$ is the orthogonal projection matrix to the column space of $\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}}$, denoted by $\mathbb{C}(\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}})$. For arbitrary $p\times 1$ vectors λ_1 and λ_2, we have $\lambda_1\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R}\in\mathbb{C}(\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R})$, and $\lambda_2\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}}\in\mathbb{C}(\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}})$. However,

$$\left(\lambda_1\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R}\right)^T\lambda_2\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}}=\lambda_1\lambda_2\boldsymbol{R}^T\bar{\boldsymbol{R}}=\boldsymbol{0},$$

suggesting that the orthogonal complement of $\mathbb{C}(\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R})$ is $\mathbb{C}(\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}})$, namely

$$\left\{\mathbb{C}(\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R})\right\}^{\perp}=\mathbb{C}(\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}}).$$

Therefore the corresponding (unique) orthogonal projection matrices sum to an identity matrix and equation

$$\boldsymbol{I}=\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{R}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1/2}+\hat{\boldsymbol{\Omega}}^{1/2}\bar{\boldsymbol{R}}\left(\bar{\boldsymbol{R}}^T\hat{\boldsymbol{\Omega}}\bar{\boldsymbol{R}}\right)^{-1}\bar{\boldsymbol{R}}^T\hat{\boldsymbol{\Omega}}^{1/2},\tag{11}$$

holds. This establishes the equivalence between $nM_n(\hat{\boldsymbol{\beta}})$ and T_n

For asymptotic independence between $nM_n(\hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\beta}}$, note that we could write $nM_n(\hat{\boldsymbol{\beta}})=\boldsymbol{u}^T\boldsymbol{N}_1\boldsymbol{u}+o_p(1)$, where

$$\boldsymbol{N}_1=\mathcal{V}^{1/2}\mathcal{L}^T\mathcal{Q}^{-T}\left(\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}\boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\right)\mathcal{Q}^{-1}\mathcal{L}\mathcal{V}^{1/2}$$

with $\boldsymbol{u}$ a $2p+q$ dimensional, mean zero and unit-variance Gaussian random vector. From (4), we have $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\boldsymbol{N}_2\boldsymbol{u}+o_p(1)$, where $\boldsymbol{N}_2=-\left(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\mathcal{Q}^{-1}\mathcal{L}\mathcal{V}^{1/2}$. Asymptotic independence holds since $\boldsymbol{N}_2\boldsymbol{N}_1=\boldsymbol{0}$. Further, it should be noted that $nM_n(\hat{\boldsymbol{\beta}})$ and the test statistic S_{Wald} are also asymptotically independent. From (9), $S_{\text{Wald}}=\boldsymbol{u}^T\boldsymbol{N}_3\boldsymbol{u}+o_p(1)$, where

$$\boldsymbol{N}_3=\mathcal{V}^{1/2}\mathcal{L}^T\mathcal{Q}^{-T}\boldsymbol{\Omega}^{-1}\boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}^T\left[\boldsymbol{C}(\boldsymbol{R}^T\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}^T\right]^{-1}\boldsymbol{C}\left(\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^T\boldsymbol{\Omega}^{-1}\mathcal{Q}^{-1}\mathcal{L}\mathcal{V}^{1/2}.$$

The asymptotic independence hold since $\boldsymbol{N}_3\boldsymbol{N}_1=\boldsymbol{0}$.

WEB APPENDIX E. DATA PREPROCESSING FOR THE PAD CASE-CONTROL DATA

The case-control study of peripheral arterial disease (PAD) was originally designed as a genome-wide association study for identifying genetic variants mediating susceptibility to PAD (The datasets were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000203.v1.p1.). PAD cases are defined as having a post-exercise ankle-brachial index no larger than 0.9, a history of lower extremity re-vascularization, or having poorly-compressible leg arteries. Controls are identified as having no history of PAD or had an ankle-brachial index between 1.0 and 1.3. The original sample consists of phenotypic information for 1688 PAD cases and 1649 disease-free controls. Genotype data (657366 SNPs) were available for 3432 samples with intentional replicates. We followed the steps outlined in Reed et al.[5] to preprocess the genotype data before

merging with the phenotype data for the purpose of (i) identifying the largest subset of unrelated individuals and (ii) estimating the (genetic) principal components to control for potential population stratification in the secondary analysis.

SNP-level filtering was conducting by filtering out SNPs with call rate less than 95%, minor allele frequency (MAF) less than 0.05 and that violating Hardy-Weinberg Equilibrium (in the controls). Sample-level filtering was conducted by filtering out individuals with more than 5% missing SNPs and excess heterozygosity (an indication of poor sample quality), which is defined by an inbreeding coefficient larger than 0.1. Linkage disequilibrium (LD) pruning was applied with a threshold value of 0.2, resulting in a subset of 59563 SNPs for principal component analysis (PCA). Based on the LD-pruned set of SNPs, related samples and duplicates were further removed by the estimating the measure of identity by descent (IBD), namely samples with IBD kinship coefficient greater than 0.1 were excluded for further analysis. PCA was performed based on the LD-pruned set of SNPs with the first two eigenvalues to be 7.6 and 3.6, indicating that approximately 0.3% of case-control status was explained by population substructure (by the largest principal component). The lack of population substructure is not surprising since the majority of the samples come from the European ancestry. For final data analysis, we removed outliers using the method suggested by Price et al.[6], which results in a final sample of 1554 cases and 1556 controls. As suggested by Reed et al.[5], LD pruning and PCA were then re-applied to the final sample. Based on assessment of the scree plot, we chose to adjust for the first genetic principal component in the secondary analysis of BMI to control for potential confounding by population substructure. Adjusting for additional genetic principal components gives very similar results which are not presented.

WEB APPENDIX F. WEB TABLES

Web Table 1 Absolute bias of $\hat{\beta}$, coverage probability of the associated 95% confidence interval and power of the specification tests when the link function is misspecified in the fitted disease model with a gamma secondary trait ($\nu = 2$).

		Common Disease			Rare Disease		
Link	Method	Abs Bias	Coverage	Power	Abs Bias	Coverage	Power
$t(1)$	GLS	0.004	0.947	0.063	0.001	0.949	0.017
	SPREG	0.007	0.945	–	0.000	0.949	–
$t(2)$	GLS	0.005	0.946	0.052	0.001	0.946	0.022
	SPREG	0.004	0.949	–	0.000	0.950	–
$t(4)$	GLS	0.002	0.946	0.104	0.002	0.947	0.064
	SPREG	0.004	0.947	–	0.004	0.947	–
$t(8)$	GLS	0.001	0.938	0.141	0.003	0.947	0.075
	SPREG	0.009	0.944	–	0.008	0.945	–
$t(16)$	GLS	0.003	0.940	0.155	0.001	0.948	0.047
	SPREG	0.013	0.939	–	0.032	0.918	–
Probit	GLS	0.005	0.932	0.167	0.007	0.950	0.222
	SPREG	0.015	0.930	–	0.135	0.421	–

Web Table 2 Average association estimates ($\hat{\beta}_1$ or $\hat{\beta}_1^*$), absolute bias of the association estimates, coverage probability of the associated 95% confidence interval and power of the specification tests when the genotype-by-disease interaction is omitted in the fitted disease model with a gamma secondary trait ($\nu = 2$). In particular, the absolute bias and coverage are with respect to $\beta_1 = -0.12$ when the secondary trait model is correctly specified and to $\beta_1^* = -0.22$ when the secondary trait model is misspecified.

		Correct Secondary Trait Model				Misspecified Secondary Trait Model			
	γ_4	Avg Est	Abs Bias	Coverage	Power	Avg Est	Abs Bias	Coverage	Power
Common Disease	-log(1.3)	-0.211	0.091	0.239	0.882	-0.315	0.095	0.174	0.928
	-log(1.2)	-0.183	0.063	0.527	0.551	-0.287	0.067	0.466	0.659
	-log(1.1)	-0.153	0.033	0.828	0.165	-0.255	0.035	0.810	0.197
	0	-0.121	0.001	0.949	0.053	-0.221	0.001	0.949	0.051
	log(1.1)	-0.093	0.027	0.866	0.227	-0.191	0.029	0.848	0.260
	log(1.2)	-0.073	0.047	0.691	0.613	-0.169	0.051	0.644	0.691
	log(1.3)	-0.056	0.064	0.516	0.882	-0.151	0.069	0.439	0.924
Rare Disease	-log(1.3)	-0.255	0.135	0.017	0.904	-0.362	-0.142	0.006	0.945
	-log(1.2)	-0.218	0.098	0.153	0.605	-0.323	0.103	0.098	0.695
	-log(1.1)	-0.174	0.054	0.613	0.196	-0.277	0.057	0.566	0.245
	0	-0.121	0.001	0.951	0.049	-0.221	0.001	0.950	0.046
	log(1.1)	-0.067	0.053	0.645	0.216	-0.164	0.056	0.600	0.257
	log(1.2)	-0.022	0.098	0.186	0.675	-0.117	0.103	0.141	0.760
	log(1.3)	-0.011	0.131	0.047	0.934	-0.084	0.136	0.030	0.963

References

1. Vaart AW. *Asymptotic Statistics*. New York: Cambridge University Press; 2000.

2. Hansen LP. Large sample properties of generalised method of moments estimators.. *Econometrica.* 1982;50(4):1029–1054.

3. Hall AR. *Generalized Method of Moments*. New York: Oxford University Press Inc.; 2004.

4. Golub GH, Van Loan CF. *Matrix Computations, Third Edition*. Baltimore, MD: The John Hopkins University Press; 1996.

5. Reed Eric, Nunez Sara, Kulp David, Qian Jing, Reilly Muredach P., Foulkes Andrea S.. A guide to genome-wide association analysis and post-analytic interrogation. *Statistics in Medicine.* 2015;34(28):3769–3792.

6. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich David. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics.* 2006;38(8):904–909.