Midterm Solutions

Part A - Multiple Choice

1) B 2) D 3) D 4) B 5) C

Part B - True/False

6. TRUE

Since the model includes an intercept, the orthogonality condition $X'e = 0$ ensures that the regression line passes through the sample means of the data, $\bar{y}$ and $\bar{X}$. Hence, when $X = \bar{X}$, $\hat{y} = \bar{X}b = \bar{y}$. You do not need to run an OLS regression, you simply need to calculate $\bar{y}$.

7. FALSE

Including an irrelevant regressor: lose efficiency under OLS.

Excluding a relevent regressor: lose unbiasedness and consistency.

Since consistency is usually considered the most important property, exclusion is generally considered to be more costly.

8. TRUE/FALSE

We can write R^2 as:

$$R^2 = 1 - \frac{e'e}{y'M_0y}$$, which is a decreasing function in $e'e$.

Since minimizing $e'e$ is the objective of OLS, including an extra variable is akin to relaxing a constraint in the optimization problem. The objective function $e'e$ must be smaller, or at least the same, when a constraint is relaxed.

9. TRUE

$$e = y - \hat{y} = X\beta + \varepsilon - Xb$$

Under our standard assumptions, $\text{plim}(b) = \beta$.

By Slutsky's theorem,

$$\text{plim}(e) = \text{plim}(X\beta) + \text{plim}(\varepsilon) - \text{plim}(Xb)$$

$$= X\beta + \varepsilon - X\beta = \varepsilon$$

10. FALSE

This model can't be estimated by OLS, but not because A.5 has been violated. In this model, there is perfect multicollinearity between M and F. Specifically, $M = 1 - F$. The assumption of full rank has been violated – $(X'X)$ can't be inverted and b can't be calculated.

11. FALSE.

$\text{plim}\left(\frac{X'X}{n}\right) = Q$ is a sufficient, but not necessary condition for the weak consistency of b. If instead we can show that b is Mean-Square consistent for β, then we also have weak consistency, since the former implies the latter. In order to show M.S. consistency we need to show that as $n \to \infty$, $\text{bias}(b) \to 0$ and $\text{var}(b) \to 0$.

Part C

12) If there is any correlation between X_1 and X_2, then

a) omitting X_2 will cause b_1 to be biased. This is because part of the effect of X_2 on y is attributed to X_1.

Now, the true value of β_3 is zero. If X_1 and X_2 are correlated, and X_1 and X_3 are correlated, then in general $E(b_3) \neq 0$. This is because part of the omitted effect of X_2 is passing through X_1 to X_3. b_3 will also be biased if X_2 and X_3 are correlated directly.

Under the fitted model:

$$b_1 = (X_1' M_3 X_1)^{-1} X_1' M_3 y \quad , \text{ where } M_3 = I - X_3 (X_3' X_3)^{-1} X_3'$$

$$b_1 = \underbrace{(X_1' M_3 X_1)^{-1} X_1' M_3 X_1}_{I} \beta_1 + (X_1' M_3 X_1)^{-1} X_1' M_3 X_2 \beta_2$$

$$+ (X_1' M_3 X_1)^{-1} X_1' M_3 \varepsilon$$

$$E(b_1) = \beta_1 + (X_1' M_3 X_1)^{-1} X_1' M_3 X_2 \beta_2 + 0.$$

So, b_1 is biased since $E(b_1) \neq \beta_1$, in general.

Similarly,

$$b_3 = (X_3' M_1 X_3)^{-1} X_3' M_1 X_1 \beta_1 + (X_3' M_1 X_3)^{-1} X_3' M_1 X_2 \beta_2$$

$$+ (X_3' M_1 X_3)^{-1} X_3' M_1 \varepsilon.$$

Since $M_1 X_1 = 0$, the first term is zero, and:

$E(b_3) = (X_3' M_1 X_3)^{-1} X_3' M_1 X_2 \beta_2$.

Since $E(b_3) \neq 0$, b_3 is biased, in general.

b) In order for b_1 to be unbiased, X_1 and X_2 must be uncorrelated ($X_1'X_2 = 0$). In addition, either X_1 and X_3 must be uncorrelated ($X_1'X_3 = 0$) OR ($X_2'X_3 = 0$).

The bias of b_1 is coming from the term:

$(X_1' M_3 X_1)^{-1} X_1' M_3 X_2 \beta_2$. Substituting in for M_3:

$$(X_1'X_1 - X_1'X_3(X_3'X_3)^{-1}X_3'X_1)^{-1}[X_1'X_2 - X_1'X_3(X_3'X_3)^{-1}X_3'X_2]\beta_2$$

This bias term only disappears when $X_1'X_2 = 0$ and either $X_1'X_3 = 0$ or $X_2'X_3 = 0$.

As for b_3, $E(b_3) = 0$ when $X_3'X_2 = 0$ and either $X_3'X_1 = 0$ or $X_1'X_2 = 0$.

c) Efficiency can go either way. We have including an irrelevant regressor which causes var(b) to increase, but have excluded a variable which causes var(b) to decrease.

We can't determine under which model the OLS estimators would have smaller variance, until we calculate the $(X'X)^{-1}$ matrix.

13. i) Fit the model, $X = Z\theta + \varepsilon$, get $\hat{\theta}$ by OLS.

Then, $\hat{X} = Z\hat{\theta} = Z(Z'Z)^{-1}Z'X$

ii) Estimating β by OLS in the following:

$$y = \hat{X}\beta + u$$

gives:

$$b = (\hat{X}'\hat{X})^{-1}\hat{X}'y = \left(X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'y$$

$$= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$$

Which is b_{IV} in the over-identified case.

14. a) If $E(\tilde{\theta}) = \dfrac{m-j}{m}\theta$,

then $E\left(\dfrac{m}{m-j}\tilde{\theta}\right) = \theta$.

If we construct a new estimator:

$$\theta^* = \frac{m}{m-j}\tilde{\theta},$$

then θ^* will be unbiased.

b) An example of where this strategy has been employed is in the estimation of σ^2.

If we use a "natural" estimator $\hat{\sigma}^2 = \frac{e'e}{n}$, we find that $E(\hat{\sigma}^2) = \frac{n-k}{n}\sigma^2$. If we instead use $\frac{n}{n-k}\hat{\sigma}^2$, we get an unbiased estimator, i.e. s^2.

15. a) $\bar{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k} = 1 - (.886142)\frac{427}{425} = 0.109688$

$t\text{-stat} = \frac{0.317758}{0.178941} = 1.776$

b) $H_0: \beta_1 = 0 \qquad H_A: \beta_1 \neq 0$

The tstat has been calculated in (a). The relevant p-value is reported in EViews output as 0.0765

At the 1% and 5% significance levels I fail to reject, but at the 10% level I reject.

c) $[1.28 - 1.96 \times 0.18 \;,\; 1.28 + 1.96 \times 0.18] = 0.927, 1.633$

If I were to repeatedly resample, the true value (β_2) would lie in this interval 95% of the time.