Statistics II | **Problem Set 3**
Professor: Marcelo J. Moreira | **Solutions**
TA: Luan Borelli | **August 31, 2023**

---

1. *We have a regressor x_i, taking values $0, 1, \ldots, 10$, say, years of education. The outcome is a continuous variable y_i, say, logarithm of income. We have a sample of n indiviuals where (y_i, x_i) are i.i.d. Assume all necessary conditional and unconditional moments exist.*

(a) *Write $y_i = \mathbb{E}[y_i|x_i] + u_i$, where $u_i \equiv y_i - \mathbb{E}[y_i|x_i]$. Show that $\mathbb{E}[u_i|x_i] = 0$.*

Solution.

$$\mathbb{E}[u_i|x_i] = \mathbb{E}[y_i - \mathbb{E}[y_i|x_i] \mid x_i] = \mathbb{E}[y_i \mid x_i] - \mathbb{E}[\mathbb{E}[y_i|x_i] \mid x_i] = \mathbb{E}[y_i \mid x_i] - \mathbb{E}[y_i \mid x_i] = 0.$$

□

(b) *Must u_i be conditionally homoskedastic, i.e., is $V[u_i|x_i]$ the same regardless of x_i? Explain your answer.*

Solution. No. $V[u_i|x_i] = \mathbb{E}[y_i^2|x_i] - \mathbb{E}[y_i|x_i]^2 = V[y_i|x_i]$. There is no reason for the variance of y_i (and hence of u_i) to be the same across every possible group (i.e., across every possible value $x_i \in \{0, 1, \ldots, 10\}$ can take). Although (y_i, x_i) are i.i.d., it is not necessarily true that $y_i|x_i$ and $y_i|x_j$ are i.i.d for all i, j. In particular, it is indeed true that $V[y_i|x_i] = V[y_i|x_j]$ if $x_i = x_j$. However, if $x_i \neq x_j$ it is perfectly possible that $V[y_i|x_i] \neq V[y_i|x_j]$. Notice that such (potential) heteroskedasticity arises precisely from conditioning. Indeed, (y_i, x_i) being i.i.d. implies that u_i is *unconditionally* homoskedastic, but not necessarily *conditionally* homoskedastic. □

(c) *We want to compute how much, on average, the outcome increases when the regressor increases from 5 to 6 years of schooling:*

$$\theta_{5,6} = \mathbb{E}[y_i|x_i = 6] - \mathbb{E}[y_i|x_i = 5].$$

Find the conditional expectation and variance of

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Solution.

$$\mathbb{E}[\hat{\theta}|X] = \frac{\sum_{i=1}^n x_i \mathbb{E}[y_i|x_i]}{\sum_{i=1}^n x_i^2} \quad \text{and } V[\hat{\theta}|X] = \frac{\sum_{i=1}^n x_i^2 V[y_i|x_i]}{\left(\sum_{i=1}^n x_i^2\right)^2}.$$

Notice that without introducing further structure we cannot go beyond this point. Moving the expectations and variances out of the summations would be incorrect, as per item **(a)**, we cannot ensure that they are equal for all i. □

(d) *Consider the estimator*

$$\hat{\theta}_{5,6} = \frac{\sum_{i\in\{i:x_i=6\}} y_i}{n_6} - \frac{\sum_{i\in\{i:x_i=5\}} y_i}{n_5} = \frac{\sum_{i=1}^n y_i I(x_i=6)}{\sum_{i=1}^n I(x_i=5)} - \frac{\sum_{i=1}^n}{\sum_{i=1}^n I(x_i=5)},$$

where n_5 and n_6 denote the number of individuals in the sample taking values $x_i = 5$ and $x_i = 6$, respectively. Find the conditional expectation and variance of $\hat{\theta}_{5,6}$.

Solution. For $j = 5, 6$, observe that

$$\begin{aligned}
\mathbb{E}\left[\left.\frac{\sum_{i=1}^n y_i I(x_i=j)}{\sum_{i=1}^n I(x_i=j)}\right| X\right] &= \frac{\sum_{i=1}^n I(x_i=j)\mathbb{E}[y_i|x_i]}{\sum_{i=1}^n I(x_i=j)} \\
&= \frac{\sum_{i=1}^n I(x_i=j)\int_{-\infty}^{\infty} y_i f_{y|x}(y_i|x=x_i)\; dy_i}{\sum_{i=1}^n I(x_i=j)} \\
&= \frac{\sum_{i\in\{i:x_i=j\}}^n I(x_i=j)\int_{-\infty}^{\infty} y_i f_{y|x}(y_i|x=j)\; dy_i}{\sum_{i\in\{i:x_i=j\}}^n I(x_i=j)} \\
&= \frac{\sum_{i\in\{i:x_i=j\}}^n I(x_i=j)\mathbb{E}[y_i|x_i=j]}{\sum_{i\in\{i:x_i=j\}}^n I(x_i=j)} \\
&= \mathbb{E}[y_i|x_i=j],
\end{aligned}$$

where the last equality follows from the i.i.d. property of (y_i, x_i) and the fact that all x_i involved in the summation are conditioned to the same value j, so we can move the expectation out of the summation and cancel out the sums.

It follows that

$$\mathbb{E}[\hat{\theta}_{5,6}|X] = \mathbb{E}[y_i|x_i=6] - \mathbb{E}[y_i|x_i=5] = \theta_{5,6}.$$

That is: $\hat{\theta}_{5,6}$ is a conditionally unbiased estimator for θ *regardless of the functional form of the conditional expectation function* $m(x_i) = \mathbb{E}[y_i|x_i]$.

Similar calculations for the variance yield

$$V[\hat{\theta}_{5,6}|X] = n_6^{-1}V[y_i|x_i=6] + n_5^{-1}V[y_i|x_i=5].$$

An alternative way of seeing this is by looking at the first representation of $\hat{\theta}_{5,6}$. Notice that (y_i, x_i)'s being i.i.d. implies that y_i's are also i.i.d. and that the sets $\{i : x_i = 6\}$ and $\{i : x_i = 5\}$ are disjoint, so we can write the variance of the sum as the sum of variances:

$$\begin{aligned}
V(\hat{\theta}_{5,6}) &= \frac{\sum_{i\in\{i:x_i=6\}} V[y_i|x_i]}{n_6^2} - \frac{\sum_{i\in\{i:x_i=5\}} V[y_i|x_i]}{n_5^2} \\
&= \frac{n_6 V[y_i|x_i=6]}{n_6^2} - \frac{n_5 V[y_i|x_i=5]}{n_5^2} \\
&= n_6^{-1}V[y_i|x_i=6] + n_5^{-1}V[y_i|x_i=5].
\end{aligned}$$

The second equality follows from the fact that in each summation all terms are being conditioned to the same value of x_i, whence by identicality of (y_i, x_i) all variances in each summation must be equal, as argued in item **(b)**. The same works for the expectation. □

(e) *Compare the MSE of $\hat{\theta}$ and $\hat{\theta}_{5,6}$ when we are interested in $\theta_{5,6}$ if the effect is homogeneous and errors are homoskedastic.*

Solution. If the effect is homogeneous, we have $\mathbb{E}[y_i|x_i] = x_i\theta^*$ for some constant θ^*. Notice that in this case

$$\mathbb{E}[\hat{\theta}|X] = \mathbb{E}[\hat{\theta}_{5,6}|X] = \theta^*,$$

so both estimators are conditionally unbiased. Therefore, by the bias-variance decomposition, the (conditional) MSEs become simply the conditional variances of each estimator:

$$MSE(\hat{\theta}) = V[\hat{\theta}|X] \text{ and } MSE(\hat{\theta}_{5,6}) = V[\hat{\theta}_{5,6}|X].$$

Under (conditional) homoskedasticity, $V[\hat{\theta}|X] = \left(\sum_{i=1}^n x_i^2\right)^{-1}\sigma^2$. Observe that both $\hat{\theta}$ and $\hat{\theta}_{5,6}$ are linear estimators. Therefore the (classical) Gauss-Markov theorem applies and we must have $V[\hat{\theta}|X] \leq V[\hat{\theta}_{5,6}|X]$, whence it follows that $MSE(\hat{\theta}) \leq MSE(\hat{\theta}_{5,6})$. □

(f) *Compare the MSE under different departures from homogeneity and homoskedasticity.*

Solution. Under homogeneity and homoskedasticity, Gauss-Markov applies and $\hat{\theta}$ dominates $\hat{\theta}_{5,6}$, as demonstrated in item **(e)**. Under homogeneity and heteroskedasticity, both $\hat{\theta}$ and $\hat{\theta}_{5,6}$ are still conditionally unbiased and hence there is still no bias-variance tradeoff. But since errors are heteroskedastic, Gauss-Markov doesn't apply anymore. The relative MSE between the two estimators becomes ambiguous and depends on the variance structure of each estimator, which without further assumptions can be anything.

Under heterogeneity and homoskedasticity, $\hat{\theta}_{5,6}$ is still conditionally unbiased, but $\hat{\theta}$ is not necessarily conditionally unbiased anymore. A bias-variance tradeoff between $\hat{\theta}$ and $\hat{\theta}_{5,6}$ arises and the relative MSE between the two estimators mainly depends on the degree of departure from homogeneity, which governs the size of the bias of $\hat{\theta}$. Under heterogeneity and heteroskedasticity, anything goes: the relative MSE between $\hat{\theta}$ and $\hat{\theta}_{5,6}$ depends simultaneously on the variance structures of both estimators, which without further assumptions can be anything, and on the degree of departure from homogeneity governing the size of the bias of $\hat{\theta}$ (and hence the bias-variance tradeoff between the two estimators).

	Homoskedasticity	**Heteroskedasticity**
Homogeneity	$MSE(\hat{\theta}) \leq MSE(\hat{\theta}_{5,6})$, by Gauss-Markov theorem.	$MSE(\hat{\theta}) \lesseqgtr MSE(\hat{\theta}_{5,6})$, depending on the variance structure.
Heterogeneity	$MSE(\hat{\theta}) \lesseqgtr MSE(\hat{\theta}_{5,6})$, depending on the degree of departure from homogeneity.	$MSE(\hat{\theta}) \lesseqgtr MSE(\hat{\theta}_{5,6})$, depending on the variance structure *and* the degree of departure from homogeneity.

□

2. [7.2, LNs] *This question is on an application of the Cramér-Wold device, and uses the (multivariate) continuity theorem. Let X_n, $1 \leq n \leq \infty$, be random vectors with characteristic function $\varphi_n(t)$. (i) if $X_n \xrightarrow{d} X$ then $\varphi_n(t) \to \varphi(t)$ for all t; and, (ii) if $\varphi_n(t)$ converges pointwise to a limit $\varphi(t)$ that is continuous at zero, then $X_n \xrightarrow{d} X$ (with X having characteristic function $\varphi(t)$.*

(a) *Use characteristic functions to prove the Crámer-Wold device: a sequence of k-dimensional random vectors S_n, $n = 1, 2, \ldots$, converges in distribution to a random vector S if and only if $\alpha' S_n \xrightarrow{d} \alpha' S$ for every fixed vector $\alpha \neq 0$.*

Solution. If $S_n \xrightarrow{d} S$, then $\varphi_n(\tau) \to \varphi(\tau)$ for all $\tau \in \mathbb{R}^k$. In particular, we can let $\tau = \alpha t$ for any arbitrary $t \in \mathbb{R}$ and every fixed vector $\alpha \in \mathbb{R}^k \backslash \{0\}$ so that

$$\mathbb{E}[e^{it\alpha' S_n}] = \mathbb{E}[e^{i(\alpha t)' S_n}] = \varphi_n(\alpha t) \to \varphi(\alpha t) = \mathbb{E}[e^{i(\alpha t)' S}] = \mathbb{E}[e^{it\alpha' S}].$$

Observe that $\varphi_n(t\alpha)$ and $\varphi(t\alpha)$ are precisely the characteristic functions of $\alpha' S_n$ and $\alpha' S$, respectively, when viewed as functions of t only (i.e., given a fixed $\alpha \in \mathbb{R}^k$). Therefore $\alpha' S_n \xrightarrow{d} \alpha' S$. Conversely, if $\alpha' S_n \to \alpha' S$ for every fixed vector $\alpha \neq 0$, then

$$\mathbb{E}[e^{i\tau\alpha' S_n}] \to \mathbb{E}[e^{i\tau\alpha' S}] \quad \text{for all } \tau \in \mathbb{R}.$$

In particular, we can let $\tau = 1$ so that

$$\varphi_n(\alpha) = \mathbb{E}[e^{i\alpha' S_n}] \to \mathbb{E}[e^{i\alpha' S}] = \varphi(\alpha) \quad \text{for all } \alpha \in \mathbb{R}^k.$$

Observe that $\varphi_n(\alpha)$ and $\varphi(\alpha)$ are precisely the characteristic functions of S_n and S, respectively, as functions of $\alpha \in \mathbb{R}^k$. Therefore $S_n \xrightarrow{d} S$. □

(b) *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random vectors of dimension k with $\mathbb{E}[X_i] = \mu$ and variance $V(X_i) = \mathbb{E}[(X_i - \mu)(X_i - \mu)'] = \Sigma$. The variance Σ is positive definite ($a'\Sigma a > 0$ for any $a \neq 0$). Find the limiting distribution of $\sqrt{n}(\bar{X}_n - \mu)$.*

Solution. Let $S_n = \sqrt{n}(\bar{X}_n - \mu)$ and observe that, by the *univariate* central limit theorem, for any $\alpha \in \mathbb{R}^k \backslash \{0\}$ we have $\alpha' S_n = \sqrt{n}(\alpha' \bar{X}_n - \alpha' \mu) \xrightarrow{d} N(0, \alpha' \Sigma \alpha)$. Notice that $N(0, \alpha' \Sigma \alpha) = \alpha' N(0, \Sigma) =: \alpha' S$, where $S \sim N(0, \Sigma)$, whence it follows, by **(a)**, that $S_n \xrightarrow{d} S$. We have just proved the *multivariate* central limit theorem. □

3. [16.7, LNs] *Let X and y be*

$$X = [X_1, X_2] = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 3 \\ 1 & 5 \\ 1 & 2 \end{bmatrix} \quad \textit{and} \quad y = \begin{bmatrix} 14 \\ 17 \\ 8 \\ 16 \\ 3 \end{bmatrix}.$$

Calculate the following: **(a)** $Q = X'X$, $|X'X|$, *and* Q^{-1}.

Solution.

$$Q = \begin{bmatrix} 5 & 16 \\ 16 & 58 \end{bmatrix}, \quad |X'X| = 34, \text{ and } \quad Q^{-1} = \begin{bmatrix} \frac{29}{17} & -\frac{8}{17} \\ -\frac{8}{17} & \frac{5}{34} \end{bmatrix}.$$

□

(b) $A = Q^{-1}X'$ *and* $\hat{\beta} = Ay$.

Solution.

$$A = \begin{bmatrix} \frac{13}{17} & -\frac{3}{17} & \frac{5}{17} & -\frac{11}{17} & \frac{13}{17} \\ -\frac{3}{17} & \frac{2}{17} & -\frac{1}{34} & \frac{9}{34} & -\frac{3}{17} \end{bmatrix} \quad \text{and } \hat{\beta} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

□

(c) N *and* $\hat{y} = Ny$.

Solution.

$$N = \begin{bmatrix} \frac{7}{17} & \frac{1}{17} & \frac{4}{17} & -\frac{2}{17} & \frac{7}{17} \\ \frac{1}{17} & \frac{5}{17} & \frac{3}{17} & \frac{7}{17} & \frac{1}{17} \\ \frac{4}{17} & \frac{3}{17} & \frac{7}{34} & \frac{5}{34} & \frac{4}{17} \\ -\frac{2}{17} & \frac{7}{17} & \frac{5}{34} & \frac{23}{34} & -\frac{2}{17} \\ \frac{7}{17} & \frac{1}{17} & \frac{4}{17} & -\frac{2}{17} & \frac{7}{17} \end{bmatrix} \quad \text{and} \quad \hat{y} = \begin{bmatrix} 8 \\ 14 \\ 11 \\ 17 \\ 8 \end{bmatrix}.$$

□

(d) M *and* $e = My$.

Solution.

$$M = \begin{bmatrix} \frac{10}{17} & -\frac{1}{17} & -\frac{4}{17} & \frac{2}{17} & -\frac{7}{17} \\ -\frac{1}{17} & \frac{12}{17} & -\frac{3}{17} & -\frac{7}{17} & -\frac{1}{17} \\ -\frac{4}{17} & -\frac{3}{17} & \frac{27}{34} & -\frac{5}{34} & -\frac{4}{17} \\ \frac{2}{17} & -\frac{7}{17} & -\frac{5}{34} & \frac{11}{34} & \frac{2}{17} \\ -\frac{7}{17} & -\frac{1}{17} & -\frac{4}{17} & \frac{2}{17} & \frac{10}{17} \end{bmatrix} \quad \text{and} \quad e = \begin{bmatrix} 6 \\ 3 \\ -3 \\ -1 \\ -5 \end{bmatrix}.$$

□

(e) $tr(N)$ *and* $tr(M)$.

Solution.

$$tr(N) = 2 \quad \text{and} \quad tr(M) = 3.$$

□

(f) *$X_2^{*\prime}X_2^*$, $X_2^{*\prime}X_2$, $X_2^{*\prime}y^*$, and $X_2^{*\prime}y$.*

Solution.

$$X_2^{*\prime}X_2^* = \frac{34}{5}, \quad X_2^{*\prime}X_2 = \frac{34}{5}, \quad X_2^{*\prime}y^* = \frac{102}{5}, \quad \text{and} \quad X_2^{*\prime}y = \frac{102}{5}.$$

□

(g) *$(X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}y$, and compare your answer with item (b).*

Solution. As in item **(b)** we obtained $\hat{\beta}_2 = 3$, by the FWL theorem the answer must be 3.

$$(X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}y = 3.$$

□

---

4. [16.27, LNs] *Suppose that the classical normal regression model applies to $\mathbb{E}[y] = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4$. Let $w = y - x_4$, $z_1 = x_1$, $z_2 = x_2 - x_4$, $z_3 = x_3 - x_4$. For a sample of 104 firms, regression y on (x_1, x_2, x_3, x_4) gives 70 as the sum of squared residuals, while regression w on (z_1, z_2, z_3) gives 80 as the sum of squared residuals.*

(a) *Test at the 5% significance level the null hypothesis $\beta_1 + \beta_2 + \beta_4 = 1$ against the two-sided alternative $\beta_2 + \beta_3 + \beta_4 \neq 1$.*

Solution. We can write the model

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + u$$

as $y - x_4 = x_1\beta_1 + (x_2 - x_4)\beta_2 + (x_3 - x_4)\beta_3 + (\beta_2 + \beta_3 + \beta_4 - 1)x_4 + u$, or, equivalently,

$$w = z_1\beta_1 + z_2\beta_2 + z_3\beta_3 + \gamma x_4 + u,$$

where $\gamma \equiv \beta_2 + \beta_3 + \beta_4 - 1$. Under this equivalent formulation, testing $\beta_1 + \beta_2 + \beta_4 = 1$ against $\beta_1 + \beta_2 + \beta_4 \neq 1$ boils down to testing $\gamma = 0$ against $\gamma \neq 0$. The F-statistic for such a test can be written as

$$F = \frac{SSE(\tilde{\beta}_{CLS}) - SSE(\hat{\beta}_{OLS})}{s^2},$$

where $SSE(\beta) = \sum_{i=1}^n (y_i - X_i'\beta)^2$ denote the sum-of-squared errors, $\tilde{\beta}_{CLS}$ is the OLS coefficient obtained from regressing w on (z_1, z_2, z_3) only (i.e., the OLS coefficient obtained under the null restriction $\gamma = 0$), and $\hat{\beta}_{OLS}$ is the OLS coefficient obtained from regressing y on (x_1, x_2, x_3, x_4). At the 5% significance level, we reject the null hypothesis if $F > 3.84$, where 3.84 is the (approximate) 95% quantile of a $\chi^2(1)$ distribution.

We have $SSE(\tilde{\beta}_{CLS}) = 80$, $SSE(\hat{\beta}_{OLS}) = 70$, and $s^2 = 0.7$.[1] It follows that

$$F = \frac{80 - 70}{0.7} \approx 14.29 > 3.84.$$

Therefore, at the 5% significance level we reject the null of $\gamma = \beta_1 + \beta_2 + \beta_4 - 1 = 0$. □

(b) *Let $v = y - x_2$, $t_1 = x_1$, $t_2 = x_3 - x_2$, $t_3 = x_4 - x_2$. If v is regressed on (t_1, t_2, t_3), what sum of squared residuals will be obtained?*

Solution. Observe that $t_1 = z_1$, $t_2 = x_3 - x_4 + x_4 - x_2 = z_3 - z_2$, and $t_3 = -z_1$. Therefore

$$\begin{aligned}
& v = t_1\beta_1 + t_2\beta_2 + t_3\beta_2 + u \\
\iff & y - x_2 = z_1\beta_1 + (z_3 - z_2)\beta_2 - \beta_3 z_1 + u \\
\iff & y - x_4 = z_1\beta_1 + (z_3 - z_2)\beta_2 - \beta_3 z_1 + x_2 - x_4 + u \\
\iff & w = z_1\beta_1 + (z_3 - z_2)\beta_2 - z_1\beta_3 + z_1 + u \\
\iff & w = z_1(1 + \beta_1 - \beta_3) + (z_3 - z_2)\beta_2 + u \\
\iff & w = z_1\gamma_1 + (z_3 - z_2)\gamma_2 + u,
\end{aligned}$$

where $\gamma_1 \equiv 1 + \beta_1 - \beta_3$ and $\gamma_2 \equiv \beta_2$. This is a regression of w on linear combinations of (z_1, z_2, z_3), so it will generate the exact same residuals as the regression of w on (z_1, z_2, z_3).[2] Therefore the sum of squared residuals will be 80. □

---

5. [8.18, Hansen] *Suppose you have two independent samples each with n observations which satisfy the models $Y_1 = X_1'\beta_1 + e_1$ with $\mathbb{E}[X_1 e_1] = 0$ and $Y_2 = X_2'\beta_2 + e_2$ with $\mathbb{E}[X_2 e_2] = 0$ where β_1 and β_2 are both $k \times 1$. You estimate β_1 and β_2 by OLS on each sample, with consistent asymptotic covariance matrix estimators $\hat{V}_{\beta_1}$ and $\hat{V}_{\beta_2}$. Consider the efficient minimum distance estimation under the restriction $\beta_1 = \beta_2$.*

(a) *Find the estimator $\tilde{\beta}$ of $\beta = \beta_1 = \beta_2$.*

Solution. Since the samples are independent, the estimators are independent and thus their joint asymptotic covariance matrix (and estimate) is block diagonal: $\begin{bmatrix} \hat{V}_{\beta_1} & 0 \\ 0 & \hat{V}_{\beta_2} \end{bmatrix}$. The minimum-distance criterion takes the form

$$\begin{aligned}
J_n(\beta) &= n \begin{bmatrix} \hat{\beta}_1 - \beta \\ \hat{\beta}_2 - \beta \end{bmatrix}' \begin{bmatrix} \hat{V}_{\beta_1} & 0 \\ 0 & \hat{V}_{\beta_2} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_1 - \beta \\ \hat{\beta}_2 - \beta \end{bmatrix} \\
&= n(\hat{\beta}_1 - \beta)'\hat{V}_{\beta_1}^{-1}(\hat{\beta}_1 - \beta) + n(\hat{\beta}_2 - \beta)'\hat{V}_{\beta_2}^{-1}(\hat{\beta}_2 - \beta).
\end{aligned}$$

[1] Recall that

$$s^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-k} = \frac{SSE(\hat{\beta}_{OLS})}{n-k}.$$

[2] Recall Exercise 5.1 from Problem Set 2.

The FOC for minimization are

$$-2n\hat{V}_{\beta_1}^{-1}(\hat{\beta}_1 - \tilde{\beta}) - 2n\hat{V}_{\beta_2}^{-1}(\hat{\beta}_2 - \tilde{\beta}) = 0$$

with solution

$$\tilde{\beta} = (\hat{V}_{\beta_1}^{-1} + \hat{V}_{\beta_2}^{-1})^{-1}(\hat{V}_{\beta_1}^{-1}\hat{\beta}_1 + \hat{V}_{\beta_2}^{-1}\hat{\beta}_2).$$

This is a weighted average of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$, with weights depending on the covariance matrices. □

(b) *Find the asymptotic distribution of $\tilde{\beta}$.*

Solution. We know that since $\beta_1 = \beta_2 = \beta$,

$$\sqrt{n}(\hat{\beta}_1 - \beta) \xrightarrow{d} Z_1 \sim N(0, V_{\beta_1}) \quad \text{and} \quad \sqrt{n}(\hat{\beta}_2 - \beta) \xrightarrow{d} Z_2 \sim N(0, V_{\beta_2}),$$

where Z_1 and Z_2 are independent. The convergence is also joint convergence. Furthermore, $\hat{V}_{\beta_1} \xrightarrow{p} V_{\beta_1}$ and $\hat{V}_{\beta_2} \xrightarrow{p} V_{\beta_2}$. It follows that

$$\begin{aligned}
\sqrt{n}(\tilde{\beta} - \beta) &= (\hat{V}_{\beta_1}^{-1} + \hat{V}_{\beta_2}^{-1})^{-1}(\hat{V}_{\beta_1}^{-1}\sqrt{n}(\hat{\beta}_1 - \beta) + \hat{V}_{\beta_2}^{-1}\sqrt{n}(\hat{\beta}_2 - \beta)) \\
&\xrightarrow{d} (V_{\beta_1}^{-1} + V_{\beta_2}^{-1})^{-1}(V_{\beta_1}^{-1}Z_1 + V_{\beta_2}^{-1}Z_2) \\
&\sim (V_{\beta_1}^{-1} + V_{\beta_2}^{-1})^{-1}N(0, V_{\beta_1}^{-1} + V_{\beta_2}^{-1}) \\
&= N\left(0, (V_{\beta_1}^{-1} + V_{\beta_2}^{-1})^{-1}\right).
\end{aligned}$$

□

(c) *How would you approach the problem if the sample sizes are different, say n_1 and n_2?*

Solution. The (approximate) variance of $\hat{\beta}_1$ is $n_1^{-1}\hat{V}_{\beta_1}$ and that of $\hat{\beta}_2$ is $n_2^{-1}\hat{V}_{\beta_2}$. Thus a minimum-distance criterion can be written as

$$\begin{aligned}
J_n(\beta) &= \begin{bmatrix} \hat{\beta}_1 - \beta \\ \hat{\beta}_2 - \beta \end{bmatrix}' \begin{bmatrix} n_1^{-1}\hat{V}_{\beta_1} & 0 \\ 0 & n_2^{-1}\hat{V}_{\beta_2} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\beta}_1 - \beta \\ \hat{\beta}_2 - \beta \end{bmatrix} \\
&= n_1(\hat{\beta}_1 - \beta)'\hat{V}_{\beta_1}^{-1}(\hat{\beta}_1 - \beta) + n_2(\hat{\beta}_2 - \beta)'\hat{V}_{\beta_2}^{-1}(\hat{\beta}_2 - \beta).
\end{aligned}$$

Minimizing, we find the solution

$$\tilde{\beta} = (n_1\hat{V}_{\beta_1}^{-1} + n_2\hat{V}_{\beta_2}^{-1})^{-1}(n_1\hat{V}_{\beta_1}^{-1}\hat{\beta}_1 + n_2\hat{V}_{\beta_2}^{-1}\hat{\beta}_2).$$

This is also a weighted average, but now the weights depend on the sample size as well. To develop an asymptotic theory we need to describe what it means for n_1, n_2 to diverge to infinity. A convenient solution is to assume that both diverge, but $n_1/n_2 \to c$, a constant which can differ from one. In practice, we simply think of c as the observed ratio n_1/n_2. Then we can treat $n_1 = cn_2$, and conduct the asymptotics as $n_2 \to \infty$. We have

$$\sqrt{n_1}(\hat{\beta}_1 - \beta) \xrightarrow{d} Z_1 \sim N(0, V_{\beta_1}) \quad \text{and} \quad \sqrt{n_2}(\hat{\beta}_2 - \beta) \xrightarrow{d} Z_2 \sim N(0, V_{\beta_2}).$$

Also,

$$\sqrt{n_2}(\hat{\beta}_1 - \beta) = \sqrt{\frac{n_2}{n_1}}\sqrt{n_1}(\hat{\beta}_1 - \beta) \xrightarrow{d} c^{-1/2}Z_1.$$

Notice that

$$\begin{aligned}\tilde{\beta} &= \left(\frac{n_1}{n_2}\hat{V}_{\beta_1}^{-1} + \hat{V}_{\beta_2}^{-1}\right)^{-1}\left(\frac{n_1}{n_2}\hat{V}_{\beta_1}^{-1}\hat{\beta}_1 + \hat{V}_{\beta_2}^{-1}\hat{\beta}_2\right) \\ &\approx \left(c\hat{V}_{\beta_1}^{-1} + \hat{V}_{\beta_2}^{-1}\right)^{-1}(c\hat{V}_{\beta_1}^{-1}\hat{\beta}_1 + \hat{V}_{\beta_2}^{-1}\hat{\beta}_2),\end{aligned}$$

whence it follows that

$$\begin{aligned}\sqrt{n_2}(\tilde{\beta} - \beta) &= (c\hat{V}_{\beta_1}^{-1} + \hat{V}_{\beta_2}^{-1})^{-1}(c\hat{V}_{\beta_1}^{-1}\sqrt{n_2}(\hat{\beta}_1 - \beta) + \hat{V}_{\beta_2}^{-1}\sqrt{n_2}(\hat{\beta}_2 - \beta)) \\ &\xrightarrow{d} (cV_{\beta_1}^{-1} + V_{\beta_2}^{-1})^{-1}(cV_{\beta_1}^{-1}c^{-1/2}Z_1 + V_{\beta_2}^{-1}Z_2) \\ &\sim (cV_{\beta_1}^{-1} + V_{\beta_2}^{-1})^{-1}N(0, cV_{\beta_1}^{-1} + V_{\beta_2}^{-1}) \\ &= N\left(0, (cV_{\beta_1}^{-1} + V_{\beta_2}^{-1})^{-1}\right).\end{aligned}$$

If you want to write in terms of n_1,

$$\sqrt{n_1}(\tilde{\beta} - \beta) = \sqrt{\frac{n_1}{n_2}}\sqrt{n_2}(\tilde{\beta} - \beta) \xrightarrow{d} c^{1/2}N\left(0, (cV_{\beta_1}^{-1} + V_{\beta_2}^{-1})^{-1}\right) = N\left(0, (V_{\beta_1}^{-1} + c^{-1}V_{\beta_2}^{-1})^{-1}\right).$$

The two are equivalent. □

---

6. [16.29, LNs] *Consider the regression*

$$y = X\beta + u,$$

where $\mathbb{E}[u|X] = 0$ *and* $V[u|X] = \sigma^2 I_N$ *with unknown* σ^2*.*

(a) *Find the Wald statistic for* $H_0 : R\beta - r = 0$*. Derive its distribution under* H_0*.*

Solution. Let R be a full rank $q \times k$ matrix. The Wald statistic for $H_0 : R\beta - r = 0$ is

$$W = (R\hat{\beta} - r)'\hat{V}_{R\hat{\beta}}^{-1}(R\hat{\beta} - r) = \sqrt{n}(R\hat{\beta} - r)'\hat{V}_{R\beta}^{-1}\sqrt{n}(R\hat{\beta} - r),$$

where $\hat{V}_{R\hat{\beta}}$ is some (consistent) covariance matrix estimator for $R\hat{\beta}$ and $\hat{V}_{R\beta} \equiv n\hat{V}_{R\hat{\beta}}$.

We know that $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathbb{E}[x_i x_i']^{-1}\sigma^2)$. Under the null, $R\beta = r$. Therefore, by the multivariate delta method,

$$\sqrt{n}(R\hat{\beta} - r) = \sqrt{n}(R\hat{\beta} - R\beta) \xrightarrow[H_0]{d} RN(0, \mathbb{E}[x_i x_i']^{-1}\sigma^2) = N(0, \underbrace{R\mathbb{E}[x_i x_i']^{-1}R'\sigma^2}_{\equiv V_{R\hat{\beta}}}) \equiv Z.$$

Since $\hat{V}_{R\hat{\beta}} \xrightarrow{p} V_{R\beta}$, it follows by the continuous mapping theorem and Slutsky's theorem that

$$W \xrightarrow[H_0]{d} Z'[R\mathbb{E}[x_i x_i']^{-1}R'\sigma^2]^{-1}Z \sim \chi^2_q.$$

Thus the Wald statistic for $H_0 : R\beta - r = 0$ is asymptotically chi-squared distributed with q degrees of freedom. □

(b) *Show that the Wald statistic for $R\beta - r = 0$ equals the largest Wald statistic among all one-dimensional tests for restrictions of the form $c'(R\beta - r) = 0$. Comment.*

Solution. The Wald statistic for restrictions of the form $c'(R\beta - r) = 0$ is

$$W_c = [c'(R\hat{\beta} - r)]'\hat{V}^{-1}_{c'R\hat{\beta}}c'(R\hat{\beta} - r).$$

Notice that

$$\hat{V}_{c'R\hat{\beta}} = c'\hat{V}_{R\hat{\beta}}c,$$

Write

$$\begin{aligned} W_c &= [c'(R\hat{\beta} - r)]'[c'\hat{V}_{R\hat{\beta}}c]^{-1}c'(R\hat{\beta} - r) \\ &= \frac{[c'(R\hat{\beta} - r)]'c'(R\hat{\beta} - r)}{c'\hat{V}_{R\hat{\beta}}c} \\ &= \frac{[c'(R\hat{\beta} - r)]^2}{c'\hat{V}_{R\hat{\beta}}c} \\ &= \frac{[v'\hat{V}^{-1/2}_{R\hat{\beta}}(R\hat{\beta} - r)]^2}{v'v} \\ \text{(Cauchy-Schwartz)} &\leq \frac{v'v[\hat{V}^{-1/2}_{R\hat{\beta}}(R\hat{\beta} - r)]'\hat{V}^{-1/2}_{R\hat{\beta}}(R\hat{\beta} - r)}{v'v} \\ &= (R\hat{\beta} - r)'\hat{V}^{-1}_{R\hat{\beta}}(R\hat{\beta} - r) = W, \end{aligned}$$

where $v = \hat{V}^{1/2}_{R\hat{\beta}}c$ and $\hat{V}^{1/2}_{R\hat{\beta}}$ is such that $\hat{V}^{1/2}_{R\hat{\beta}}\hat{V}^{1/2}_{R\hat{\beta}} = \hat{V}_{R\hat{\beta}}$.

Observe that, in particular, one could set $c = e_i$ to be a canonical selector vector that selects the i-th restriction in $Rb - r$. Suppose instead of the original procedure of jointly testing all the restrictions, $H_0 : Rb - r = 0$, one proposes a new different procedure: testing each restriction separately, by performing a sequence of p independent tests $H_0 : e_i'(Rb - r) = 0$, $i = 1, \ldots, p$, with associated Wald statistic W_i, and then claiming that the joint restrictions are rejected if at least one of the separate tests rejects the null. The above result tells us that $W_i \leq W$ for all $i = 1, \ldots, q$. We reject the null when, for a given $1 - \alpha$ quantile k, the Wald statistic is greater than k. Thus for some i and some α we could have quantiles k_1 and k_q such that $k_1 < W_i < W < k_q$, where k_1 and k_q are the $1 - \alpha$ quantiles of the χ^2_1 and χ^2_q distributions, respectively. In this case, the new procedure would reject the null, while the original one would not. This shows that the new proposed procedure is problematic. Indeed, it could even be used for cheating: for example, one could purposely seek a significance level such that $k_1 < W_i < W < k_q$ for *some* i so that we reject the null based on the new procedure when actually — based on the original correct one — we should not! □

7. [7.18, Hansen] *Suppose an economic model suggests*

$$m(x) = \mathbb{E}[Y|X = x] = \beta_0 + \beta_1 x + \beta_2 x^2,$$

where $X \in \mathbb{R}$. *You have a random sample* (Y_i, X_i), $i = 1, \ldots, n$.

(a) *Describe how to estimate* $m(x)$ *at a given value* x.

Solution. The model is a linear regression. We know this because the question specifies the conditional mean of y_i given x_i. This is a regression, so it does not need to be assumed. Furthermore, as the conditional variance is unspecified we assume heteroskedasticity.

Write

$$\begin{aligned} g(x) &= \mathbb{E}[y_i|x_i = x] \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 \\ &= z'\beta, \end{aligned}$$

where

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{and} \quad z = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}.$$

As the equation is linear in the parameters, it can be estimated by least-squares:

$$y_i = \hat\beta_0 + \hat\beta_1 x_i + \hat\beta_2 x_i^2 + \hat e_i = \hat\beta' z_i + \hat e_i,$$

where

$$\hat\beta = \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \\ \hat\beta_2 \end{bmatrix} \quad \text{and} \quad z = \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix}.$$

The estimate for the conditional mean function at x is simply

$$\hat g(x) = \hat\beta_0 + \hat\beta_1 x + \hat\beta x^2 = z'\hat\beta.$$

Notice that since the model is a regression, FGLS would also be possible. However, since the model did not suggest a functional form for the variance, FGLS would require an approximate variance equation, and it may just be easiest and best to use least-squares. □

(b) *Describe (be specifc) an appropriate confidence interval for* $m(x)$.

Solution. We know that

$$\sqrt{n}(\hat\beta - \beta) \xrightarrow{d} N(0, V),$$

where $V = Q^{-1}\Omega Q^{-1}$. Thus

$$\sqrt{n}(\hat g(x) - g(x)) = \sqrt{n}(z'\hat\beta - z'\beta) = z'\sqrt{n}(\hat\beta - \beta) \xrightarrow{d} N(0, z'Vz).$$

An estimator for V is

$$\hat{V} = \hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1},$$

where

$$\hat{Q} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' \quad \text{and} \quad \hat{\Omega} = \sum_{i=1}^{n} z_i z_i' \hat{e}_i^2.$$

A standard error for $\hat{g}(x) = z'\hat{\beta}$ is $(n^{-1}z'\hat{V}z)^{1/2}$. An asymptotic 95% confidence interval is

$$\begin{aligned}\hat{g}(x) \pm 2(n^{-1}z'\hat{V}z)^{1/2} &= z'\hat{\beta} \pm 2(n^{-1}z'\hat{V}z)^{1/2} \\ &= \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 \pm 2(n^{-1}z'\hat{V}z)^{1/2}.\end{aligned}$$

Since the function $z'\hat{\beta}$ is linear in the parameters, this Wald confidence interval approach is appropriate. □

---

10. [16.31, LNs] *Consider the model*

$$y_i = x_i'\beta + u_i,$$

where (x_i', u_i) are iid with $u|x_i \sim N(0, \sigma^2)$. We want to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ using the statistic

$$W = \frac{(\hat{\beta} - \beta_0)'X'X(\hat{\beta} - \beta_0)}{\sigma^2}.$$

(a) *Is this statistic pivotal? Explain your answer.*

Solution. Under the null, yes. Observe that $\hat{\beta} - \beta_0 = (X'X)^{-1}X'u$, whence

$$W = (X'u)'(X'X\sigma^2)^{-1}X'u.$$

Notice that $X'u \sim N(0, X'X\sigma^2)$. Therefore $W \sim \chi_k^2$, which does not depend on (β, σ). □

(d) *Show that*

$$\hat{W} = \frac{(\hat{\beta} - \beta_0)'X'X(\hat{\beta} - \beta_0)/k}{e'e/(n-k)}$$

has a $F(k, n-k)$ distribution under the null.

Solution. We have already verified that, under the null, when divided by σ^2, the numerator term has a χ_k^2 distribution. For the denominator, notice that $e'e/\sigma^2 = u'Mu/\sigma^2$. Spectral decomposition on M gives $M = H\Lambda H'$, where $H'H = I_n$. Thus

$$\begin{aligned}\frac{e'e}{\sigma^2} &= \left(\frac{H'u}{\sigma}\right)' \Lambda \left(\frac{H'u}{\sigma}\right) \\ &= \left(\frac{H'u}{\sigma}\right)' \begin{bmatrix} I_{n-k} & 0 \\ 0 & 0 \end{bmatrix} \left(\frac{H'u}{\sigma}\right).\end{aligned}$$

Let $w \equiv H'u/\sigma$. Notice that $w \sim N(0, I_n)$. Partition $w = (w_1', w_2')'$, where $w_1 \sim N(0, I_{n-k})$. It follows that

$$\frac{e'e}{\sigma^2} = w_1'w_1 \sim \chi^2_{n-k}.$$

Write

$$\hat{W} = \frac{\overbrace{\frac{(\hat{\beta}-\beta_0)'X'X(\hat{\beta}-\beta_0)}{\sigma^2}}^{\sim\chi^2_k}/k}{\underbrace{\frac{e'e}{\sigma^2}}_{\sim\chi^2_{n-k}}/(n-k)} \sim F(k, n-k).$$

Notice that numerator and denominator terms are independent. Indeed, the numerator term can be written as $u'Nu/\sigma^2$, while the denominator term can be written as $u'Mu/\sigma^2$. As N and M are orthogonal, both terms are orthogonal and hence independent (due to normality). Therefore $\hat{W} \sim F(k, n-k)$. □

(e) *Show that $\hat{W}$ asymptotically has a chi-square distribution with k degrees of freedom.*

Solution. Consider the above expression for $\hat{W}$. Notice that $e'e/(n-k) \xrightarrow{p} \sigma^2$, so the denominator asymptotically converges to 1. Furthermore, observe that the numerator asymptotically converges in distribution to $Z'[\mathbb{E}[x_i x_i']^{-1}\sigma^2]^{-1}Z/k \sim \chi^2_k/k$, where $Z \sim N(0, \mathbb{E}[x_i x_i']^{-1}\sigma^2)$.[3] Therefore, by Slutsky, it follows that $k\hat{W} \xrightarrow{d} \chi^2_k$. □

---

[3]This is a particular case of Exercise 6, item **(a)**. Here, $R = I_k$ and $r = \beta_0$.