Department of Economics **University of Manitoba**

ECON 7010: Econometrics I

FINAL EXAM, Dec. 12th, 2020

Instructor: **Ryan Godwin**
Instructions: **Upload your answers to UM Learn.**
Time Allowed: **2 hours.**
Number of Pages: **3**

There are a total of 100 marks.

Answer all questions. Each question is worth 20 marks.

1.) The assumptions underlying the classical linear regression model are:

A.1	Linearity
A.2	Full rank: rank(X) = k
A.3	Errors have zero mean: $E(\varepsilon) = 0$
A.4	Spherical errors
A.5	The process that generates X is unrelated to the process that generates ε
A.6	Normality of errors

Under these assumptions, OLS is a good estimator. Additionally, $s^2(X'X)^{-1}$ is a good estimator for the variance-covariance matrix of the OLS estimator. For each of the assumptions listed above (if applicable): (i) describe the consequences of the assumption being violated (in terms of estimation, hypothesis testing, etc.), and (ii) describe how the problem might be corrected, or how an alternative estimator might be used to correct the problem (if possible).

See Final 2013 answer key.

2.) Suppose that we have a linear multiple regression model,

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

where the regressors may be random and may be correlated with the errors, even asymptotically. Accordingly, we decide to use a (generalized) IV estimator, where Z is the $(n \times L)$ matrix of (possibly random) instruments.

a) Prove, algebraically, that the IV estimator is equivalent to the following two-step estimator:

(i) Regress X on Z by OLS, and get the predicted matrix, $\hat{X}$.
(ii) Fit the following artificial model by OLS: $y = \hat{X}\beta + v$.

See pg. 75 of the lecture notes.

b) Prove that the generalized IV estimator collapses to the simple one, if X and Z have the same dimensions.

The generalized IV estimator is:

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

If X and Z have the same dimension then we may expand out the inverse and write:

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

The middle terms become identity and disappear:

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

Which is the simple IV estimator.

c) Briefly describe how you would test to see if IV estimation is needed.

See section 8.4.1 of the lecture notes.

d) Is there a way to test if the assumption that $plim(Z'\varepsilon/n) = 0$ is satisfied? Explain.

There is no way (that I know of) to test this assumption, without additional information. See section 8.4.2.

3.) Consider following wage equation:

$$log(wage) = \beta_0 + \beta_1 Educ + \beta_2 Exper + \beta_3 Marr + \epsilon$$

where $wage$ is hourly wage, $Educ$ is years of education, $Exper$ is job experience in years, and $Marr$ is a dummy variable that is equal to 1 if individual is married and zero if single.

a) Using the estimation results in the table below calculate the **F-statistics** in order to test the following hypothesis:

$$H_o: \beta_3 = 0.5$$

Explain your reasoning.

```
##
## =======================================================
##                              Dependent variable:
##                     -----------------------------------
##                                    lwage
## -------------------------------------------------------
## educ                             0.09***
##                                   (0.01)
##
## exper                            0.01***
##                                  (0.002)
##
## married                          0.19***
##                                   (0.04)
##
## Constant                          0.22**
##                                   (0.11)
##
## -------------------------------------------------------
## Observations                        526
## R2                                  0.28
## Adjusted R2                         0.27
## Residual Std. Error           0.45 (df = 522)
## F Statistic                66.27*** (df = 3; 522)
```

```
## =====================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
##                         Standard Errors are in parenthesis
```

When J = 1, the F-statistic is equal to the square of the t-statistic. The t-stat is:

$$t = \frac{0.19 - 0.5}{0.04}$$

b) Assuming that you cannot reject the following hypothesis

$$H_o: \beta_1 - 9\beta_2 = 0, and \beta_3 = 1$$

Derive the restricted model. Why is the LS estimator of the restricted model more appropriate than the unrestricted model? Explain

The restricted model could be estimated as:

$$log(wage) - Marr = \beta_0 + \beta_2(Exper - 9Educ) + \epsilon$$

The restricted model is more appropriate in the sense that it is more *efficient* (i.e. the estimators have smaller variance).

c) As we learned in class the hypothesis given in section (b) can be re-written as

$$H_o: R\beta = q$$

We also saw that $m \sim N(0, V(m))$ where $m = Rb - q$

Calculate the **Wald Statistic** that can be used to test the hypothesis in section (b) assuming that a consistent estimate of $V(m)^{-1}$ is given below.

5443	-481
-481	572

$$Rb - q = \begin{bmatrix} 0 \\ -0.81 \end{bmatrix}$$

So,

$$W = \begin{bmatrix} 0 & -0.81 \end{bmatrix} \begin{bmatrix} 5443 & -481 \\ -481 & 572 \end{bmatrix} \begin{bmatrix} 0 \\ -0.81 \end{bmatrix} = 375$$

4.) Suppose that $V(\boldsymbol{\varepsilon}) \neq \sigma^2 I$. Instead, heteroskedasticity is present and $V(\boldsymbol{\varepsilon}) = \sigma^2\Omega$.

a) Explain the implications of ignoring heteroskedasticity.

See the highlighted box on pg. 102.

b) Derive $V(\boldsymbol{b})$ for when $V(\boldsymbol{\varepsilon}) = \sigma^2\Omega$.

See Section 11.1.

c) Describe White's test for heteroskedasticity.

See the numbered list on pg. 107 of Section 11.3.1.

d) Suppose that you have an additional variable, z_i, which describes the form of heteroskedasticity, such that:

$$\Omega = \begin{bmatrix} z_1 & 0 & 0 & \dots & 0 \\ 0 & z_2 & 0 & \dots & 0 \\ 0 & 0 & z_3 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & & z_n \end{bmatrix}$$

Explain how to transform the model into one that is characteristic of homoskedasticity, and how you could obtain the Generalized Least Squares (GLS) estimates from the transformed model.

If we replace z_i with n_j, then we have the same situation as in 11.4.3. Each observation should be multiplied by $\frac{1}{\sqrt{z_i}}$.

e) Prove that the transformed model from (d) above satisfies A.4.

Show that $V(P\varepsilon) = \sigma^2 I$ as in the highlighted box in Section 11.4.2.

5.)

a) Suppose that we wish to estimate a regression model by least squares, but that this model is non-linear in the parameters. Will we need to use a numerical algorithm to obtain the parameter estimates? If so, explain why.

There is a possibility that we can transform the model to one that is linear in the parameters, and still use LS. It might also be possible to approximate the non-linear relationship using polynomials, logarithms, or splines, and still use LS. (See Sections 10.1 – 10.3). In other cases, we can use NLS. In NLS, usually the FOCs for the estimators are not solvable and then will require a numerical algorithm (see the bottom half of Section 10.4).

b) Illustrate how the Newton-Raphson algorithm can find the value of θ that minimizes the function:

$$f(\theta) = \frac{\theta^4}{2} - \frac{\theta^3}{3}$$

See Example 10.2 for something very similar.

END.