Statistics II | **Problem Set 4**
Professor: Marcelo J. Moreira | **Solutions**
TA: Luan Borelli | **September 09, 2024**

---

1. *The OLS estimator $\hat{\beta}_n$ minimizes $(y - X\beta)'(y - X\beta)$. Now, consider the constrained estimator $\tilde{\beta}_n$ which minimizes $(y - X\beta)'(y - X\beta)$ subject to $R'\beta = c$.*

(a) *Show that*

$$\tilde{\beta}_n = \hat{\beta}_n - (X'X)^{-1}R\left(R'(X'X)^{-1}R\right)^{-1}\left(R'\hat{\beta}_n - c\right).$$

Solution. The constrained least squares estimator is

$$\begin{aligned}\tilde{\beta}_n &= \arg\min_{R'\beta=c}(y - X\beta)'(y - X\beta)\\ &= \arg\min_{R'\beta=c}\left(y'y - 2y'X\beta + \beta'X'X\beta\right).\end{aligned}$$

This problem is equivalent to finding the critical points of the Lagrangian

$$\mathcal{L}(\beta, \lambda) = \frac{1}{2}\left(y'y - 2y'X\beta + \beta'X'X\beta\right) + \lambda'(R'\beta - c)$$

over (β, λ) where λ is a vector of Lagrangian multipliers. The solution is a saddle point. The Lagrangian is minimized over β while maximized over λ. The first-order conditions for the solution are

$$\begin{aligned}&\frac{\partial}{\partial\beta}\mathcal{L}(\tilde{\beta}_n, \tilde{\lambda}) = -X'y + X'X\tilde{\beta}_n + R\tilde{\lambda} = 0,\\ \text{and}\quad &\frac{\partial}{\partial\lambda}\mathcal{L}(\tilde{\beta}_n, \tilde{\lambda}) = R'\tilde{\beta}_n - c = 0.\end{aligned}$$

Premultiplying the former by $R'(X'X)^{-1}$ we obtain

$$-R'\hat{\beta}_n + R'\tilde{\beta}_n + R'(X'X)^{-1}R\tilde{\lambda} = 0.$$

Imposing $R'\tilde{\beta}_n - c = 0$ and solving for $\tilde{\lambda}$ we find

$$\tilde{\lambda}_n = \left[R'(X'X)^{-1}R\right]^{-1}(R'\hat{\beta}_n - c).$$

Substituting this expression into the first first-order condition we obtain

$$\tilde{\beta}_n = \hat{\beta}_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}(R'\hat{\beta}_n - c),$$

as desired. □

(b) *When errors are homoskedastic, show that*

$$V\left(\hat{\beta}_n|X\right) - V\left(\tilde{\beta}_n|X\right) = \sigma^2(X'X)^{-1}R\left(R'(X'X)^{-1}R\right)^{-1}R'(X'X)^{-1} \geq 0.$$

Solution. When errors are (conditionally) homoskedastic, we have $\mathbb{E}[uu'|X] = I_n\sigma^2$. Write $\hat{\beta}_n = \beta + (X'X)^{-1}X'u$. Then

$$\begin{aligned} V\left(\hat{\beta}_n|X\right) &= V\left((X'X)^{-1}X'u|X\right) = \mathbb{E}[(X'X)^{-1}X'uu'X(X'X)^{-1}|X] \\ &= (X'X)^{-1}X'\mathbb{E}[uu'|X]X(X'X)^{-1} = (X'X)^{-1}\sigma^2. \end{aligned}$$

Further, observe that

$$\begin{aligned} \tilde{\beta}_n &= \hat{\beta}_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}(R'\hat{\beta}_n - c) \\ &= \hat{\beta}_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'\hat{\beta}_n + (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}c \\ &= [I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R']\hat{\beta}_n + (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}c \\ &= [I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'](\beta + (X'X)^{-1}X'u) + (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}c \\ &= \underbrace{[I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R']\beta}_{\text{Conditionally nonstochastic}} + [I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'](X'X)^{-1}X'u \\ &\quad + \underbrace{(X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}c}_{\text{Conditionally nonstochastic}}. \end{aligned}$$

Note that, conditional on X, the only stochastic component is the second term. As a result, the remaining terms can be disregarded when computing the conditional variance. Therefore,

$$\begin{aligned} V\left(\tilde{\beta}_n|X\right) &= V\left([I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'](X'X)^{-1}X'u\right) \\ &= \mathbb{E}[[I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'](X'X)^{-1}X'uu'X(X'X)^{-1} \\ &\qquad \times [I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R']'|X] \\ &= [I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'](X'X)^{-1}X'\mathbb{E}[uu'|X]X(X'X)^{-1} \\ &\qquad \times [I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R']' \\ &= [I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'](X'X)^{-1}[I_n - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R']'\sigma^2 \\ &= [(X'X)^{-1} - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'(X'X)^{-1}][I_n - R\left[R'(X'X)^{-1}R\right]^{-1}R'(X'X)^{-1}]\sigma^2 \\ &= (X'X)^{-1}\sigma^2 - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'(X'X)^{-1}\sigma^2 \\ &\quad - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'(X'X)^{-1}\sigma^2 \\ &\quad + (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}\underbrace{R'(X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}}_{\text{Identity}}R'(X'X)^{-1}\sigma^2 \\ &= \underbrace{(X'X)^{-1}\sigma^2}_{V(\hat{\beta}|X)} - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}R'(X'X)^{-1}\sigma^2. \end{aligned}$$

Therefore,

$$V\left(\hat{\beta}_n|X\right) - V\left(\tilde{\beta}_n|X\right) = \sigma^2(X'X)^{-1}R\left(R'(X'X)^{-1}R\right)^{-1}R'(X'X)^{-1},$$

as desired. To verify that this matrix is positive semidefinite, simply note that

$$\begin{aligned}
&\sigma^2(X'X)^{-1}R\left(R'(X'X)^{-1}R\right)^{-1}R'(X'X)^{-1}\\
&=\sigma^2(X'X)^{-1/2}(X'X)^{-1/2}R\left(R'(X'X)^{-1/2}(X'X)^{-1/2}R\right)^{-1}R'(X'X)^{-1/2}(X'X)^{-1/2}\\
&=\sigma^2(X'X)^{-1/2\prime}N_{(X'X)^{-1/2}R}(X'X)^{-1/2},
\end{aligned}$$

where

$$N_{(X'X)^{-1/2}R}\equiv R\left(R'(X'X)^{-1/2}(X'X)^{-1/2}R\right)^{-1}R'(X'X)^{-1/2}$$

is the projection matrix of $(X'X)^{-1/2}R$. Since $V\left(\hat{\beta}_n|X\right)-V\left(\tilde{\beta}_n|X\right)$ is a sandwich form in a positive semidefinite matrix,[1] it must be positive semidefinite. □

---

2. *Suppose that*

$$y_i=\beta_1+\exp(x_i\beta_2)+u_i,$$

where (x_i,u_i) *are iid with* $\mathbb{E}[u_i|x_i]=0$ *and* $V[u_i|x_i]=\sigma^2$. *We find the estimator* $\hat{\beta}_n$ *for* $\beta=(\beta_1,\beta_2)$ *by minimizing*

$$Q_n(\beta)\equiv\sum_{i=1}^n(y_i-\beta_1-\exp(x_i\beta_2))^2.$$

Find the asymptotic distribution of $\sqrt{n}(\hat{\beta}_n-\beta)$.

Solution. To simplify notation, let $h(x_i,\beta)\equiv\beta_1+\exp(x_i\beta_2)$. The first-order conditions are

$$\begin{aligned}
\nabla_\beta Q_n(\beta)&=-2\sum_{i=1}^n\nabla_\beta h(x_i,\beta)[y_i-h(x_i,\beta)]\\
&=-2\sum_{i=1}^n\begin{bmatrix}1\\\exp(x_i\beta_2)x_i\end{bmatrix}[y_i-\beta_1-\exp(x_i\beta_2)]=0.
\end{aligned}$$

A first-order Taylor expansion of $\nabla_\beta Q_n(\beta)$ around the true parameter value $\beta^*\equiv(\beta_1^*,\beta_2^*)$ gives us

$$0=\nabla_\beta Q_n(\hat{\beta})\approx\nabla_\beta Q_n(\beta^*)+\nabla_{\beta\beta'}Q_n(\beta^*)(\hat{\beta}-\beta^*),\tag{1}$$

where the Hessian matrix $\nabla_{\beta\beta'}Q_n(\beta^*)$ is given by

$$\begin{aligned}
\nabla_{\beta\beta'}Q_n(\beta^*)&=-2\sum_{i=1}^n\nabla_{\beta\beta'}h(x_i,\beta)[y_i-h(x_i,\beta)]+2\sum_{i=1}^n\nabla_\beta h(x_i,\beta)\nabla_\beta h(x_i,\beta)'\\
&=-2\sum_{i=1}^n\begin{bmatrix}0&0\\0&\exp(x_i\beta_2)x_i^2\end{bmatrix}[y_i-h(x_i,\beta)]+2\sum_{i=1}^n\begin{bmatrix}1&\exp(x_i\beta_2)x_i\\\exp(x_i\beta_2)x_i&\exp(x_i\beta_2)^2x_i^2\end{bmatrix}.
\end{aligned}$$

[1] Recall that every projection matrix is positive semidefinite.

We can multiply (1) by $\sqrt{n}$ and isolate $\sqrt{n}(\hat{\beta} - \beta^*)$ to obtain

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta^*) &\approx -\left[\nabla_{\beta\beta'} Q_n(\beta^*)\right]^{-1} \sqrt{n} \nabla_\beta Q_n(\beta^*) \\ &= -\left[\frac{1}{n} \nabla_{\beta\beta'} Q_n(\beta^*)\right]^{-1} \sqrt{n} \left[\frac{1}{n} \nabla_\beta Q_n(\beta^*) - \mathbb{E}[\nabla_\beta Q_n(\beta^*)]\right],\end{aligned}$$

where the last equality follows from multiplying and dividing the expression by n and using the fact that $\mathbb{E}[\nabla_\beta Q_n(\beta^*)] = 0$. This holds because $\nabla_\beta Q_n(\beta^*) = -2\sum_{i=1}^n \nabla_\beta h(x_i, \beta) u_i$ and $\mathbb{E}[u_i] = 0$.

Now, by the law of large numbers and the continuous mapping theorem, we have that

$$\begin{aligned}\left[\frac{1}{n} \nabla_{\beta\beta'} Q_n(\beta^*)\right]^{-1} &\xrightarrow{p} \mathbb{E}\left[-2\nabla_{\beta\beta'} h(x_i, \beta)[y_i - h(x_i, \beta)] + 2\nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]^{-1} \\ &= \left[-2\nabla_{\beta\beta'} h(x_i, \beta) \underbrace{\mathbb{E}[u_i]}_{=0} + 2\mathbb{E}\left[\nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]\right]^{-1} \\ &= \frac{1}{2} \mathbb{E}\left[\nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]^{-1} \qquad (2) \\ &= \frac{1}{2} \begin{bmatrix} 1 & \mathbb{E}[\exp(x_i\beta_2)x_i] \\ \mathbb{E}[\exp(x_i\beta_2)x_i] & \mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] \end{bmatrix}^{-1}. \qquad (3)\end{aligned}$$

Furthermore, by the central limit theorem,

$$\begin{aligned}\sqrt{n}\left[\frac{1}{n} \nabla_\beta Q_n(\beta^*) - \mathbb{E}[\nabla_\beta Q_n(\beta^*)]\right] &\xrightarrow{d} Z \sim N\left(0, 4\mathbb{E}\left[u_i^2 \nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]\right) \qquad (4) \\ \text{(LIE)} &= N\left(0, 4\mathbb{E}\left[\sigma^2 \nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]\right) \\ &= N\left(0, 4\sigma^2 \begin{bmatrix} 1 & \mathbb{E}[\exp(x_i\beta_2)x_i] \\ \mathbb{E}[\exp(x_i\beta_2)x_i] & \mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] \end{bmatrix}\right). \qquad (5)\end{aligned}$$

Therefore, from (2) and (4), by Slutsky's theorem,

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta^*) &\xrightarrow{d} \frac{1}{2} \mathbb{E}\left[\nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]^{-1} Z \\ &= N(0, \mathbb{E}\left[\nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]^{-1} \mathbb{E}\left[u_i^2 \nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right] \mathbb{E}\left[\nabla_\beta h(x_i, \beta) \nabla_\beta h(x_i, \beta)'\right]^{-1}). \qquad (6)\end{aligned}$$

Or, more specifically, using (3) and (5),

$$\begin{aligned}N\Bigg(0, \sigma^2 &\begin{bmatrix} 1 & \mathbb{E}[\exp(x_i\beta_2)x_i] \\ \mathbb{E}[\exp(x_i\beta_2)x_i] & \mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} 1 & \mathbb{E}[\exp(x_i\beta_2)x_i] \\ \mathbb{E}[\exp(x_i\beta_2)x_i] & \mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] \end{bmatrix} \begin{bmatrix} 1 & \mathbb{E}[\exp(x_i\beta_2)x_i] \\ \mathbb{E}[\exp(x_i\beta_2)x_i] & \mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] \end{bmatrix}^{-1}\Bigg)\end{aligned}$$

$$= N\left(0, \sigma^2 \begin{bmatrix} 1 & \mathbb{E}[\exp(x_i\beta_2)x_i] \\ \mathbb{E}[\exp(x_i\beta_2)x_i] & \mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] \end{bmatrix}^{-1}\right)$$

$$= N\left(0, \frac{\sigma^2}{\mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] - \mathbb{E}[\exp(x_i\beta_2)x_i]^2} \begin{bmatrix} \mathbb{E}[\exp(x_i\beta_2)^2 x_i^2] & -\mathbb{E}[\exp(x_i\beta_2)x_i] \\ -\mathbb{E}[\exp(x_i\beta_2)x_i] & 1 \end{bmatrix}^{-1}\right).$$

Notice that the asymptotic distribution obtained in (6) is quite general for nonlinear least squares estimators, holding for any nonlinear function h satisfying standard regularity conditions. The above expression is simply a specialization for the case $h(x_i, \beta) = \beta_1 + \exp(x_i\beta_2)$ under homoskedastic errors. □

---

3. [17.2, LNs] *Consider the model*

$$y_i = I(x_i'\beta^* + u_i > 0),$$

where β^* is the true parameter and (x_i', u_i) are iid with $u_i|x_i$ having a $N(0, 1)$ distribution with cdf Φ.

(a) *Find $\mathbb{E}[y_i|x_i; \beta]$, the conditional expectation of y_i given x_i.*

Solution. The event $y_i = 1$ is the same as $x_i'\beta^* + u_i > 0$. Thus

$$\begin{aligned} P[y_i = 1 \mid x_i] &= P[x_i'\beta^* + u_i > 0 \mid x_i] = P[u_i > -x_i'\beta^* \mid x_i] \\ &= 1 - P[u_i < -x_i'\beta^* \mid x_i] = 1 - \Phi(-x_i'\beta^*) = \Phi(x_i'\beta^*). \end{aligned}$$

Therefore

$$\mathbb{E}[y_i|x_i; \beta] = \Phi(x_i'\beta^*) \cdot 1 + [1 - \Phi(x_i'\beta^*)] \cdot 0 = \Phi(x_i'\beta^*).$$

□

(b) *Find $V[y_i \mid x_i; \beta]$, the conditional variance of y_i given x_i. Hint: y_i is a Bernoulli random variable.*

Solution. Observe that the event $y_i^2 = 1$ is also the same as $x_i'\beta^* + u_i > 0$. Therefore just as in **(a)** we have $\mathbb{E}[y_i^2|x_i; \beta] = \Phi(x_i'\beta^*)$. It follows that

$$\begin{aligned} V[y_i|x_i; \beta] &= \mathbb{E}[y_i^2|x_i; \beta] - \mathbb{E}[y_i|x_i; \beta]^2 \\ &= \Phi(x_i'\beta^*) - \Phi(x_i'\beta^*)^2 \\ &= \Phi(x_i'\beta^*)[1 - \Phi(x_i'\beta^*)]. \end{aligned}$$

□

(c) *Consider the estimator $\hat{\beta}$ which minimizes*

$$\sum_{i=1}^{n} (y_i - \mathbb{E}[y_i|x_i, \beta])^2.$$

Find the limiting distribution of $\sqrt{n}(\hat{\beta} - \beta^*)$.

Solution. Notice that since $\mathbb{E}[y_i|x_i, \beta] = \Phi(x_i'\beta)$, $\hat{\beta}$ can be understood as a nonlinear least squares estimator with $h(x_i, \beta) = \Phi(x_i'\beta)$ for a Probit model $y_i = \Phi(x_i'\beta) + e_i$. Therefore, applying the results from Exercise 2 for the asymptotic distribution of nonlinear least squares estimators,

$$\text{Avar}[\hat{\beta}] = \mathbb{E}[h_{\beta i}h_{\beta i}']^{-1}\mathbb{E}[h_{\beta i}h_{\beta i}'e_i^2]\mathbb{E}[h_{\beta i}h_{\beta i}']^{-1},$$

where $h_{\beta i} \equiv \partial h(x_i, \beta^*)/\partial\beta$. For $h(x_i, \beta) = \Phi(x_i'\beta)$ we have

$$h_{\beta i} = x_i\phi(x_i'\beta^*).$$

Thus, by observing that $V[y_i|x_i;\beta] = V[e_i|x_i;\beta] = \Phi(x_i'\beta^*)[1 - \Phi(x_i'\beta^*)]$,

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta^*) &\xrightarrow{d} N(0, \mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2 e_i^2 x_i x_i']\mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}) \\
\text{(LIE)} &= N(0, \mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2 V[y_i|x_i;\beta^*] x_i x_i']\mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}) \\
&= N(0, \mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2 \Phi(x_i'\beta^*)[1 - \Phi(x_i'\beta^*)] x_i x_i']\mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}).
\end{aligned}$$

□

(d) *Consider the estimator $\tilde{\beta}$ which minimizes*

$$\sum_{i=1}^{n} \frac{(y_i - \mathbb{E}[y_i|x_i, \beta])^2}{V[y_i|x_i;\beta^*]}.$$

Find the limiting distribution of $\sqrt{n}(\tilde{\beta} - \beta^*)$.

Solution. Observe that

$$\sum_{i=1}^{n} \frac{(y_i - \mathbb{E}[y_i|x_i, \beta])^2}{V[y_i|x_i;\beta^*]} = \sum_{i=1}^{n} \left(\frac{y_i}{V[y_i|x_i;\beta^*]^{1/2}} - \frac{\mathbb{E}[y_i|x_i, \beta]}{V[y_i|x_i;\beta^*]^{1/2}}\right)^2.$$

In a similar fashion to what was done in item **(c)**, $\tilde{\beta}$ can be understood as a nonlinear least squares estimator for a rescaled (by $V[y_i|x_i;\beta^*]^{1/2}$) Probit model

$$\frac{y_i}{V[y_i|x_i;\beta^*]^{1/2}} = \frac{\Phi(x_i'\beta)}{V[y_i|x_i;\beta^*]^{1/2}} + \varepsilon_i.$$

This is equivalent to

$$y_i = \Phi(x_i'\beta) + V[y_i|x_i;\beta^*]^{1/2}\varepsilon_i.$$

Thus from **(c)** we have the relation $e_i = V[y_i|x_i;\beta^*]^{1/2}\varepsilon_i$, whence it follows that

$$\sqrt{n}(\hat{\beta} - \beta^*) \xrightarrow{d} N(0, \mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2 V[y_i|x_i;\beta^*]\varepsilon_i^2 x_i x_i']\mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}).$$

But notice that $V[\varepsilon_i] = V[e_i]/V[y_i|x_i;\beta^*] = V[y_i|x_i;\beta^*]/V[y_i|x_i;\beta^*] = 1$. Therefore by LIE

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta^*) &\xrightarrow{d} N(0, \mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2 V[y_i|x_i;\beta^*] x_i x_i']\mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}) \\
&= N(0, \mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}\mathbb{E}[\phi(x_i'\beta^*)^2 \Phi(x_i'\beta^*)[1 - \Phi(x_i'\beta^*)]\mathbb{E}[\phi(x_i'\beta^*)^2 x_i x_i']^{-1}).
\end{aligned}$$

This is exactly the same asymptotic distribution obtained in item **(c)**. □

(e) *Compare the asymptotic variance of $\hat{\beta}$ and $\tilde{\beta}$. Explain your answer.*

Solution. The asymptotic variances of $\hat{\beta}$ and $\tilde{\beta}$ are exactly the same. Division by $V[y_i|x_i;\beta^*]$ just provides a useful normalization that makes the error variance unitary. □

---

4. [17.3, LNs] *Suppose that the logistic, rather than probit, model applies. So $\mathbb{E}[y|X] = G(x'\theta)$, with $G(a) = \frac{\exp(a)}{1+\exp(a)}$. Show that the ZES-rule estimator c satisfies $X'u = 0$, where $u = \{u_i\}$ with $u_i = y_i - G(x_i'c)$.*

Solution. In a logistic model, y_i is a binary random variable that takes the value 1 or 0. Therefore, its distribution follows a Bernoulli distribution with the conditional probability mass function being

$$f(y_i|x_i) = G(x_i'\theta)^y[1 - G(x_i'\theta)]^{1-y}.$$

Thus

$$\ln f(y_i|x_i) = y \ln G(x_i'\theta) + (1-y)\ln\left(1 - G(x_i'\theta)\right).$$

The score is then given by

$$\begin{aligned}\frac{\partial}{\partial\theta}\ln f(y_i|x_i) &= y\frac{G'(x_i'\theta)}{G(x_i'\theta)} - (1-y)\frac{G'(x_i'\theta)}{1-G(x_i'\theta)} \\ &= \frac{G'(x_i'\theta)\,[y_i - G(x_i'\theta)]}{G(x_i'\theta)\,[1 - G(x_i'\theta)]}.\end{aligned}$$

But notice that

$$G'(x_i'\theta) = \frac{\exp(x_i'\theta)x_i}{[1+\exp(x_i'\theta)]^2} \quad \text{and} \quad G(x_i'\theta)\,[1 - G(x_i'\theta)] = \frac{\exp(x_i'\theta)}{[1+\exp(x_i'\theta)]^2}.$$

Thus

$$\frac{G'(x_i'\theta)}{G(x_i'\theta)[1-G(x_i'\theta)]} = x_i,$$

whence

$$\frac{\partial}{\partial\theta}\ln f(y_i|x_i) = x_i[y_i - G(x_i'\theta)].$$

The ZES-rule estimator c is obtained by setting the sample analog of

$$\mathbb{E}[x_i[y_i - G(x_i'c)]] = \mathbb{E}[x_i u_i]$$

to zero and solving for c. That is, the ZES-rule estimator solves

$$n^{-1}\sum_{i=1}^{n} x_i u_i = 0,$$

or, equivalently, by letting $X \equiv (x_1, x_2, \ldots, x_n)'$ and $u \equiv (u_1, u_2, \ldots, u_n)$,

$$X'u = 0.$$

□

5. [18.1, LNs] *This question is about the GMM estimator. Define*

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^n g(X_i,\theta)' A_n \frac{1}{n}\sum_{i=1}^n g(X_i,\theta)/2,$$

where $A_n \xrightarrow{p} A$ *and* $\theta \in \mathbb{R}$.

(a) *Rewrite the proof done in class that the GMM estimator is asymptotically normal for the case* $\theta \in \mathbb{R}^k$*:*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, (\Gamma_0' A \Gamma_0)^{-1} \Gamma_0' A V_0 A \Gamma_0 (\Gamma_0' A \Gamma_0)^{-1}\right),$$

where $\Gamma_0 \equiv \mathbb{E}\left[\frac{\partial}{\partial\theta} g(X_i,\theta)\right]$ *and* $V_0 \equiv \mathbb{E}\left[g(X_i,\theta)g(X_i,\theta)'\right]$.

Solution. Being an interior solution to the problem of maximizing $Q_n(\theta)$, $\hat{\theta}$ satisfies the first order conditions

$$-\left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\hat{\theta})\right)' A_n \left(n^{-1}\sum_{i=1}^n g(X_i,\hat{\theta})\right) = 0. \tag{7}$$

Assuming g is C^1, a first-order Taylor series expansion of $n^{-1}\sum_{i=1}^n g(X_i,\hat{\theta})$ gives us

$$n^{-1}\sum_{i=1}^n g(X_i,\hat{\theta}) \approx n^{-1}\sum_{i=1}^n g(X_i,\theta_0) + \left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\theta_0)\right)(\hat{\theta} - \theta_0),$$

so that, plugging back this expression into the first-order conditions,

$$\begin{aligned} &-\left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\hat{\theta})\right)' A_n \left(n^{-1}\sum_{i=1}^n g(X_i,\theta_0)\right) \\ &\quad -\left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\hat{\theta})\right)' A_n \left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\theta_0)\right)(\hat{\theta} - \theta_0) \approx 0. \end{aligned}$$

Multiplying the above expression by $\sqrt{n}$ and solving for $\sqrt{n}(\hat{\theta} - \theta_0)$ we obtain

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) \approx -&\left[\left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\hat{\theta})\right)' A_n \left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\theta_0)\right)\right]^{-1} \\ &\times \left(n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\hat{\theta})\right)' A_n \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0)\right). \end{aligned}$$

By the law of large numbers we have that

$$n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\hat{\theta}) \xrightarrow{p} \Gamma_0 \quad \text{and} \quad n^{-1}\sum_{i=1}^n \nabla_\theta g(X_i,\theta_0) \xrightarrow{p} \Gamma_0,$$

where $\Gamma_0 \equiv \mathbb{E}[\nabla_\theta g(w_t;\theta_0)]$. Further, by the central limit theorem we have that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i,\theta_0) \xrightarrow{d} N(0,V_0),$$

where $V_0 \equiv \mathbb{E}[g(X_i,\hat{\theta})g(X_i,\hat{\theta})']$. Finally, by hypothesis, we have that $A_n \xrightarrow{p} A$, so by the continuous mapping and Slutsky's theorems we conclude that

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} (\Gamma_0' A\Gamma_0)^{-1}\Gamma_0' A \cdot N(0,V_0) = N\left(0, (\Gamma_0' A\Gamma_0)^{-1}\Gamma_0' A V_0 A\Gamma_0(\Gamma_0' A\Gamma_0)^{-1}\right),$$

as desired. □

(b) *Show that A minimizes the variance based on the limiting distribution of $\sqrt{n}(\hat{\theta}_n-\theta)$ is $A = V_0^{-1}$.*

Solution. When $A = V_0^{-1}$, the asymptotic variance reduces to $(\Gamma_0' V_0^{-1}\Gamma_0)^{-1}$. We want to show that for any A,

$$(\Gamma_0' A\Gamma_0)^{-1}\Gamma_0' A V_0 A\Gamma_0(\Gamma_0' A\Gamma_0)^{-1} - (\Gamma_0' V_0^{-1}\Gamma_0)^{-1} \succeq 0,$$

where "$\succeq$" denotes "is positive semidefinite". Recall that for matrices F and G,

$$G - F \succeq 0 \iff F^{-1} - G^{-1} \succeq 0.$$

Therefore, we can alternatively show that

$$\Gamma_0' V_0^{-1}\Gamma_0 - \Gamma_0' A\Gamma_0(\Gamma_0' A V_0 A\Gamma_0)^{-1}\Gamma_0' A\Gamma_0 \succeq 0.$$

By taking the matrix square root of V_0^{-1} and wisely pre- and post-multiplying terms by $V_0^{-1/2}$ and $V_0^{1/2}$, the left-hand side of the expression above can be rewritten as

$$\begin{aligned}
&\Gamma_0' V_0^{-1/2} V_0^{-1/2}\Gamma_0 - \Gamma_0' V_0^{-1/2} V_0^{1/2} A\Gamma_0(\Gamma_0' A V_0 A\Gamma_0)^{-1}\Gamma_0' A V_0^{1/2} V_0^{-1/2}\Gamma_0 \\
&= \Gamma_0' V_0^{-1/2}[I - V_0^{1/2} A\Gamma_0(\Gamma_0' A V_0 A\Gamma_0)^{-1}\Gamma_0' A V_0^{1/2}]V_0^{-1/2}\Gamma_0 \\
&= (V_0^{-1/2}\Gamma_0)' M_{V_0^{1/2}A\Gamma_0} V_0^{-1/2}\Gamma_0,
\end{aligned}$$

where $M_{V_0^{1/2}A\Gamma_0} \equiv I - V_0^{1/2}A\Gamma_0(\Gamma_0' A V_0 A\Gamma_0)^{-1}\Gamma_0' A V_0^{1/2}$ is the annihilator matrix of $V_0^{1/2}A\Gamma_0$, which is positive semidefinite.[2] Thus the expression is a sandwich form in a positive semidefinite matrix; whence it follows that it must be positive semidefinite.[3] This concludes the proof that $A = V_0^{-1}$ minimizes the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta)$. □

[2] Recall that annihilator matrices are always positive semidefinite.

[3] Not satisfied? Let $H \equiv V_0^{-1/2}\Gamma_0$. Then the expression becomes $H' M_{V_0^{1/2}A\Gamma_0} H$. Or, using the idempotency and symmetry of the annihilator matrix, $H M_{V_0^{1/2}A\Gamma_0}(H M_{V_0^{1/2}A\Gamma_0})'$. Now, for any z we have $z' H M_{V_0^{1/2}A\Gamma_0}(H M_{V_0^{1/2}A\Gamma_0})' z = \|M_{V_0^{1/2}A\Gamma_0} H' z\|^2 \geq 0$, proving the positive semidefiniteness.

---

6. [18.2, LNs] *Consider the population model $y = X\beta + u$, where $E(u|X) = 0$. Show that the OLS estimator can be seen as a GMM estimator.*

Solution. Partition $X = (x_i, \dots, x_n)'$ and consider the $\mathbb{R}^k \to \mathbb{R}^k$ moment function $g(x_i, \beta) = x_i(y_i - x_i'\beta)$. This moment function yields k valid moment conditions for the k-dimensional parameter vector β. Indeed, since $y_i - x_i'\beta = u_i$, by the law of iterated expectations we have

$$\mathbb{E}[g(x_i, \beta)] = \mathbb{E}[x_i(y_i - x_i'\beta)] = \mathbb{E}[x_i u_i] = \mathbb{E}[x_i \mathbb{E}[u_i|X]] = 0.$$

The sample analog for the k moment conditions above is

$$n^{-1} \sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}) = 0.$$

Rearranging,

$$\left(\sum_{i=1}^{n} x_i x_i' \right) \hat{\beta} = \sum_{i=1}^{n} x_i y_i.$$

Isolating $\hat{\beta}$,

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i y_i,$$

or, equivalently, in matrix form,

$$\hat{\beta} = (X'X)^{-1} X'y,$$

which is precisely the OLS estimator. □

---

7. [18.4, LNs] *Suppose that you have k unbiased estimators of a parameter θ:*

$$T_j = \frac{1}{n} \sum_{i=1}^{n} g_j(X_i), j = 1, \dots, k,$$

where X_i are iid random variables. Let V be the $k \times k$ covariance matrix for $\sqrt{n}T$, where $T = (T_1, \dots, T_k)'$. Assume that V is known and nonsingular.

(a) *Find the estimator with minimum variance among the class of unbiased estimators of the form $a'T = \sum_{j=1}^{k} a_j T_j$.*

Solution. First, notice that unbiasedness of $a'T$ implies that we must have

$$\mathbb{E}[a'T] = \sum_{j=1}^{k} a_j \mathbb{E}[T_j] = \left(\sum_{j=1}^{k} a_j \right) \theta = \theta,$$

whence it follows that we must have $\sum_{j=1}^{k} a_j = 1$. Let $g(X_i) \equiv (g_1(X_i), g_2(X_i), \ldots, g_k(X_i))'$. We want to minimize

$$\begin{aligned} V[a'T] &= a'V\left[n^{-1}\sum_{i=1}^{n} g(X_i)\right] a \\ &= a'\left(n^{-2}\sum_{i=1}^{n} V[g(X_i)]\right) a \\ &= \frac{a'V[g(X_i)]a}{n} \end{aligned}$$

subject to the unbiasedness constraint $\mathbf{1}'a = \sum_{j=1}^{k} a_j = 1$, where $\mathbf{1}$ is a k-dimensional vector of ones. That is, we want to solve the problem

$$\min_{a} \frac{a'V[g(X_i)]a}{2n} \quad \text{s.t.} \quad \mathbf{1}'a = 1.$$

The Lagrangian for this problem is

$$\mathcal{L}(a, \lambda) = \frac{a'V[g(X_i)]a}{2n} + \lambda\left[1 - \mathbf{1}'a\right].$$

The first order conditions are

$$\begin{aligned} n^{-1}V[g(X_i)]\hat{a} - \lambda\mathbf{1} &= 0, \\ 1 - \mathbf{1}'\hat{a} &= 0. \end{aligned}$$

The former condition yields $\hat{a} = n\lambda V[g(X_i)]^{-1}\mathbf{1}$, while the latter $\mathbf{1}'\hat{a} = 1$. Thus,

$$\mathbf{1}'\hat{a} = n\lambda\mathbf{1}'V[g(X_i)]^{-1}\mathbf{1} = 1,$$

whence

$$\lambda = n^{-1}\frac{1}{\mathbf{1}'V[g(X_i)]^{-1}\mathbf{1}}.$$

Therefore,

$$\hat{a} = \frac{V[g(X_i)]^{-1}\mathbf{1}}{\mathbf{1}'V[g(X_i)]^{-1}\mathbf{1}}.$$

The estimator for θ with minimum variance among the class of unbiased estimators of the form $a'T$ is then

$$\hat{\theta} = \frac{\mathbf{1}'V[g(X_i)]^{-1}}{\mathbf{1}'V[g(X_i)]^{-1}\mathbf{1}}T.$$

□

(b) *Show that the estimator in (a) coincides with the GMM estimator based on k moment conditions $\mathbb{E}[g_j(X_i) - \theta] = 0$, $j = 1, \ldots, k$.*

Solution. The (optimal) GMM estimator based on the k moment conditions $\mathbb{E}[g(X_i)-\mathbf{1}\theta]=0$ solves the problem

$$\min_\theta \frac{1}{2}\left(n^{-1}\sum_{i=1}^n[g(X_i)-\mathbf{1}\theta]\right)' V[g(X_i)-\mathbf{1}\theta]^{-1}\left(n^{-1}\sum_{i=1}^n[g(X_i)-\mathbf{1}\theta]\right)$$
$$=\min_\theta \frac{1}{2}\left(n^{-1}\sum_{i=1}^n[g(X_i)-\mathbf{1}\theta]\right)' V[g(X_i)]^{-1}\left(n^{-1}\sum_{i=1}^n[g(X_i)-\mathbf{1}\theta]\right).$$

The first-order conditions for an interior solution of this problem are

$$0=\left(n^{-1}\sum_{i=1}^n[-\mathbf{1}]\right)' V[g(X_i)]^{-1}\left(n^{-1}\sum_{i=1}^n[g(X_i)-\mathbf{1}\hat{\theta}]\right)$$
$$=-\mathbf{1}'V[g(X_i)]^{-1}\left(n^{-1}\sum_{i=1}^n g(X_i)-\mathbf{1}\hat{\theta}\right)$$
$$=-n^{-1}\mathbf{1}'V[g(X_i)]^{-1}\sum_{i=1}^n g(X_i)+\mathbf{1}'V[g(X_i)]^{-1}\mathbf{1}\hat{\theta},$$

whence

$$\mathbf{1}'V[g(X_i)]^{-1}\mathbf{1}\hat{\theta}=\mathbf{1}'V[g(X_i)]^{-1}n^{-1}\sum_{i=1}^n g(X_i)=\mathbf{1}'V[g(X_i)]^{-1}T,$$

and hence

$$\hat{\theta}=\frac{\mathbf{1}'V[g(X_i)]^{-1}}{\mathbf{1}'V[g(X_i)]^{-1}\mathbf{1}}T.$$

□

---

8. *Assume that X_i are iid with marginal density $g(x_i,\theta)$.*

(a) *Find the score based on the joint density.*

Solution. Since X_i are iid, the joint density is given by

$$L(X_i,\theta)\equiv\prod_{i=1}^n g(x_i,\theta).$$

Taking the natural logarithm,

$$\ln L(X_i,\theta)=\sum_{i=1}^n \ln g(x_i,\theta).$$

The score based on the joint density is then given by

$$\frac{\partial}{\partial\theta}\ln L(X_i,\theta)=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln g(x_i,\theta).$$

□

(b) *Consider the GMM objective function in which you include the moment based on the score,* $\mathbb{E}[\partial \ln g(x_i,\theta^*)/\partial\theta]=0$ *and additional moments* $\mathbb{E}[h(x_i,\theta^*)]=0$.

Solution. Define the moment function

$$\rho(x_i,\theta)\equiv\begin{bmatrix}\nabla_\theta \ln g(x_i,\theta)\\ h(x_i,\theta)\end{bmatrix},$$

where $\nabla_\theta \ln g(x_i,\theta)$ is a $k\times 1$ score vector and $h(x_i,\theta)$ a $q\times 1$ vector of additional moment functions. This gives us a set of moment conditions

$$\mathbb{E}[\rho(x_i,\theta)]=\begin{bmatrix}\mathbb{E}[\nabla_\theta \ln g(x_i,\theta)]\\ \mathbb{E}[h(x_i,\theta)]\end{bmatrix}=0$$

that can be used to establish a GMM objective function by considering the sample analog

$$\rho_n(\theta)\equiv n^{-1}\sum_{i=1}^{n}\rho(x_i,\theta)=\begin{bmatrix}n^{-1}\sum_{i=1}^{n}\nabla_\theta \ln g(x_i,\theta)\\ n^{-1}\sum_{i=1}^{n}h(x_i,\theta^*)\end{bmatrix}$$

and defining

$$Q_n(\theta)\equiv\rho_n(\theta)'W_n\rho_n(\theta),$$

for some consistent, positive semidefinite, weighting matrix W_n. □

(c) *Compare the asymptotic variance of the GMM estimator with that of the MLE estimator.*

Solution. From 5(a), the asymptotic variance of the GMM estimator is given by

$$(\Gamma_0'A\Gamma_0)^{-1}\Gamma_0'AV_0A\Gamma_0(\Gamma_0'A\Gamma_0)^{-1},$$

where $\Gamma_0\equiv\mathbb{E}\left[\nabla_\theta\rho(x_i,\theta)\right]$ and $V_0\equiv\mathbb{E}\left[\rho(x_i,\theta)\rho(x_i,\theta)'\right]$. And, as we know, the asymptotic variance of the MLE estimator is given by the inverse of the Fisher information matrix,

$$\mathcal{I}^{-1}\equiv\mathbb{E}[\nabla_\theta \ln g(x_i,\theta)\nabla_{\theta'}\ln g(x_i,\theta)]^{-1}.$$

Observe that

$$\begin{aligned}V_0&=\mathbb{E}\begin{bmatrix}\nabla_\theta \ln g(x_i,\theta)\\ h(x_i,\theta)\end{bmatrix}\begin{bmatrix}\nabla_\theta \ln g(x_i,\theta)' & h(x_i,\theta)'\end{bmatrix}\\&=\begin{bmatrix}\mathbb{E}[\nabla_\theta \ln g(x_i,\theta)\nabla_{\theta'}\ln g(x_i,\theta)] & \mathbb{E}[\nabla_\theta \ln g(x_i,\theta)h(x_i,\theta)']\\ \mathbb{E}[\nabla_\theta \ln g(x_i,\theta)h(x_i,\theta)'] & \mathbb{E}[h(x_i,\theta)h(x_i,\theta)']\end{bmatrix}\\&=\begin{bmatrix}\mathcal{I} & C[\nabla_\theta \ln g(x_i,\theta),h(x_i,\theta)]\\ C[\nabla_\theta \ln g(x_i,\theta),h(x_i,\theta)] & V[h(x_i,\theta)]\end{bmatrix}\end{aligned}$$

and that

$$\Gamma_0 = \begin{bmatrix} \mathbb{E}[\nabla_{\theta\theta'} \ln g(x_i, \theta)] \\ \mathbb{E}[\nabla_{\theta'} h(x_i, \theta)] \end{bmatrix} = \begin{bmatrix} -\mathcal{I} \\ \mathbb{E}[\nabla_{\theta'} h(x_i, \theta)] \end{bmatrix}$$

As discussed in Exercise 5(b), when the optimal weighting matrix $A = V_0^{-1}$ is considered, the asymptotic variance of the GMM estimator simplifies to $(\Gamma_0' V_0^{-1} \Gamma_0)^{-1}$, whence

$$\begin{aligned} &(\Gamma_0' V_0^{-1} \Gamma_0)^{-1} \\ &= \begin{bmatrix} -\mathcal{I}' & \mathbb{E}[\nabla_{\theta'} h(x_i, \theta)]' \end{bmatrix} \begin{bmatrix} \mathcal{I} & C[\nabla_\theta \ln g(x_i, \theta), h(x_i, \theta)] \\ C[\nabla_\theta \ln g(x_i, \theta), h(x_i, \theta)] & V[h(x_i, \theta)] \end{bmatrix}^{-1} \begin{bmatrix} -\mathcal{I} \\ \mathbb{E}[\nabla_{\theta'} h(x_i, \theta)] \end{bmatrix}. \end{aligned}$$

Without loss of generality, let $k = q = 1$. In this case θ is a scalar parameter, and hence $\mathcal{I}$, $\nabla_\theta \ln g(x_i, \theta)$, and $h(x_i, \theta)$ scalar functions. Then we can compute the above expression by hand to obtain

$$\begin{aligned} (\Gamma_0' V_0^{-1} \Gamma_0)^{-1} &= \frac{V[h(x_i, \theta)]\mathcal{I} - C[\nabla_\theta \ln g(x_i, \theta), h(x_i, \theta)]^2}{V[h(x_i, \theta)]\mathcal{I} + 2C[\nabla_\theta \ln g(x_i, \theta), h(x_i, \theta)]\mathbb{E}[\nabla_\theta h(x_i, \theta)] + \mathbb{E}[\nabla_\theta h(x_i, \theta)]^2} \mathcal{I}^{-1} \\ &=: F\mathcal{I}^{-1}. \end{aligned} \tag{8}$$

From this expression, we can see that under the optimal weighting matrix, the asymptotic variances of the GMM and ML estimators are related through the term F. It is evident that when $F = 1$, the asymptotic variances are the same. In the next section, we will show that a sufficient condition for this equality to hold is that the likelihood is correctly specified. □

(d) *What happens with item (c) if the likelihood is correctly specified?*

Solution. When the likelihood is correctly specified — that is, when the data x_i is indeed generated from a distribution with marginal density $g(x_i, \theta)$ — we have the identity

$$0 = \mathbb{E}[h(x_i, \theta^*)] = \int h(x_i, \theta^*) g(x_i, \theta^*) dx_i.$$

Differentiating this identity, assuming differentiation under the integral is allowed, gives

$$\begin{aligned} 0 &= \int h(x_i, \theta^*) g(x_i, \theta^*) dx_i \\ &= \int \nabla_\theta h(x_i, \theta^*) g(x_i, \theta^*) dx_i + \int h(x_i, \theta^*) \nabla_{\theta'} g(x_i, \theta) dx_i \\ &= \mathbb{E}[\nabla_\theta h(x_i, \theta^*)] + \mathbb{E}[h(x_i, \theta^*) \nabla_{\theta'} \ln g(x_i, \theta^*)]. \end{aligned}$$

Thus,

$$\mathbb{E}[\nabla_\theta h(x_i, \theta^*)] = -\mathbb{E}[h(x_i, \theta^*) \nabla_{\theta'} \ln g(x_i, \theta^*)] = -C[\nabla_\theta \ln g(x_i, \theta), h(x_i, \theta)].$$

The term F in (8) then simplifies as

$$\begin{aligned} F &= \frac{V[h(x_i,\theta)]\mathcal{I} - C[\nabla_\theta \ln g(x_i,\theta), h(x_i,\theta)]^2}{V[h(x_i,\theta)]\mathcal{I} + 2C[\nabla_\theta \ln g(x_i,\theta), h(x_i,\theta)]\mathbb{E}[\nabla_\theta h(x_i,\theta)] + \mathbb{E}[\nabla_\theta h(x_i,\theta)]^2} \\ &= \frac{V[h(x_i,\theta)]\mathcal{I} - \mathbb{E}[\nabla_\theta h(x_i,\theta)]^2}{V[h(x_i,\theta)]\mathcal{I} - 2\mathbb{E}[\nabla_\theta h(x_i,\theta)]^2 + \mathbb{E}[\nabla_\theta h(x_i,\theta)]^2} \\ &= \frac{V[h(x_i,\theta)]\mathcal{I} - \mathbb{E}[\nabla_\theta h(x_i,\theta)]^2}{V[h(x_i,\theta)]\mathcal{I} - \mathbb{E}[\nabla_\theta h(x_i,\theta)]^2} \\ &= 1. \end{aligned}$$

Therefore, the expression for the asymptotic variance of the GMM estimator obtained in (8) simplifies to

$$(\Gamma_0' V_0^{-1} \Gamma_0)^{-1} = \mathcal{I}^{-1}.$$

In other words, the asymptotic variances of the GMM and ML estimators become identical. This is expected, as it is possible to show that:

1. The maximum likelihood estimator is the most efficient estimator in the entire class of GMM estimators [see Newey and McFadden (1994), Theorem 5.1, for a proof]

2. Including additional moments in a GMM specification *never* hurts (i.e., never increases) the asymptotic variance of a GMM estimator [see Hall (2005), Theorem 6.1, for a proof]

A logical consequence of these two facts is that if a correctly specified score is included in the moment conditions, the asymptotic variance of the resulting estimator must be the inverse of the Fisher information matrix. Since a correctly specified score contains all the relevant information from the data, any additional moment condition is informationally redundant. □

---

9. *Consider the moment conditions*

$$\mathbb{E}\begin{bmatrix} X_i - \theta \\ (X_i - \theta)^3 \end{bmatrix} = 0.$$

The observations X_i are iid.

(a) *Find the asymptotic variance of the GMM estimator based on the weighting matrix $W_n \to W$.*

Solution. Let $g(X_i,\theta) = (X_i - \theta, (X_i - \theta)^3)'$. As derived in Exercise 5, the asymptotic variance is

$$(\Gamma_0' W \Gamma_0)^{-1} \Gamma_0' W V_0 W \Gamma_0 (\Gamma_0' W \Gamma_0)^{-1},$$

where

$$\Gamma_0 = \mathbb{E}\left[\frac{\partial}{\partial\theta} g(X_i,\theta)\right] = \begin{bmatrix} -1 \\ -3\mathbb{E}[(X_i - \theta)^2] \end{bmatrix} := \begin{bmatrix} -1 \\ -3\mu_2 \end{bmatrix},$$

and

$$V_0 \equiv V \begin{bmatrix} X_i - \theta \\ (X_i - \theta)^3 \end{bmatrix} = \begin{bmatrix} \mathbb{E}[(X_i - \theta)^2] & \mathbb{E}[(X_i - \theta)^4] \\ \mathbb{E}[(X_i - \theta)^4] & \mathbb{E}[(X_i - \theta)^6] \end{bmatrix} =: \begin{bmatrix} \mu_2 & \mu_4 \\ \mu_4 & \mu_6 \end{bmatrix}.$$

W is some generic limiting positive semidefinite weighting matrix. □

(b) *Find the asymptotic variance of the efficient GMM estimator.*

Solution. For the efficient GMM estimator, we must impose the optimal weighting matrix

$$W = V_0^{-1} = \frac{1}{\mu_2\mu_6 - \mu_4^2} \begin{bmatrix} \mu_6 & -\mu_4 \\ -\mu_4 & \mu_2 \end{bmatrix}.$$

As discussed in Exercise 5, in this case the GMM asymptotic variance simplifies to

$$\begin{aligned}(\Gamma_0' V_0^{-1} \Gamma_0)^{-1} &= \left(\begin{bmatrix} -1 & -3\mu_2 \end{bmatrix} \frac{1}{\mu_2\mu_6 - \mu_4^2} \begin{bmatrix} \mu_6 & -\mu_4 \\ -\mu_4 & \mu_2 \end{bmatrix} \begin{bmatrix} -1 \\ -3\mu_2 \end{bmatrix} \right)^{-1} \\ &= \left(\frac{1}{\mu_2\mu_6 - \mu_4^2} \begin{bmatrix} -\mu_6 + 3\mu_2\mu_4 & \mu_4 - 3\mu_2^2 \end{bmatrix} \begin{bmatrix} -1 \\ -3\mu_2 \end{bmatrix} \right)^{-1} \\ &= \left(\frac{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3}{\mu_2\mu_6 - \mu_4^2} \right)^{-1} = \frac{\mu_2\mu_6 - \mu_4^2}{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3}.\end{aligned}$$

□

(c) *Show that the asymptotic variance of the (efficient) GMM estimator is smaller or equal than that based only on* $\mathbb{E}[X_i - \theta] = 0$. *Explain why they are equal when* X_i *are normal with mean* θ.

Solution. As discussed in Exercise 8(d), more information never hurts. The elements of the population moment condition can be viewed as pieces of information about θ. It can be shown that for a $q \times 1$ moment function $g(X_i, \theta)$, and any partition $g(X_i, \theta) = (g_1(X_i, \theta), g_2(X_i, \theta))'$, where $g_i(X_i, \theta)$ is $q_i \times 1$ and $q_1 + q_2 = q$, a GMM estimator based on the full set of moment conditions $\mathbb{E}[g(X_i, \theta)] = 0$ will always have an asymptotic variance that is at least as small as that of a GMM estimator based on a subset of moment conditions, $\mathbb{E}[g_1(X_i, \theta)] = 0$. In particular, consider the efficient GMM estimator based solely on the moment condition $\mathbb{E}[X_i - \theta] = 0$. For this case, we have $\Gamma_0 = -1$ and $V_0 = \mathbb{E}[(X_i - \theta)^2] =: \mu_2$. Thus, the asymptotic variance simplifies to μ_2. This result implies that we must have

$$\frac{\mu_2\mu_6 - \mu_4^2}{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3} \leq \mu_2.$$

Indeed, rearranging we obtain

$$9\mu_2^4 - 6\mu_2^2\mu_4 + \mu_4^2 = (3\mu_2 - \mu^4)^2 \geq 0.$$

The fact that more correct information never hurts does not imply, however, that it always helps. When additional moment conditions are redundant, there is no efficiency gain from adding them to the original set of moment conditions. To illustrate this, consider the case where $X_i \sim N(\theta, \sigma^2)$. In this case, we have that

$$\mu_2 = \sigma^2$$
$$\mu_4 = 3\sigma^4$$
$$\mu_6 = 15\sigma^6.$$

Thus

$$\frac{\mu_2\mu_6 - \mu_4^2}{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3} = \frac{15\sigma^8 - 9\sigma^8}{15\sigma^6 - 18\sigma^6 + 9\sigma^6} = \frac{6\sigma^8}{6\sigma^6} = \sigma^2 = \mu_2.$$

Therefore, when X_i is normally distributed, adding the third central moment to the moment conditions does not provide any efficiency gain. This is a latent implication of the symmetry of the normal distribution: symmetry implies that all odd central moments contain the same amount of information about the distribution. Thus, the third central moment is redundant relative to the first central moment. □

---

10. *Consider the moment conditions*

$$\mathbb{E}[g(X_i, \theta)] = 0 \iff \theta = \theta_0.$$

The observations X_i are iid and the variance of $g(X_i, \theta)$ is the matrix $V(\theta)$. Let $\tilde{\theta}_n$ be a preliminary consistent estimator of θ_0. Assume $W(\theta) = V(\theta)^{-1}$ is a differentiable function in a neighborhood of θ_0.

(a) *Show that*

$$W_n(\theta) = \left(n^{-1}\sum_{i=1}^{n} g(X_i, \theta)g(X_i, \theta)'\right)^{-1} \xrightarrow{p} V(\theta)^{-1}.$$

Solution. By the Law of Large Numbers,

$$\sum_{i=1}^{n} g(X_i, \theta)g(X_i, \theta)' \xrightarrow{p} \mathbb{E}[g(X_i, \theta)g(X_i, \theta)'] =: V(\theta).$$

Thus, by the Continuous Mapping Theorem,

$$\left(\sum_{i=1}^{n} g(X_i, \theta)g(X_i, \theta)'\right)^{-1} \xrightarrow{p} V(\theta)^{-1}.$$

□

(b) *Argue that $W_n(\tilde{\theta}_n)$ converges in probability to $V(\theta_0)^{-1}$.*

Solution. Consistency of $\tilde{\theta}_n$ implies $\tilde{\theta}_n \xrightarrow{p} \theta_0$. Since $W(\theta)$ is differentiable in a neighborhood of θ_0, it must be continuous at that point. The continuous mapping theorem then implies

$$W_n(\tilde{\theta}_n) \xrightarrow{p} W(\theta_0) = V(\theta_0)^{-1}.$$

□

(c) *Consider three alternative estimators*

$$\hat{\theta}_1 = \arg\min_{\theta} \bar{g}_n(\theta)' W_n(\tilde{\theta}_n) \bar{g}_n(\theta),$$
$$\hat{\theta}_2 = \arg\min_{\theta} \bar{g}_n(\theta)' W_n(\theta_0) \bar{g}_n(\theta),$$
$$\hat{\theta}_3 = \arg\min_{\theta} \bar{g}_n(\theta)' W_n(\theta) \bar{g}_n(\theta).$$

Assuming all three estimators are consistent, show that all three standardized estimators $\sqrt{n}(\hat{\theta}_i - \theta_0)$ *have the same distribution. Find that distribution.*

Solution. For $\hat{\theta}_1$ and $\hat{\theta}_2$, the proof follows the same steps as presented in Exercise 5(a). For the third estimator, the continuously updating estimator, the first-order conditions become[4]

$$2\left[n^{-1}\sum_{i=1}^{n}\nabla_{\theta} g(X_i,\hat{\theta}_3)\right]' W_n(\hat{\theta}_3)\left[n^{-1}\sum_{i=1}^{n} g(X_i,\hat{\theta}_3)\right] - \left[\frac{\partial\ \mathrm{vec}W(\hat{\theta}_3)^{-1}}{\partial\theta'}\right]'\left[W(\hat{\theta}_3)\otimes W(\hat{\theta}_3)\right] \times \mathrm{vec}\left[\left(n^{-1}\sum_{i=1}^{n} g(X_i,\hat{\theta}_3)\right)\left(n^{-1}\sum_{i=1}^{n} g(X_i,\hat{\theta}_3)\right)'\right] = 0.$$

Furthermore, since X_i are iid, it can be shown that

$$\frac{\partial\ \mathrm{vec}W(\hat{\theta}_3)^{-1}}{\partial\theta'} = n^{-1}\sum_{i=1}^{n}\left\{[I\otimes g(X_i,\theta)] + [g(X_i,\theta)\otimes I]\right\}\frac{\partial g(X_i,\theta)}{\partial\theta'}.$$

Notice that the first term in the sum of the first-order conditions is similar to the first-order conditions for a non-CUE criterion, such as those presented in (7) in Exercise 5. Therefore, a sufficient condition for the asymptotic equivalence of CUE and non-CUE estimators is that the second term converges in probability to zero. Under this condition, the CUE first-order conditions will reduce to the usual non-CUE first-order conditions. Without loss of generality, consider the scalar case. In this scenario, it can be shown (see Donald and Newey, 2000) that the second term simplifies to[5]

$$\Xi_n(\hat{\theta}) \equiv \left(n^{-1}\sum_{i=1}^{n} g(X_i,\hat{\theta}_3)\right)' W_n(\hat{\theta}_3)\left(n^{-1}\sum_{i=1}^{n} g(X_i,\hat{\theta}_3)\nabla_{\theta'} g(X_i,\hat{\theta}_3)\right) W_n(\hat{\theta}_3)\left(n^{-1}\sum_{i=1}^{n} g(X_i,\hat{\theta}_3)\right)$$

But notice that, assuming consistency (i.e., assuming $\hat{\theta}_3 \xrightarrow{p} \theta_0$), by Slutsky's theorem this terms converges in probability to

$$\Xi(\theta_0) \equiv \underbrace{\mathbb{E}[g(X_i,\theta_0)]'}_{=0} W(\theta_0)\mathbb{E}[g(X_i,\theta_0)\nabla_{\theta'} g(X_i,\theta_0)] W(\theta_0)\underbrace{\mathbb{E}[g(X_i,\theta_0)]}_{=0} = 0.$$

[4] See Hall (2005), p.116. These equations can be derived using Dhrymes (1984). See Proposition 99, p.115, and Proposition 106, p.124.

[5] When θ is a vector, then this expression is a valid expression of the first-order condition for a single element of θ where the derivative terms would involve derivatives for a particular element of the θ vector. Consequently this expression and all of the results to follow can easily be obtained in the case where θ is a vector.

With this result, deriving the asymptotic distribution of the CUE becomes straightforward. The first-order conditions are

$$2\left[n^{-1}\sum_{i=1}^{n}\nabla_\theta g(X_i,\hat\theta_3)\right]' W_n(\hat\theta_3)\left[n^{-1}\sum_{i=1}^{n} g(X_i,\hat\theta_3)\right] - \Xi_n(\hat\theta) = 0.$$

Following 5(a), a first-order Taylor expansion of $\sum_{i=1}^{n} g(X_i,\hat\theta_3)$ around θ_0 allows us to write

$$\begin{aligned}\sqrt{n}(\hat\theta_3-\theta_0) \approx &-\left[\left(n^{-1}\sum_{i=1}^{n}\nabla_\theta g(X_i,\hat\theta_3)\right)' W_n(\hat\theta)\left(n^{-1}\sum_{i=1}^{n}\nabla_\theta g(X_i,\theta_0)\right)\right]^{-1}\\ &\times\left(n^{-1}\sum_{i=1}^{n}\nabla_\theta g(X_i,\hat\theta_3)\right)' W_n(\hat\theta)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i,\theta_0)\right)\\ &-\sqrt{n}\left[\left(n^{-1}\sum_{i=1}^{n}\nabla_\theta g(X_i,\hat\theta_3)\right)' W_n(\hat\theta)\left(n^{-1}\sum_{i=1}^{n}\nabla_\theta g(X_i,\theta_0)\right)\right]^{-1}\Xi_n(\hat\theta).\end{aligned}$$

As we known from 5(a), assuming consistency of $\hat\theta_3$, the former term converges to a normal distribution with its asymptotic variance being equal to

$$\begin{aligned}&(\Gamma_0'W(\theta_0)\Gamma_0)^{-1}\Gamma_0'W(\theta_0)V_0W(\theta_0)\Gamma_0(\Gamma_0'W(\theta_0)\Gamma_0)^{-1}\\ &=(\Gamma_0'V_0^{-1}\Gamma_0)^{-1}\Gamma_0'V_0^{-1}V_0V_0^{-1}\Gamma_0(\Gamma_0'V_0^{-1}\Gamma_0)^{-1}\\ &=(\Gamma_0'V_0^{-1}\Gamma_0)^{-1},\end{aligned}$$

while the latter term, as argued before, converges to zero. Therefore,

$$\sqrt{n}(\hat\theta_3-\theta_0)\xrightarrow{d} N(0,(\Gamma_0'V_0^{-1}\Gamma_0)^{-1}),$$

concluding the proof. □

(c) *Show that*

$$n\bar g_n(\hat\theta_1)'W_n(\tilde\theta_n)\bar g_n(\hat\theta_1) \quad \text{and} \quad n\bar g_n(\hat\theta_3)'W_n(\theta)\bar g_n(\hat\theta_3)$$

converge in distribution to a chi-square with $d-k$ degrees of freedom. This is the Sargan-Hansen J-test: it rejects when the statistic above is larger than the $1-\alpha$ quantile of a chi square with $d-k$ degrees of freedom. Argue that this rejects more when there is no θ_0 such that $\mathbb{E}[g(X_i,\theta)]=0$.

Solution. A first-order Taylor expansion of $\bar g_n(\hat\theta_1)$ around θ_0 gives

$$\bar g_n(\hat\theta_1)\approx \bar g_n(\theta_0)+\nabla_\theta\bar g_n(\theta_0)(\hat\theta_1-\theta_0).$$

Plugging this expression into the first-order conditions we obtain

$$0=\nabla Q_n(\hat\theta_1)\approx\nabla_\theta\bar g_n(\hat\theta_1)'W_n(\tilde\theta_n)[\bar g_n(\theta_0)+\nabla_\theta\bar g_n(\theta_0)(\hat\theta_1-\theta_0)],$$

whence it follows that

$$\hat{\theta}_1 - \theta_0 \approx - \left[\nabla_\theta \bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n) \nabla_\theta \bar{g}_n(\theta_0)\right]^{-1} \nabla_\theta \bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n) \bar{g}_n(\theta_0).$$

Plugging this expression back into the initial Taylor expansion equation we can write

$$\sqrt{n}\bar{g}_n(\hat{\theta}_1) \approx [I_d - \nabla_\theta \bar{g}_n(\theta_0)[\nabla_\theta \bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n) \nabla_\theta \bar{g}_n(\theta_0)]^{-1} \nabla_\theta \bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n)] \sqrt{n}\bar{g}_n(\theta_0).$$

Under standard regularity conditions, the term inside brackets converges in probability to $I_d - \Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0' V_0^{-1}$ and $\sqrt{n}\bar{g}_n(\theta_0) \xrightarrow{d} \xi \sim N(0, V_0)$, whence it follows, by Slutsky's theorem, that

$$\sqrt{n}\bar{g}_n(\hat{\theta}_1) \xrightarrow{d} [I_d - \Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0' V_0^{-1}]\xi.$$

Therefore

$$\begin{aligned} n\bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n)\bar{g}_n(\hat{\theta}_1) &= \sqrt{n}\bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n)\sqrt{n}\bar{g}_n(\hat{\theta}_1) \\ &\xrightarrow{d} \xi'[I_d - \Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0' V_0^{-1}]' V_0^{-1}[I_d - \Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0' V_0^{-1}]\xi \\ &= \xi'[I_d - V_0^{-1}\Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0'] V_0^{-1}[I_d - \Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0' V_0^{-1}]\xi \\ &= \xi'[V_0^{-1} - V_0^{-1}\Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0' V_0^{-1}]\xi \\ &= (V_0^{-1/2}\xi)'[I_d - V_0^{-1/2}\Gamma_0[\Gamma_0' V_0^{-1}\Gamma_0]^{-1}\Gamma_0' V_0^{-1/2}] V_0^{-1/2}\xi \\ &= (V_0^{-1/2}\xi)' M_{V_0^{-1/2}\Gamma_0}(V_0^{-1/2}\xi) \\ &= Z' M_{\Omega^{-1/2}G} Z, \end{aligned}$$

where $Z \sim N(0, I_d)$. Spectral decomposition on $M_{V_0^{-1/2}G}$ gives

$$Z' M_{V_0^{-1/2}\Gamma_0} Z = Z' H \Lambda H' Z = (H'Z)' \Lambda H' Z$$

where Λ is the diagonal matrix of eigenvalues of $M_{V_0^{-1/2}\Gamma_0}$ and H is such that $H'H = I_d$. The latter implies $H'Z \sim N(0, I_d)$. Since $M_{V_0^{-1/2}\Gamma_0}$ is an annihilator matrix, it has $d - k$ eigenvalues equal one and k eigenvalues equal zero. Thus we can write

$$n\bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n)\bar{g}_n(\hat{\theta}_1) = (H'Z)' \begin{bmatrix} I_{d-k} & 0 \\ 0 & 0 \end{bmatrix} H'Z.$$

Let $w \equiv H'Z$ and partition $w = (w_1, w_2)'$, where $w_1 \sim N(0, I_{d-k})$. It follows that

$$n\bar{g}_n(\hat{\theta}_1)' W_n(\tilde{\theta}_n)\bar{g}_n(\hat{\theta}_1) = [w_1' : w_2'] \begin{bmatrix} I_{d-k} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = w_1' w_1 \sim \chi^2_{d-k},$$

concluding the proof. Using the same arguments as in part (a), this proof can be easily adapted to obtain the same result for $\hat{\theta}_3$, the continuously updating estimator. It suffices to replace the first-order conditions for the CUE first-order conditions into the argument and use the fact that the additional term will converge in probability to zero, as demonstrated in part (a). The remainder of the proof remains unchanged. □