Econ 7010 - Final - Fall 2022

Ryan T. Godwin

The exam is 180 minutes long, and has 100 marks. You may not use any outside materials, only writing implements. Write your answers in the booklet provided.

Short Answer - Answer 10 out of 12 questions. Only the first 10 questions will be marked. 30% total, each question worth 3%.

1. Prove that the least squares estimator is unbiased and consistent, stating any assumptions that you use.

   See section 5.1 and 7.2 for the proofs of unbiasedness and consistency.

2. Derive the variance-covariance matrix of the LS estimator, under standard assumptions.

   See the math leading up to equation 5.3 in section 5.3.

Use the following R code and output for **Questions 3 to 5**. R code was used to estimate a wage model:

```
cps.mod <- lm(log(wage) ~ education + gender + age + experience
                        + gender * education, data = cps)
summary(cps.mod)
```

The results are:

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            0.53764    0.70887   0.758 0.448521
education              0.18311    0.11333   1.616 0.106753
gendermale             0.69499    0.20315   3.421 0.000672 ***
age                   -0.06472    0.11345  -0.570 0.568616
experience             0.07754    0.11355   0.683 0.494959
education:gendermale  -0.03362    0.01531  -2.196 0.028545 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4509 on 528 degrees of freedom
Multiple R-squared:  0.2769,    Adjusted R-squared:  0.2701
F-statistic: 40.44 on 5 and 528 DF,  p-value: < 2.2e-16
```

3. Test the hypothesis that the effect of education on wage is zero.

This would require an F-test to test the joint significance of both the `education` and `education:gendermale` variables.

---

4. If we added a variable to the regression, would R^2 increase? Briefly explain.

See section 4.5.2.

---

5. What is the purpose of the interaction term?

The interaction term allows for a difference in the effect of education on wage, between men and women.

---

6. Explain the interpretation of a confidence interval.

See section 6.2.4.

---

7. Describe the consequences, and solutions for, heteroskedasticity.

Heteroskedasticity leads the LS estimator to be inefficient (but still biased and consistent), and leads to inconsistent estimators for $V(b)$. This means that standard errors, test statistics, p-values, and confidence intervals will all be wrong. Hypothesis testing will be invalid.

The solutions are to: (i) fix both efficiency and the inconsistency of the standard errors by using GLS (as in the case of clustering) or FGLS; or (ii) ignore the inefficiency of LS and use a "robust" estimator for the standard errors (such as White's).

---

8. Describe a situation where IV estimation might be needed.

See Table 1 on page 82.

---

9. Show how to write the null hypothesis $H_0 : \beta_3 = \beta_4, \beta_1 = 2\beta_2$ in terms of R and $\boldsymbol{q}$, where $k = 4$.

$$R = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 1 & -2 & 0 & 0 \end{bmatrix} \quad ; \quad \boldsymbol{q} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

---

10. Explain how to calculate an F-test statistic, by estimating two different models.

We can view a null hypothesis as something which imposes restrictions on a model. Estimating the restricted model under the null hypothesis and the unrestricted under the alternative hypothesis, and comparing their R^2, allows us to assess the validity of the null. Formally, we can calculate the F-statistic in equation 9.9, which follows an F distribution with J (the number of restrictions) and $n - k$ degrees of freedom.

---

11. Describe two ways to allow for non-linear effects between X and $\boldsymbol{y}$, while still using LS.

Sections 10.1, 10.2, 10.3, each describe a way to approximate non-linear relationships between variables.

---

12. What properties must an instrument have in order to work in IV estimation?

An instrumental variable, Z, must be:

(a) Correlated with the endogenous variables X.
- This is sometimes called the "relevance" of an IV.
- This condition can be tested.

(b) Uncorrelated with the error term, or equivalently, uncorrelated with the dependent variable other than through its correlation with X.
- This is sometimes called the "exclusion" restriction.
- This restriction cannot be easily tested.

Long Answer - Answer 5 out of 6 questions. Only the first 5 questions will be marked. 70% total, each question worth 14%, each part worth 3.5%.

1. The RLS estimator is:

$$\boldsymbol{b}_* = \boldsymbol{b} - \left(X'X\right)^{-1} R' \left[R\left(X'X\right)^{-1} R'\right]^{-1} (R\boldsymbol{b} - \boldsymbol{q})$$

a) When will the RLS and LS estimators be identical? What does this imply for the null hypothesis?

RLS and LS will coincide when $R\boldsymbol{b} = \boldsymbol{q}$. This would mean that the restrictions (e.g. from a null hypothesis) would have to exactly line up with what is estimated from LS.

---

b) Prove that the RLS estimator is unbiased, carefully stating any important assumptions that are required.

See Theorem 9.2 in Section 9.4.

---

c) Explain why the RLS estimator is more efficient than the LS estimator. What assumptions are required for this result?

The RLS estimator is more efficient because there are fewer β to estimate (a restricted model always has fewer things that need to be estimated). When there are fewer β in the model, the estimators have lower variance, because all n sample information is now focusing on fewer estimates. It is as if the sample size has increased.

Importantly, the relative efficiency of the RLS estimator does not depend on A.3 or A.5. That is, even if we impose false restrictions and make the estimators biased and inconsistent, the variance still decreases.

---

d) In practice, how is the RLS estimator calculated?

In practice, the RLS formula is not used. The restrictions are simply substituted into the population model, and the restricted model is estimated by LS.

---

2. a) Show that the 2SLS procedure leads to the IV estimator.

The two stages of 2SLS are:

i. Regress X on the instruments Z. Get the predicted values from this regression. That is, we need $\hat{X} = P_Z X$.
ii. Estimate the population model using $\hat{X}$ instead of X, using LS. That is, estimate the model $y = \hat{X}\boldsymbol{\beta} + \epsilon = P_Z X \boldsymbol{\beta} + \epsilon$.

The LS estimator from stage 2 is:

$$b = \left((P_Z X)'(P_Z X)\right)^{-1}(P_Z X)'\boldsymbol{y} \tag{1}$$

$$= \left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1}X'Z\left(Z'Z\right)^{-1}Z'\boldsymbol{y} \tag{2}$$

which is the "over-identified" formula on the formula sheet.

---

b) Prove that the IV estimator is consistent, stating any assumptions that you use.

Assuming that Z is a valid instrument, and that the Z matrix has full rank:

$$\begin{array}{l} \text{plim}\left(\frac{1}{n}Z'Z\right) = Q_{ZZ} \quad ; \quad \text{p.d. and finite} \\ \text{plim}\left(\frac{1}{n}Z'X\right) = Q_{ZX} \quad ; \quad \text{p.d. and finite} \\ \text{plim}\left(\frac{1}{n}Z'\boldsymbol{\epsilon}\right) = \mathbf{0} \end{array}$$

Then, the IV estimator is *consistent*:

$$\begin{aligned} \boldsymbol{b}_{IV} &= \left(Z'X\right)^{-1}Z'\boldsymbol{y} = \left(Z'X\right)^{-1}Z'(X\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= \left(Z'X\right)^{-1}Z'X\boldsymbol{\beta} + \left(Z'X\right)^{-1}Z'\boldsymbol{\epsilon} \\ &= \boldsymbol{\beta} + \left(Z'X\right)^{-1}Z'\boldsymbol{\epsilon} \\ &= \boldsymbol{\beta} + \left(\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{n}Z'\boldsymbol{\epsilon}\right) \end{aligned}$$

and so:

$$\begin{aligned} \text{plim}\left(\boldsymbol{b}_{IV}\right) &= \boldsymbol{\beta} + \left[\text{plim}\left(\frac{1}{n}Z'X\right)\right]^{-1}\text{plim}\left(\frac{1}{n}Z'\boldsymbol{\epsilon}\right) \\ &= \boldsymbol{\beta} + Q_{ZX}^{-1}\mathbf{0} = \boldsymbol{\beta} \end{aligned}$$

---

c) Explain the intuition behind the Hausman test, for testing if IV is needed.

The Hausman test statistic is constructed by comparing the difference between the IV estimator and the LS estimator: $(\boldsymbol{b}_{IV} - \boldsymbol{b})$ (see the formula sheet). If the null is correct, and there is no endogeneity, then both estimators are consistent, and the test statistic should be small. If there is endogeneity then LS is inconsistent and the difference between IV and LS will make the test statistic large. As always, a large test statistic leads to a small p-value, and to rejection of the null.

---

d) Explain what influences the precision (efficiency) of the IV estimator. Hint: The asymptotic distribution of the simple IV estimator is:

$$\sqrt{n}\left(\boldsymbol{b_{IV}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left[\mathbf{0}, \sigma^2 Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right]$$

where $\text{plim}(\frac{1}{n}Z'X)^{-1} = Q_{ZX}^{-1}$, for example.

The asymptotic efficiency of $\boldsymbol{b}_{IV}$ will be higher the more highly correlated are Z and X.

---

3. Consider the usual population model, except that there are two groups, A and B. There is a dummy variable in the data that differentiates group membership: $D_i = 1$ if the observation is from A, and $D_i = 0$ if the observation is from B. The groups A and B only determine the **variance** in the model (the dummy $\boldsymbol{D}$ is not correlated with $\boldsymbol{y}$ and is not needed in the X matrix). Specifically,

$$\text{var}\left(\epsilon_i \mid D_i = 0\right) = \sigma^2$$
$$\text{var}\left(\epsilon_i \mid D_i = 1\right) = 2\sigma^2$$

a) Explain how $\boldsymbol{D}$ can be used to implement GLS, through the "weighted least squares" interpretation.

The idea behind WLS is that observations with lower variance should receive more weight than observations with higher variance. This improves the efficiency of the estimator. The D variable tells which observations have twice as much variance as the others. To equalize the variance of all observations (to recover A.4), we could multiple observations in the $D = 1$ group by $1/\sqrt{2}$, for example. This would equalize the variance across groups.

---

b) What are the Σ and P matrices?

The Σ matrix is still unknown since σ^2 is unknown, but the Ω matrix would contain diagonal elements equal to 1 when $D_i = 0$ and equal to 2 when $D_i = 1$. The P matrix would contain the reciprocal of the square roots of these values.

---

c) Show that $\text{V}(P\epsilon) = \sigma^2 I_n$.

$$\begin{aligned} V[P\boldsymbol{\epsilon}] &= PV(\boldsymbol{\epsilon})P' \\ &= P\left(\sigma^2\Omega\right)P' = \sigma^2 P\Omega P' \end{aligned}$$

Because P is both square and non-singular, note that:

$$\begin{aligned} \sigma^2 P\Omega P' &= \sigma^2 P\left(\Omega^{-1}\right)^{-1} P' \\ &= \sigma^2 P\left(P'P\right)^{-1} P' \\ &= \sigma^2 P P^{-1}\left(P'\right)^{-1} P' = \sigma^2 I \end{aligned}$$

---

d) If you didn't know that $\text{var}\left(\epsilon_i \mid D_i = 1\right) = 2 \times \text{var}\left(\epsilon_i \mid D_i = 0\right)$ *exactly* (but you knew that the variances were different between the two groups), how might you implement FGLS? (You can just describe, without any math, how you would approach the problem).

You could estimate the model by LS, and collect the residuals. You could take the average squared residual for both groups. This would give a consistent estimator for proportionality of the variance between the two groups. The Ω matrix would then be estimated using the average squared residuals of the two groups, and FGLS could be implemented.

---

4. a) Derive the Newton-Raphson algorithm, either through a Taylor-series expansion, or graphically.

See Figure 10.4 on page 96.

---

b) When would you need to use a numerical algorithm such as Newton-Raphson?

See the opening paragraph in Section 10.4: when the model is non-linear in the parameters, and can not be linearized or approximated using polynomials, logs, splines, etc.

---

c) Why might different starting values for the Newton-Raphson algorithm, lead to different solutions, or to no solution at all?

There may not be a global minimum, different starting values could lead to different local minima and thus different estimates. If the algorithm hits a point where the second derivative is zero, or if the algorithm oscillates (bounces back and forth between values), then the algorithm may never converge.

---

d) Explain why we might want to estimate the gravity model:

$$T_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \eta_{ij}$$

by NLS instead of by log-linearizing the model and using LS, where the α are parameters to be estimated and η_{ij} are the error terms.

See the two paragraphs below equation 10.9 on page 97.

---

5. a) Derive the variance of ϵ_t, where ϵ_t follows an AR(1) process.

See page 116, where we take the variance of equation 12.1.

---

b) Suppose that you estimate the model $y_t = \beta y_{t-1} + \epsilon_t$, but that ϵ_t is AR(1). Show that b is inconsistent.

See section 12.3. Alternatively, in equation 12.3, iterate y_t back one period. See that y_{t-1} depends on ϵ_{t-1}. Since ϵ_t also depends on ϵ_{t-1}, there is endogeneity (A.5 is violated).

---

c) What does it mean for a process to have "infinite memory"?

An AR process has infinite memory; the entire past history of shocks has determined the current value of ϵ_t. See equation 12.1. An AR process can always be written as an infinite order MA process.

---

d) Explain what a spurious regression is, in the context of two random walks.

A spurious regression is one in which a significant relationship between variables is estimated, where none truly exist. In the context of time series, this usually refers to the situation where two independent random walks are regressed one on the other. Even if they truly have no relation to each other, LS will estimate the relationship to be more and more significant and with a highere and higher R^2, as the sample size grows.

---

6. The probability function for the Poisson distribution is:

$$f\left(y_i \mid \lambda\right) = \frac{\lambda^{y_i}}{e^{\lambda} y_i!} \quad ; \quad y_i = 0, 1, 2, \ldots \quad ; \quad \lambda > 0$$

a) What is the joint log likelihood function for data that is Poisson distributed?

b) Derive the MLE for λ.

c) What are the properties of MLEs?

d) Explain how you would estimate the effect that an x regressor variable has on a y variable (where y is Poisson distributed).

END.