Midterm 2 - 2014 - Answer Key

1.) FALSE. The null is $\text{plim}\left(\frac{X'\varepsilon}{n}\right) = 0$. A small p-value would indicate we should reject the null, and use I.V. instead of OLS.

2.) TRUE/FALSE! Which test is better depends on the situation. The F-test should be used if it can be used, since it provides exact results for a finite sample. If any of the assumptions underlying the appropriate distribution of the F-test are violated, however, the Wald test should be used instead. The Wald test is likely applicable to a wider variety of situations.

3.) TRUE.
$$\begin{aligned} Z'e_{IV} &= Z'(y - X(Z'X)^{-1}Z'y) \\ &= Z'y - Z'y = 0. \end{aligned}$$

If Z contains a column of "ones", the I.V. residuals will sum to zero.

4.) FALSE. While the limiting distribution of both estimators will have zero variance, we can control for the rate at which the sampling distribution of the estimators collapse. For example, if $\text{var}(\hat{\theta}) \to 0$ ~~as $n \to \infty$~~ at a rate of n^{-1}, we can instead consider the variance of $\sqrt{n}\,\hat{\theta}$. If the sampling distribution of another consistent estimator, $\tilde{\theta}$, is also collapsing at rate n^{-1}, then the asymptotic efficiency of $\hat{\theta}$ and $\tilde{\theta}$ may be compared.

5.) FALSE. OLS will be inconsistent if a relevant variable is excluded, AND the excluded variable is correlated with one or more of the included variables.

I.V. may be used, as long as the instruments are uncorrelated with the unobservable variable, and correlated with the included endogenous variables.

6.a) $b = (X'X)^{-1}X'(X\beta + \varepsilon)$

$$= \beta + (X'X)^{-1}X'\varepsilon$$

$$= \beta + \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\varepsilon}{n}\right)$$

Assuming that:

$\operatorname{plim}\left(\frac{X'X}{n}\right) = Q$ (p.s.d. and finite),

$\operatorname{plim}\left(\frac{X'\varepsilon}{n}\right) = 0$ (the X data and ε are asymptotically uncorrelated),

then (using Slutsky's theorem):

$\operatorname{plim}(b) = \beta + Q^{-1}\cdot 0 = \beta$,

and OLS is consistent.

b) Irrelevant regressor – b is inefficient, but is unbiased and consistent.

Excluding a relevant regressor – b is more precise (smaller variance), but is biased and inconsistent if the excluded variable is correlated with the included ones.

7. a) We can define a dummy variable, D:

$$D = 0 \text{ if } YEAR \leq 1973$$
$$D = 1 \text{ if } YEAR > 1973$$

One model we could specify is:

$$\ln GAS = \beta_1 + \beta_2 \ln GASP + \beta_3 \ln PNC + \beta_4 \ln PUC$$
$$+ \beta_5 \ln Income/Pop + \beta_6 D \ln GASP$$

Notice that the effects of all variables (except GASP) are the same across 1973. We could also add $\beta_7 D$ to the model, which would allow for the intercept to change across the break, while maintaining the belief that all other effects remain constant.

An appropriate null hypothesis is then:

$$H_0: \beta_6 = 0 \quad \text{vs.} \quad H_A: \beta_6 \neq 0,$$

where the F-test or Wald test may be used.

b) For the Chow test, we would want the dummy variable D to interact fully with all regressors in the model, and then jointly test the significance of all estimated coefficients which are associated with D.

8.) a)

$$e_*'e_* = \left[(Rb-q)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' + e'\right] \cdot$$

$$\left[e + X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb-q)\right]$$

$$= (Rb-q)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}\underbrace{X'e}_{=0 \text{ if intercept}}$$

$$+ \underbrace{e'X}_{=0}(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb-q)$$

$$+ e'e$$

$$+ (Rb-q)'\overbrace{[R(X'X)^{-1}R']^{-1}R\overbrace{(X'X)^{-1}X'X}^{I}(X'X)^{-1}R'}^{I} \cdot$$

$$[R(X'X)^{-1}R']^{-1}(Rb-q)$$

$$= 0 + 0 + e'e + (Rb-q)'[R(X'X)^{-1}R']^{-1}(Rb-q)$$

Hence, $e_*'e_* - e'e = (Rb-q)'[R(X'X)^{-1}R']^{-1}(Rb-q)$,

and $F = \dfrac{(e_*'e_* - e'e)/J}{s^2}$.

The F-test ignores the restrictions being imposed, so s^2 is calculated from the unrestricted regression:

$$s^2 = \frac{e'e}{n-K}$$

b) If the model includes an intercept, then:

$$R^2_U = 1 - \frac{e'e}{SST}$$

$$R^2_R = 1 - \frac{e_*'e_*}{SST}$$

SST is the same for both measures.

So, $e_*'e_* = SST(1 - R^2_R)$,

and $e'e = SST(1 - R^2_U)$.

$$e_*'e_* - e'e = SST(R^2_U - R^2_R).$$

So the F-stat: $F = \dfrac{(e_*'e_* - e'e)/J}{e'e/(n-k)}$ may be rewritten as:

$$F = \frac{SST\,(R^2_U - R^2_R)\,/\,J}{SST(1-R^2_U)\,/\,(n-k)} = \frac{(R^2_U - R^2_R)\,/\,J}{(1-R^2_U)\,/\,(n-k)}$$

9a) The I.V. estimator is:

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

which is generally a biased estimator, so $\hat{\beta}$ is likely to be biased as well. $\hat{\beta}$ is consistent however:

$$\hat{\beta} = [A + Z'X]^{-1}Z'X\beta + [A + Z'X]^{-1}\cancel{Z}'\varepsilon$$

$$= \left[\frac{A}{n} + \frac{Z'X}{n}\right]^{-1}\frac{Z'X}{n}\beta + \left[\frac{A}{n} + \frac{Z'X}{n}\right]^{-1}\frac{Z'\varepsilon}{n}$$

Since A is non-random, $\text{plim}\left(\frac{A}{n}\right) = \lim_{n\to\infty}\left(\frac{A}{n}\right) = 0.$

So, $\text{plim}(\hat{\beta}) = [0 + Q_{ZX}]^{-1}Q_{ZX}\beta + [0 + Q_{ZX}]^{-1}\cdot 0$

$$= \beta$$

When comparing the asymptotic variance of this estimator with the I.V. estimator, efficiency could go either way.

b) In order for $\hat{\beta}$ to still be consistent, we need $\text{plim}\left(\frac{A}{n}\right) = 0.$