**Statistics II** **Problem Set 2**
**Professor:** Marcelo J. Moreira **Solutions**
**TA:** Luan Borelli **August 10, 2023**

---

**1.** *Show that $tr(C'D) = vec(C)'vec(D)$ and that $tr(C'D) = tr(DC')$ for $p \times q$ matrices $C$ and $D$.*

*Solution.* Let $c_i$ and $d_i$ denote the $i$-th columns of $C$ and $D$, respectively. By partitioning $C = \begin{bmatrix} c_1 & c_2 & \cdots & c_q \end{bmatrix}$ and $D = \begin{bmatrix} d_1 & d_2 & \cdots & d_q \end{bmatrix}$ we have

$$C'D = \begin{bmatrix} c_1' \\ c_2' \\ \vdots \\ c_q' \end{bmatrix} \begin{bmatrix} d_1 & d_2 & \cdots & d_q \end{bmatrix} = \begin{bmatrix} c_1'd_1 & c_1'd_2 & \cdots & c_1'd_q \\ c_2'd_1 & c_2'd_2 & \cdots & c_2'd_q \\ \vdots & \vdots & \ddots & \vdots \\ c_q'd_1' & c_q'd_2 & \cdots & c_q'd_q \end{bmatrix}$$

Thus

$$tr(C'D) = \sum_{j=1}^{q} c_j'd_j = \begin{bmatrix} c_1' & c_2' & \cdots & c_q' \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_q \end{bmatrix} = vec(C)'vec(D).$$

It is also easy to see that

$$tr(DC') = tr\left(\sum_{j=1}^{q} d_j c_j'\right) = \sum_{j=1}^{q} tr(d_j c_j') = \sum_{j=1}^{q}\sum_{i=1}^{p} d_{ij}c_{ij} = \sum_{i=1}^{q}\sum_{j=1}^{p} c_{ij}d_{ij} = \sum_{i=1}^{q} c_j'd_j = tr(C'D),$$

where $c_{ij}$ and $d_{ij}$ denote the $(i,j)$ entries of $C$ and $D$, respectively. □

---

**2.** *Let $A$ and $B$ be two symmetric and positive definite $k \times k$ matrices. Prove that $(I_k - A)$ is positive definite if and only if $(A^{-1} - I_k)$ is positive definite. Extend this result to show that $(A - B)$ is positive definite if and only if $(B^{-1} - A^{-1})$ is positive definite.*

*Solution.* For this proof we will be extensively using the well-known fact that if $X \succ 0$, then $C'XC \succ 0$ for any conformable $C$, where "$\succ 0$" here reads "is positive definite". This is just a notation.

First, we shall prove that $I_k - A \succ 0 \iff A^{-1} - I_k \succ 0$.

($\implies$) Since $A$ is symmetric, so is $A^{-1}$. Therefore there exists $A^{-1/2}$ symmetric such that $A^{-1} = A^{-1/2}A^{-1/2}$. Notice that $I_k = A^{-1/2}AA^{-1/2}$.[1] We have that

$$A^{-1} - I_k = A^{-1/2}A^{-1/2} - A^{-1/2}AA^{-1/2} = A^{-1/2}(I_k - A)A^{-1/2} \succ 0.$$

[1] Observe that

$$A^{-1/2}AA^{-1/2} = A^{-1/2}(A^{-1})^{-1}A^{-1/2} = A^{-1/2}(A^{-1/2}A^{-1/2})^{-1}A^{-1/2} = A^{-1/2}(A^{-1/2})^{-1}(A^{-1/2})^{-1}A^{-1/2} = I_k.$$

The conclusion follows from the fact that $I_k - A \succ 0$, by hypothesis.

($\Longleftarrow$) Similarly, since $A$ is symmetric, we can find $A^{1/2}$ such that $A = A^{1/2}A^{1/2}$. Notice that $I_k = A^{1/2}A^{-1}A^{1/2}$. We have that

$$A^{-1} - I_k = A^{1/2}A^{-1}A^{1/2} - A^{1/2}I_kA^{1/2} = A^{1/2}(A^{-1} - I_k)A^{1/2} \succ 0.$$

The conclusion follows from the fact that $A^{-1} - I_k \succ 0$, by hypothesis.

Now we shall use the above result to prove that $A - B \succ 0 \iff B^{-1} - A^{-1} \succ 0$.

($\Longrightarrow$) Observe that

$$A - B \succ 0 \implies A^{-1/2}(A - B)A^{-1/2} \succ 0 \implies I_k - A^{-1/2}BA^{-1/2} \succ 0.$$

From the previous result, this implies $A^{1/2}B^{-1}A^{1/2} - I_k \succ 0$. Therefore

$$B^{-1} - A^{-1} = A^{-1/2}A^{1/2}B^{-1}A^{1/2}A^{-1/2} - A^{-1/2}I_kA^{-1/2} = A^{-1/2}(A^{1/2}B^{-1}A^{1/2} - I_k)A^{-1/2} \succ 0.$$

($\Longleftarrow$) Since $B$ is symmetric, there exists $B^{1/2}$ symmetric such that $B = B^{1/2}B^{1/2}$. We have that $B^{-1} - A^{-1} \succ 0$, whence $B^{1/2}(B^{-1} - A^{-1})B^{1/2} \succ 0$. Thus $I_k - B^{1/2}A^{-1}B^{1/2} \succ 0$.

The result from the first part implies $B^{-1/2}AB^{-1/2} - I_k \succ 0$. Therefore

$$A - B = B^{1/2}B^{-1/2}AB^{-1/2}B^{1/2} - B^{1/2}I_kB^{1/2} = B^{1/2}(B^{-1/2}AB^{-1/2} - I_k)B^{1/2} \succ 0.$$

□

---

**3.** *Let $n_{ij}$ and $m_{ij}$ be the $(i, j)$ elements of $N = X(X'X)^{-1}X'$ and $M = I - N$.*

**(a)** *Show that $0 \leq n_{ii} \leq 1$ and $0 \leq m_{ii} \leq 1$.*

*Solution.* Let $e_i$ denote the $i$-th canonical vector. By spectral decomposition of $N$, we can write

$$n_{ii} = e_i'Ne_i = e_i'S\Lambda S'e_i = (S'e_i)'\Lambda S'e_i = v'\Lambda v,$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $S$ is orthogonal and $v \equiv S'e_i$. Here, we arrange eigenvalues in increasing order: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Observe that $v'v = e_i'SS'e_i = e_i'e_i = 1$, so by denoting $v_i$ as the $i$-th element of $v$, we can write

$$\lambda_1 = \lambda_1 v'v = \sum_{i=1}^n \lambda_1 v_i^2 \leq \underbrace{\sum_{i=1}^n \lambda_i v_i^2}_{v'\Lambda v} \leq \sum_{i=1}^n \lambda_n v_i^2 = \lambda_n v'v = \lambda_n.$$

Since $N$ is idempotent, all of its eigenvalues are either 0 or 1. Therefore $0 \leq n_{ii} \leq 1$. Since $m_{ii} = 1 - n_{ii}$, it also follows that $0 \leq m_{ii} \leq 1$.

An alternative one-line proof is

$$0 = \lambda_1 = \min_x \frac{x'Nx}{x'x} \leq \frac{\overbrace{e_i'Ne_i}^{n_{ii}}}{\underbrace{e_i'e_i}_{1}} \leq \max_x \frac{x'Nx}{x'x} = \lambda_n = 1.$$

Both inequalities follow from the Rayleigh quotient. □

**(b)** *Find all of the eigenvalues of $N$ and $M$.*

*Solution.* As argued in **(a)**, since $N$ is idempotent, all of its eigenvalues are either 0 or 1. The same holds for $M$, as it is also idempotent. Here I shall prove this result. Let $A$ be any idempotent matrix. By eigendecomposition, $A = H\Lambda H^{-1}$, whence

$$AA = H\Lambda H^{-1}H\Lambda H^{-1} = H\Lambda\Lambda H^{-1} = H\Lambda^2 H^{-1}.$$

Therefore $\Lambda = \Lambda^2$, and hence $\lambda_i = \lambda_i^2$ for all $i = 1, \ldots, n$. Thus each $\lambda_i$ must be equal to either zero or one. □

**(c)** *Interpret geometrically the vectors $Ny$ and $My$.*

*Solution.* Let $Y = X\beta + e$. Observe that

$$NY = X(X'X)^{-1}X'Y = X\hat{\beta}_{OLS} = \hat{Y}.$$

$$\text{and} \quad MY = (I - X(X'X)^{-1}X')Y = Y - X(X'x)^{-1}X'Y = Y - X\hat{\beta}_{OLS} = \hat{e}.$$

Therefore

$$NY + MY = \hat{Y} + \hat{e}. \quad (= Y)$$

Observe that $NY + MY = (N + M)Y = IY = Y$, so $\hat{Y} = NY$ is the "part" of $Y$ that is in the column space of $X$, while $\hat{e} = MY$ is the "part" of $Y$ that is orthogonal to the column space of $X$. To visualize, examine Figure 1. This displays the case $n = 3$ and $k = 2$. Displayed are three vectors $Y$, $X_1$, and $X_2$, which are each elements of $\mathbb{R}^3$. The plane created by $X_1$ and $X_2$ is the column space of $X$. Regression-fitted values are linear combinations of $X_1$ and $X_2$ and so lie on this plane. The fitted value $\hat{Y}$ is the vector on this plane closest to $Y$. The residual $\hat{e} = Y - \hat{Y}$ is the difference between the two. The angle between the vectors $\hat{Y}$ and $\hat{e}$ is 90°, and therefore they are orthogonal as shown.

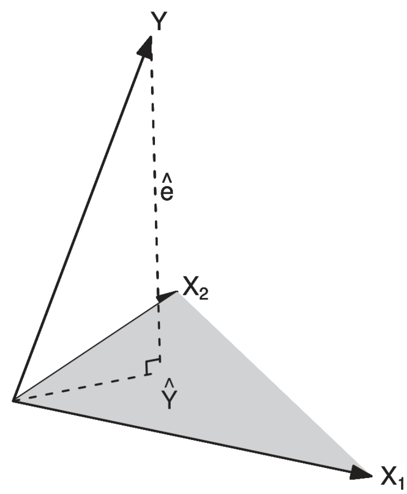


Figure 1: Projection of $Y$ onto $X_1$ and $X_2$.

□

**(d)** *Show that the null space of $N$ is the column space of $M$.*

*Solution.* Let $v \in \mathbb{R}^n$. If $v$ is in the column space of $M$, then $Mx = v$ for some $x$. Hence $Nv = NMx = 0$. Thus $v$ is in the null space of $N$. Conversely, if $v$ is in the null space of $N$, then $Nv = 0$ and hence $-Nv = 0$, whence $v - Nv = v$. Thus $(I - N)v = v$. Therefore $Mv = v$, which means that $v$ is in the column space of $M$. □

---

**4.** *Suppose that the $n \times k$ matrix $X = (X_1, X_2)$ has full column rank. Let $X_2^* = M_1 X_2$ be the $n \times k_2$ matrix of residuals from the auxiliary regression of $X_2$ on $X_1$. Show the following:*

**(a)** *rank*$(X_2^*) = k_2$.

*Solution.* Being the annihilator matrix associated with $X_1$, $M_1$ has rank $n - k_1$. Since $X$ is full column rank, so is $X_2$. Therefore $X_2$ has rank $k_2$. It follows that

$$\text{rank}(X_2^*) \leq \min\{\text{rank}(M_1), \text{rank}(X_2)\} = \min\{n - k_1, k_2\}.$$

If $k_2 > n - k_1$, then $k_1 + k_2 = k > n$, contradicting $X$ being full column rank. Thus we must have $k_2 \leq n - k_1$, so

$$\text{rank}(X_2^*) \leq k_2.$$

Suppose by contradiction $\text{rank}(X_2^*) < k_2$. Then $X_2^*$ is not full rank and hence there exists $\alpha = (\alpha_1, \cdots, \alpha_{k_2}) \in \mathbb{R}^{k_2} \backslash \{\mathbf{0}\}$ such that

$$X_2^* \alpha = 0 \iff M_1 X_2 \alpha = 0 \iff X_2 \alpha = N_1 X_2 \alpha.$$

Since $N_1 X_2 \alpha$ is a projection of $X_2\alpha$ onto the column space of $X_1$, there exists $w \in \mathbb{R}^{k_1}$ such that $X_1 w = N_1 X_2 \alpha$. Therefore

$$X_2\alpha = X_1 w \iff X_1 w - X_2\alpha = 0 \iff \underbrace{\begin{bmatrix} X_1 & X_2 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} w \\ -\alpha \end{bmatrix}}_{\equiv v} = 0 \iff Xv = 0.$$

But since $v = (w, -\alpha) \neq 0$ (recall that $\alpha \neq 0$), this contradicts $X$ being full rank. □

**(b)** *$N_X - N_1$ is symmetric and idempotent.*

*Solution.* For symmetry, just notice that

$$\begin{aligned} [N_X - N_1]' &= [X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1']' \\ &= [X(X'X)^{-1}X']' - [X_1(X_1'X_1)^{-1}X_1']' \\ &= X(X'X)^{-1}X' - X_1(X_1'X_1)^{-1}X_1' = N_X - N_1. \end{aligned}$$

For idempotency, write

$$N_X - N_1 = N_X - I + I - N_1 = I - N_1 - (I - N_X) = M_1 - M_X$$

and notice that

$$\begin{aligned} (M_1 - M_X)^2 &= M_1^2 - M_1 M_X - M_X M_1 + M_X^2 \\ \text{(Idempotency)} &= M_1 - M_1 M_X - M_X M_1 + M_X \\ &= M_1 - M_X. \end{aligned}$$

The last equality follows from the fact that

$$M_1 M_X = M_X M_1 = M_X.$$

Indeed, since $M_X X_1 = \mathbf{0}_{n\times k_1}$,[2]

$$M_X M_1 = M_X - M_X X_1 (X_1'X_1)^{-1} X_1' = M_X,$$

and

$$\begin{aligned} M_1 M_X &= M_X - X_1(X_1'X_1)^{-1}X_1' M_X \\ &= M_X - X_1(X_1'X_1)^{-1}(M_X' X_1)' \\ &= M_X - X_1(X_1'X_1)^{-1}(M_X X_1)' = M_X. \end{aligned}$$

□

[2] $X_1$ is in the column space of $X$, so this is expected. Observe that

$$MX = M[X_1 : X_2] = [MX_1 : MX_2] = \mathbf{0}_{n\times k} \iff MX_1 = 0_{n\times k_1} \text{ and } MX_2 = \mathbf{0}_{n\times k_2}.$$

**(c)** *$N_X$ - $N_1 = N_{X_2^*}$, that is, the orthogonal projection on the span of $X_2^*$.*

*Solution.* Define

$$A \equiv \begin{bmatrix} I_{k_1} & -(X_1'X_1)^{-1}X_1'X_2 \\ \mathbf{0} & I_{k_2} \end{bmatrix}$$

and consider the orthogonalizing transformation $Z \equiv XA$. Observe that

$$Z(Z'Z)^{-1}Z' = XA(A'X'XA)^{-1}A'X' = XAA^{-1}(X'X)^{-1}A'^{-1}A'X' = X(X'X)^{-1}X' = N_X.$$

We have

$$\begin{aligned} Z = XA &= \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} I_{k_1} & -(X_1'X_1)^{-1}X_1'X_2 \\ \mathbf{0} & I_{k_2} \end{bmatrix} \\ &= \begin{bmatrix} X_1 & -X_1(X_1'X_1)^{-1}X_1'X_2 + X_2 \end{bmatrix} \\ &= \begin{bmatrix} X_1 & X_2 - N_1X_2 \end{bmatrix} \\ &= \begin{bmatrix} X_1 & M_1X_2 \end{bmatrix} = \begin{bmatrix} X_1 & X_2^* \end{bmatrix}, \end{aligned}$$

and hence

$$Z' = A'X' = \begin{bmatrix} X_1' \\ X_2^{*\prime} \end{bmatrix}.$$

Therefore

$$\begin{aligned} Z'Z = A'X'XA &= \begin{bmatrix} X_1' \\ X_2^{*\prime} \end{bmatrix} \begin{bmatrix} X_1 & X_2^* \end{bmatrix} \\ &= \begin{bmatrix} X_1'X_1 & X_1'X_2^* \\ X_2^{*\prime}X_1 & X_2^{*\prime}X_2^* \end{bmatrix} \\ &= \begin{bmatrix} X_1'X_1 & \mathbf{0} \\ \mathbf{0} & X_2^{*\prime}X_2^* \end{bmatrix}, \end{aligned}$$

where the last equality follows from the fact that $X_1$ is orthogonal to $X_2^*$. Indeed, $X_1'X_2^* = X_1'M_1X_2 = (M_1X_1)'X_2 = 0$. It follows that

$$(Z'Z)^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (X_2^{*\prime}X_2^*)^{-1} \end{bmatrix}$$

and thus

$$\begin{aligned} N_X = Z(Z'Z)^{-1}Z' &= \begin{bmatrix} X_1 & X_2^* \end{bmatrix} \begin{bmatrix} (X_1'X_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (X_2^{*\prime}X_2^*)^{-1} \end{bmatrix} \begin{bmatrix} X_1' \\ X_2^{*\prime} \end{bmatrix} \\ &= X_1(X_1'X_1)^{-1}X_1' + X_2^*(X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime} \\ &= N_1 + N_{X_2^*}, \end{aligned}$$

whence $N_X - N_1 = N_{X_2^*}$, as desired. □

---

**5.1.** [3.2, Hansen] *Consider the OLS regression of the $n \times 1$ vector $Y$ on the $n \times k$ matrix $X$. Consider an alternative set of regressors $X = XC$, where $C$ is a $k \times k$ non-singular matrix. Thus, each column of $Z$ is a mixture of some of the columns of $X$. Compare the OLS estimates and residuals from the regression of $Y$ on $X$ to the OLS estimates from the regression of $Y$ on $Z$.*

*Solution.* Let $\hat{\beta}$ and $\tilde{\beta}$ denote the OLS estimators of $Y$ on $X$ and $Y$ on $XC$, respectively. We have

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

and

$$\begin{aligned}\tilde{\beta} &= [(XC)'(XC)]^{-1}(XC)'Y \\ &= (C'X'XC)^{-1}C'X'Y \\ &= C^{-1}(X'X)^{-1}C'^{-1}C'X'Y \\ &= C^{-1}(X'X)^{-1}X'Y \\ &= C^{-1}\hat{\beta}.\end{aligned}$$

Let $M$ and $M_C$ denote the annihilator matrices associated with $X$ and $XC$, respectively. Observe that

$$\begin{aligned}M_C &= I_n - XC[(XC)'(XC)]^{-1}(CX)' \\ &= I_n - XC(C'X'XC)^{-1}C'X' \\ &= I_n - XCC^{-1}(X'X)^{-1}C'^{-1}C'X' \\ &= I_n - X(X'X)^{-1}X' = M.\end{aligned}$$

Therefore $\tilde{u} = M_C Y = MY = \hat{u}$; that is, the residuals of both regressions are equal. □

---

**5.2.** [3.12, Hansen] *A dummy variable takes on only the values 0 and 1. It is used for categorical variables. Let $D_1$ and $D_2$ be vectors of 1's and 0's, with the $i^{th}$ element of $D_1$ equaling 1 and that of $D_2$ equaling 0 if the person is a man, and the reverse if the person is a woman. Suppose that there are $n_1$ men and $n_2$ women in the sample. Consider fitting the following three equations by OLS*

$$Y = \mu + D_1\alpha_1 + D_2\alpha_2 + e \tag{1}$$

$$Y = D_1\alpha_1 + D_2\alpha_2 + e \tag{2}$$

$$Y = \mu + D_1\phi + e \tag{3}$$

*Can all three equations (1), (2), and (3) be estimated by OLS? Explain if not.*

*Solution.* Equation (1) cannot be estimated by OLS. Notice that $\mathbf{1}_n = D_1 + D_2$, so we have perfect multicollinearity. This implies that the design matrix $X \equiv \begin{bmatrix} \mathbf{1}_n & D_1 & D_2 \end{bmatrix}$ is not full rank and hence the moment matrix $X'X$ is not invertible. Therefore, the OLS estimator $(X'X)^{-1}X'Y$ is not well-defined. Equations (2) and (3) can be estimated by OLS, since $\begin{bmatrix} D_1 & D_2 \end{bmatrix}$ and $\begin{bmatrix} \mathbf{1}_n & D_1 \end{bmatrix}$ are both full rank. □

**(a)** *Compare regressions (2) and (3). Is one more general than the other? Explain the relationship between the parameters in (2) and (3).*

*Solution.* No. Equations (2) and (3) convey the same information. Indeed, since $D_2 = 1 - D_1$, we can rewrite equation (2) as $Y = D_1\alpha_1 + (1-D_1)\alpha_2 + e = \alpha_2 + D_1(\alpha_1 - \alpha_2) + e$. Therefore we have the relationship $\mu = \alpha_2$ and $\phi = \alpha_2 - \alpha_1$. Note that this is a one-to-one relationship. □

**(b)** *Compute $\mathbf{1}_n' D_1$ and $\mathbf{1}_n' D_2$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones.*

*Solution.* Let $d_{1i}$ and $d_{2i}$ denote the $i^{th}$ elements of $D_1$ and $D_2$, respectively. Since the $i^{th}$ element of $D_1$ is one only if the person is a man and there are $n_1$ men in the sample,

$$\mathbf{1}_n' D_1 = \sum_{i=1}^{n} d_{1i} = n_1.$$

Analogously, since the $i^{th}$ element of $D_2$ is one only if the person is a woman and there are $n_2$ women in the sample,

$$\mathbf{1}_n' D_2 = \sum_{i=1}^{n} d_{2i} = n_2.$$

□

---

**8.** [4.14, Hansen] *Take a regression model $Y = X\beta + e$ with $\mathbb{E}[e|X] = 0$ and i.i.d. observations $(Y_i, X_i)$ and scalar $X$. The parameter of interest is $\theta = \beta^2$. Consider the OLS estimators $\hat{\beta}$ and $\hat{\theta} = \hat{\beta}^2$.*

**(a)** *Find $\mathbb{E}[\hat{\theta}|X]$ using our knowledge of $\mathbb{E}[\hat{\beta}|X]$ and $V_{\hat{\beta}} = V[\hat{\beta}|X]$. Is $\hat{\theta}$ biased for $\theta$?*

*Solution.* We have that $V[\hat{\beta}|X] = \mathbb{E}[\hat{\beta}^2|X] - \mathbb{E}[\hat{\beta}|X]^2 = \mathbb{E}[\hat{\theta}|X] - \beta^2 = \mathbb{E}[\hat{\theta}|X] - \theta$, whence it follows that $\mathbb{E}[\hat{\theta}|X] = \theta + V_{\hat{\beta}} \neq \theta$. Therefore $\hat{\theta}$ is biased for $\theta$. □

**(b)** *Suggest an (approximate) biased-corrected estimator $\hat{\theta}^*$ using an estimator $\hat{V}_{\hat{\beta}}$ for $V_{\hat{\beta}}$.*

*Solution.* $\hat{\theta}^* = \hat{\theta} - \hat{V}_{\hat{\beta}}$, where $\hat{V}_{\hat{\beta}}$ is some estimator for the covariance matrix $V_{\hat{\beta}}$. Recall that $V_{\hat{\beta}} = (X'X)^{-1}(X'DX)(X'X)^{-1}$, where $D = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. If the squared errors were observable, we could define an ideal unbiased estimator for $V_{\hat{\beta}}$ as

$$\hat{V}_{\hat{\beta}}^{\text{ideal}} = (X'X)^{-1}(X'\tilde{D}X)(X'X)^{-1}$$

and obtain an *exact* biased-corrected estimator. Indeed,

$$\begin{aligned}\mathbb{E}[\hat{V}_{\hat{\beta}}^{\text{ideal}}|X] &= (X'X)^{-1}\left(\sum_{i=1}^{n} X_i X_i' \mathbb{E}[e_i^2|X]\right)(X'X)^{-1} \\ &= (X'X)^{-1}\left(\sum_{i=1}^{n} X_i X_i' \sigma_i^2\right)(X'X)^{-1} \\ &= (X'X)^{-1}(X'DX)(X'X)^{-1} = V_{\hat{\beta}}.\end{aligned}$$

But since the errors $e_i^2$ are unobserved, $\hat{V}_{\hat{\beta}}^{\text{ideal}}$ is not a feasible estimator, so we must resort to some approximation of $\hat{V}_{\hat{\beta}}^{\text{ideal}}$ — for example, by replacing the squared errors $e_i^2$ with the squared residuals $\hat{e}_i^2$. In this sense, the suggested estimator $\hat{\theta}^*$ is, in general, just an *approximate* biased-corrected estimator for $\theta$. □

**(c)** *For $\hat{\theta}^*$ to be potentially unbiased, which estimator for $V_{\hat{\beta}}$ is most appropriate? Under which conditions is $\hat{\theta}^*$ unbiased?*

*Solution.* The HC2 estimator $\hat{V}_{\hat{\beta}}^{\text{HC2}} = (X'X)^{-1}\left(\sum_{i=1}^{n}(1-h_{ii})^{-1}X_iX_i'\hat{e}_i^2\right)(X'X)^{-1}$ is most appropriate, as it is unbiased for $V_{\hat{\beta}}$ under conditional homoskedasticity while remaining a valid estimator under heteroskedasticity. You can check that under homoskedasticity $\mathbb{E}[\hat{V}_{\hat{\beta}}^{\text{HC2}}|X] = V_{\hat{\beta}}$ and hence, when considering the biased-corrected estimator in **(b)** with this covariance estimator, $\mathbb{E}[\hat{\theta}^*|X] = \theta$. □

---

**9.** [4.20, Hansen] *Take the model in vector notation*

$$\begin{aligned}Y &= X\beta + e \\ \mathbb{E}[e|X] &= 0 \\ \mathbb{E}[ee'|X] &= \Sigma.\end{aligned}$$

*Assume for simplicity that $\Sigma$ is known. Consider the OLS and GLS estimators $\hat{\beta} = (X'X)^{-1}X'Y$ and $\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$. Compute the (conditional) covariance between $\hat{\beta}$ and $\tilde{\beta}$: $\mathbb{E}[(\hat{\beta}-\beta)(\tilde{\beta}-\beta)'|X]$. Compute the (conditional) covariance for $\hat{\beta}-\tilde{\beta}$: $\mathbb{E}[(\hat{\beta}-\tilde{\beta})(\hat{\beta}-\beta)'|X]$.*

*Solution.* We know that $\hat{\beta}-\beta = (X'X)^{-1}X'e$ and $\tilde{\beta}-\beta = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}e$. Thus

$$\begin{aligned}\mathbb{E}[(\hat{\beta}-\beta)(\tilde{\beta}-\beta)'|X] &= \mathbb{E}[(X'X)^{-1}X'ee'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}|X] \\ &= (X'X)^{-1}X'\mathbb{E}[ee'|X]\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} \\ &= (X'X)^{-1}X'\Sigma\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} \\ &= (X'X)^{-1}X'X(X'\Sigma^{-1}X)^{-1} \\ &= (X'\Sigma^{-1}X)^{-1}.\end{aligned}$$

Furthermore, we know that

$$\mathbb{E}[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|X] = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

and

$$\mathbb{E}[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'|X] = (X'\Sigma^{-1}X)^{-1}.$$

Thus

$$\begin{aligned}
\mathbb{E}[(\hat{\beta} - \tilde{\beta})(\hat{\beta} - \tilde{\beta})'|X] &= \mathbb{E}[((\hat{\beta} - \beta) - (\tilde{\beta} - \beta))((\hat{\beta} - \beta) - (\tilde{\beta} - \beta))'|X] \\
&= \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X] + \mathbb{E}[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'|X] \\
&- \mathbb{E}[(\hat{\beta} - \beta)(\tilde{\beta} - \beta)'|X] - \mathbb{E}[(\tilde{\beta} - \beta)(\hat{\beta} - \beta)'|X] \\
&= (X'X)^{-1}X'\Sigma X(X'X)^{-1} + (X'\Sigma^{-1}X)^{-1} \\
&- (X'\Sigma^{-1}X)^{-1} - (X'\Sigma^{-1}X)^{-1} \\
&= (X'X)^{-1}X'\Sigma X(X'X)^{-1} - (X'\Sigma^{-1}X)^{-1} \\
&= V[\hat{\beta}] - V[\tilde{\beta}].
\end{aligned}$$

□

---

**10.** *Consider the model $y = x_1\beta_1 + x_2\beta_2 + u$, where $\beta_1$ and $\beta_2$ are scalars, $x_1$ and $x_2$ are fixed vectors, and $u \sim N(0, \sigma^2 I_n)$. Let $\hat{\beta}_2$ be the OLS estimator for $\beta_2$. Compute $V[\hat{\beta}_2]$ and show what happens when $(x_1'x_2)^2 \to \|x_1\|^2 \cdot \|x_2\|^2$. Comment on this result.*

*Solution.* By FWL we have

$$\hat{\beta}_2 = (x_2'M_1x_2)^{-1}x_2'M_1y = \beta_2 + (x_2'M_1x_2)^{-1}x_2'M_1u.$$

Taking the variance we obtain

$$\begin{aligned}
V[\hat{\beta}_2] &= (x_2'M_1x_2)^{-1}x_2'M_1(\sigma^2 I_n)M_1x_2(x_2'M_1x_2)^{-1} = (x_2'M_1x_2)^{-1}\sigma^2 \\
&= \left[x_2'(I_n - x_1(x_1'x_1)^{-1}x_1')x_2\right]^{-1}\sigma^2 = \left[x_2'x_2 - x_2'x_1(x_1'x_1)^{-1}x_1'x_2\right]^{-1}\sigma^2 \\
&= \left[x_2'x_2 - \frac{(x_1'x_2)^2}{(x_1'x_1)}\right]^{-1}\sigma^2 = \left[\|x_2\|^2 - \frac{(x_1'x_2)^2}{\|x_1\|^2}\right]^{-1}\sigma^2.
\end{aligned}$$

As $(x_1'x_2)^2 \to \|x_1\|^2\|x_2\|^2$, $\|x_2\|^2 - \frac{(x_1'x_2)^2}{\|x_1\|^2} \to 0$ and hence $V[\hat{\beta}_2] \to \infty$. □

---

**11.** *Suppose two researchers are interested in the linear relation between the production of an agricultural product $y$ and fertilizer $z$. They have quarterly data on these variables from $m$ years and a total of $n = 4m$ observations. The researchers are concerned with seasonal patterns in these variables. Researcher John proposes that first each variable is deseasonalized in the following way: calculate the seasonal means $\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4$, and express each observation as a deviation from its seasonal mean: $y_{th}^* = y_{th} - \bar{y}_h$, where $y_{th}$ is the value of $y$ for year $t$ and quarter $h$, and $z_{th}^* = z_{th} - \bar{z}_h$, where $z_{th}$ is the value of $z$ for year $t$ and quarter $h$.*

*Then regress $y^*$ on $z^*$. On the other hand, researcher Robert propoes to regress $y$ on $X$ and $z$, where $X = (x_1, x_2, x_3, x_4)$ and $x_h$ is the $h^{th}$ quarter seasonal dummy*

$$x_h = \begin{cases} 1 & \textit{in quarter } h \\ 0 & \textit{otherwise} \end{cases}.$$

**(a)** *Show that the two competing methods proposed by John and Robert yield the same estimator for the fertilizer effect $z$ on $y$.*

*Solution.* Partition data in quarters: $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4)$ and $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4)$, where each $\mathbf{y}_h$ and $\mathbf{z}_h$, $h = 1, ..., 4$ is $m \times 1$. Let $\bar{\mathbf{z}}_h = \mathbf{1}\bar{z}_h$ and $\bar{\mathbf{y}}_h = \mathbf{1}\bar{y}_h$ for $h = 1, \ldots, h$, where $\mathbf{1}$ is a $m \times 1$ vector of ones. John estimates

$$\mathbf{y}^* = \mathbf{z}^*\beta + \mathbf{u},$$

or

$$\begin{bmatrix} \mathbf{y}_1 - \bar{\mathbf{y}}_1 \\ \mathbf{y}_2 - \bar{\mathbf{y}}_2 \\ \mathbf{y}_3 - \bar{\mathbf{y}}_3 \\ \mathbf{y}_4 - \bar{\mathbf{y}}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 - \bar{\mathbf{z}}_1 \\ \mathbf{z}_2 - \bar{\mathbf{z}}_2 \\ \mathbf{z}_3 - \bar{\mathbf{z}}_3 \\ \mathbf{z}_4 - \bar{\mathbf{z}}_4 \end{bmatrix} \beta + \mathbf{u}.$$

The OLS estimator for $\beta$ is given by

$$\hat{\beta} = (\mathbf{z}^{*\prime}\mathbf{z}^*)^{-1}\mathbf{z}^{*\prime}\mathbf{y}^*.$$

Robert estimates

$$\mathbf{y} = \mathbf{z}\beta_1 + X\beta_2 + \mathbf{u}.$$

By FWL, $\hat{\beta}_1 = (\mathbf{z}'M_X\mathbf{z})^{-1}\mathbf{z}'M_X y$, where $M_X = I - X(X'X)^{-1}X'$. Partition

$$X = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix},$$

where $\mathbf{1}$ and $\mathbf{0}$ are $m \times 1$ vectors of ones and zeros, respectively. Observe that

$$\begin{aligned} X'X &= \begin{bmatrix} \mathbf{1}' & \mathbf{0}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{1}' & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{1}' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{1}' \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{0}'\mathbf{0} & \mathbf{0}'\mathbf{0} & \mathbf{0}'\mathbf{0} \\ \mathbf{0}'\mathbf{0} & \mathbf{1}'\mathbf{1} & \mathbf{0}'\mathbf{0} & \mathbf{0}'\mathbf{0} \\ \mathbf{0}'\mathbf{0} & \mathbf{0}'\mathbf{0} & \mathbf{1}'\mathbf{1} & \mathbf{0}'\mathbf{0} \\ \mathbf{0}'\mathbf{0} & \mathbf{0}'\mathbf{0} & \mathbf{0}'\mathbf{0} & \mathbf{1}'\mathbf{1} \end{bmatrix} = \begin{bmatrix} m & 0 & 0 & 0 \\ 0 & m & 0 & 0 \\ 0 & 0 & m & 0 \\ 0 & 0 & 0 & m \end{bmatrix} = I_4 m. \end{aligned}$$

Furthermore,

$$XX' = \begin{bmatrix} \mathbf{1}\mathbf{1}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' \\ \mathbf{0}\mathbf{0}' & \mathbf{1}\mathbf{1}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' \\ \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{1}\mathbf{1}' & \mathbf{0}\mathbf{0}' \\ \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{1}\mathbf{1}' \end{bmatrix}.$$

Therefore

$$\begin{aligned} M_X\mathbf{z} &= \mathbf{z} - m^{-1}XX'\mathbf{z} \\ &= \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \mathbf{z}_4 \end{bmatrix} - m^{-1} \begin{bmatrix} \mathbf{1}\mathbf{1}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' \\ \mathbf{0}\mathbf{0}' & \mathbf{1}\mathbf{1}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' \\ \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{1}\mathbf{1}' & \mathbf{0}\mathbf{0}' \\ \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{0}\mathbf{0}' & \mathbf{1}\mathbf{1}' \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \mathbf{z}_4 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \mathbf{z}_4 \end{bmatrix} - \begin{bmatrix} m^{-1}\mathbf{1}\mathbf{1}'\mathbf{z}_1 \\ m^{-1}\mathbf{1}\mathbf{1}'\mathbf{z}_2 \\ m^{-1}\mathbf{1}\mathbf{1}'\mathbf{z}_3 \\ m^{-1}\mathbf{1}\mathbf{1}'\mathbf{z}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \mathbf{z}_4 \end{bmatrix} - \begin{bmatrix} \mathbf{1}m^{-1}\sum_{j=1}^m z_{j,1} \\ \mathbf{1}m^{-1}\sum_{j=1}^m z_{j,2} \\ \mathbf{1}m^{-1}\sum_{j=1}^m z_{j,3} \\ \mathbf{1}m^{-1}\sum_{j=1}^m z_{j,4} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \mathbf{z}_4 \end{bmatrix} - \begin{bmatrix} \mathbf{1}\bar{z}_1 \\ \mathbf{1}\bar{z}_2 \\ \mathbf{1}\bar{z}_3 \\ \mathbf{1}\bar{z}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \mathbf{z}_4 \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \\ \bar{\mathbf{z}}_3 \\ \bar{\mathbf{z}}_4 \end{bmatrix} = \mathbf{z}^*, \end{aligned}$$

where by $z_{j,h}$ we denote the year-$j$ quarter-$h$ observation. Analogously, $M_X\mathbf{y} = \mathbf{y}^*$. Therefore

$$\begin{aligned} \hat{\beta}_1 &= (\mathbf{z}'M_X\mathbf{z})^{-1}\mathbf{z}'M_X\mathbf{y} \\ &= (\mathbf{z}'M_XM_x\mathbf{z})^{-1}\mathbf{z}'M_XM_X\mathbf{y} \\ &= [(M_X\mathbf{z})'M_x\mathbf{z}]^{-1}(M_X\mathbf{z})'M_X\mathbf{y} \\ &= (\mathbf{z}^{*\prime}\mathbf{z}^*)^{-1}\mathbf{z}^{*\prime}\mathbf{y}^* = \hat{\beta}, \end{aligned}$$

whence it follows that the two competing methods proposed by John and Robert yield the same estimator for the effect of $z$ on $y$. □

**(b)** *Mark instead suggests the intercept could be relevant, therefore they should regress on* $\boldsymbol{1} = (1, \ldots, 1)'$, *$X$ and $Z$. The researchers comment this would not be a good idea. Why?*

*Solution.* Notice that $\mathbf{1}_{4m} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_4$, so if we were to add an intercept we would have perfect multicollinearity. This would imply that the full design matrix $D \equiv \begin{bmatrix} \mathbf{1}_{4m} & \mathbf{z} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{bmatrix}$ would not be full rank and hence the moment matrix $D'D$ would not be invertible. Therefore, the OLS estimator $(D'D)^{-1}D'\mathbf{y}$ would not be well-defined. □

**(c)** *Now Mark and Robert are interested in testing the hypothesis that there is no seasonal pattern in the data. Mark proposes to regress $y$ on $X$ only and test if the coefficient of $X$ equals to zero. Robert, however, proposes to regress $y$ on $X$ and $z$ and test if the coefficient of $X$ equals to zero. Which of the two methods would you choose? Explain your answer.*

*Solution.* As $\mathbf{z}$ is relevant for explaining $\mathbf{y}$, omitting it from the regression would imply omitted variable bias. So, in principle, I would prefer Robert's suggestion. We must, however, always be aware of the bias-variance tradeoff when deciding whether or not to include new regressors in a model. □

**(d)** *The research assistant loads the database on Stata. The database contains the variables $y$, $x_1$, $x_2$, $x_3$, $x_4$, $z$ described above. Write the commands he would use to run the regression of $y$ on $x_1$, $x_2$, $x_3$, $x_4$, $z$.*

*Solution.* `regress y x1 x2 x3 x4 z, noconstant` □

---

**12.** *For each of the following statements, indicate whether it is true or false, and justify your answer.*

**(a)** *The random variable $t = b'b$ is an unbiased estimator of the parameter $\beta'\beta$.*

*Solution.* False. $\mathbb{E}[b'b] > \mathbb{E}[b]'\mathbb{E}[b] = \beta'\beta$, by Jensen's inequality and unbiasedness of $b$. □

**(b)** *Since $\hat{y} = Ny$, it follows that $y = N^{-1}\hat{y}$.*

*Solution.* False. $N$ has dimensions $n \times n$ and rank $k$. Therefore it is not full rank and hence is not invertible. □

**(c)** *Since $\mathbb{E}[\hat{y}] = \mathbb{E}[y]$, the sum of the residuals is 0.*

*Solution.* False. The sum of the residuals can be written as

$$\mathbf{1}_n' My = (M\mathbf{1}_n)'y,$$

where $\mathbf{1}_n$ is a $n$-dimensional vector of ones. If $\mathbf{1}_n$ is not in the column space of $X$, which is something that could perfectly well happen even though $\mathbb{E}[\hat{y}] = \mathbb{E}[y]$, the above expression may not be zero. One could easily construct several numerical examples. Take, for instance, $X = (2, 1)$ and $Y = (4, 7)$. We have $\hat{\beta} = (X'X)^{-1}X'y = 3$ and $\hat{u} = Y - X\hat{\beta} = (-2, 4)$. Therefore $\sum_{i=1}^{2} \hat{u}_i = 2 \neq 0$. If, however, $\mathbf{1}_n$ is in the column space of $X$ (which is the case when the regression contains a constant regressor, for example), then $M\mathbf{1}_n = 0$ and therefore the sum of the residuals is necessarily zero. □

**(d)** *If $b_1$ and $b_2$ are the first two elements of $b$, $t_1 = b_1 + b_2$, and $t_2 = b_1 - b_2$, then $V(t_1) \geq V(t_2)$.*

*Solution.* False. Consider the $k = 2$ case. We have that $V[t_1] = V[b_1] + V[b_2] + 2Cov(b_1, b_2)$ and $V[t_2] = V[b_1] + V[b_2] - 2Cov(b_1, b_2)$. Therefore $V[t_1] - V[t_2] = 4Cov(b_1, b_2)$. One could easily find an example in which $Cov(b_1, b_2) < 0$. Take for instance $X = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ and assume $u \sim N(0, I_2)$. We have

$$(X'X)^{-1}X' = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix},$$

and hence

$$b = \beta + (X'X)^{-1}X'u = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Therefore,

$$b_1 = \beta_1 + u_1 - u_2 \quad \text{and } b_2 = \beta_2 - u_1 + 2u_2.$$

It follows that $V[b_1] = 2$, $V[b_2] = 5$, $V[t_1] = V[\beta_1 + \beta_2 + u_2] = 1$, and $V[t_2] = V[\beta_1 - \beta_2 + 2u_1 - 3u_2] = 13$. Therefore $V[t_2] > V[t_1]$. Note that $Cov(b_1, b_2) = -3$. □

**(e)** *The LS coefficients $b = Ay$ are uncorrelated with the residuals $e = My$.*

*Solution.* In general, this statement is false. However, if we impose homoskedasticity, it becomes true. Let $A \equiv (X'X)^{-1}X'$. Assume $X$ to be nonstochastic and let $\Sigma$ denote the covariance matrix of $u$. Observe that $b - \beta = Au$ and $\mathbb{E}[e] = 0$. Therefore, the covariance between $b$ and $e$ is given by

$$\begin{aligned}
\mathbb{E}[(b-\beta)(e-\mathbb{E}[e])'] &= \mathbb{E}[Aue'] = \mathbb{E}[Auu'M'] \\
&= A\mathbb{E}[uu']M \\
&= (X'X)^{-1}X'\Sigma M \\
&= (X'X)^{-1}X'\Sigma(I_n - X(X'X)^{-1}X') \\
&= (X'X)^{-1}X'\Sigma - (X'X)^{-1}X'\Sigma X(X'X)^{-1}X' \neq 0.
\end{aligned}$$

If we impose homoskedasticity, then $\Sigma = I_n\sigma^2$ and hence

$$\begin{aligned}
\mathbb{E}[(b-\beta)(e-\mathbb{E}[e])'] &= (X'X)^{-1}X'\sigma^2 - (X'X)^{-1}X'X(X'X)^{-1}X'\sigma^2 \\
&= (X'X)^{-1}X'\sigma^2 - (X'X)^{-1}X'\sigma^2 = 0.
\end{aligned}$$

□

---

**13.** *Prove Theorem 4.6 of Hansen (2021).*

*Solution.* Different versions of Hansen's book have different Theorems 4.6. If you considered the MSFE theorem, the proof is in the text, just above the statement. If you've considered the Generalized Gauss-Markov Theorem, see the Technical Proofs section of the most recent versions of the book (for example, the 2022 version). It seems that Theorem 4.5 of the newer versions is equivalent to Theorem 4.6 of the older versions of the book. □

---

**14.** *Prove Theorems 5.4 and 5.7 of Hansen (2021).*

*Solution.* Check Hansen's book. The proofs are in the text, just above the statements. □

---