# Studienzentrum Gerzensee Doctoral Program in Economics
# Midterm Econometrics Exam

Bo Honore and Mark Watson

July 23, 1999

*Instructions:*

Write your Identification Number in the space provided below. (Don't give us your name, just your ID number.)

Sketch of Answers

ID Number: ____________________________________________

There are 120 points on this 120 minute exam. The number of possible points for each problem is shown in parentheses; unless we state otherwise, each part of a problem will have the same number of possible points. If you need additional space use the back of the exam sheet. Feel free to use your notes and any textbooks that you may find useful.

1. (15 Points) Complete the following calculations

   (a) Suppose $X \sim N(4, 25)$. Find $P(-1 < X < 5)$

$$P(-1 < X < 5) = P\left(\frac{-1-4}{5} < \frac{X-4}{5} < \frac{5-4}{5}\right)$$

$$= P(-1 < Z < .2) = .42$$

   (b) Suppose $X \sim N(20, 16)$. Find $P(X > 16)$

$$P\left(\frac{X-20}{4} > \frac{16-20}{4}\right) = P(Z > -1) = .84$$

(c) Suppose $X \sim \chi^2_3$. Find $P(X \geq 6.25)$

.10

(d) Suppose $X \sim F_{2,8}$. Find $P(X \leq 8.65)$

.99

(e) Suppose $X \sim t_{15}$. Find $P(1.34 \leq X \leq 2.6)$

.09

2. (15 Points) Suppose that $X_i \sim iid\ N_p(0, \Sigma)$ for $i = 1, ..., n$ where $X_i$ is a $n \times 1$ vector. Let $\alpha$ denote a non-zero $p \times 1$ vector and let

$$Y_n = \frac{\alpha' X_n X_n' \alpha}{\frac{1}{n-1} \sum_{i=1}^{n-1} \alpha' X_i X_i' \alpha}$$

(a) Prove that $Y_n \sim F_{1,n-1}$. (Hint: recognize that $\alpha' X_n$ is a scalar.)

$\alpha' X_i \sim iid\ N(0, \sigma^2)$ with $\sigma^2 = \alpha' \Sigma \alpha$

$\Rightarrow A = \left(\frac{\alpha' X_n}{\sigma}\right)^2 = \frac{\alpha' X_n X_n' \alpha}{\sigma^2} \sim \chi^2_1$

$B = \sum_{i=1}^{n-1} \left(\frac{\alpha X_i X_i' \alpha}{\sigma^2}\right) \sim \chi^2_{n-1}$

; these two R.V.s are independent.

The result then follows immediately by noting

$Y_n = \frac{A/1}{B/N-1}$

(b) Sketch a proof that $Y_n \xrightarrow{d} Y$ where $Y \sim \chi^2_1$ (as $n \to \infty$).

$D = \frac{1}{n-1} \sum a_i$ where $a_i = \left(\frac{\alpha' X_i}{\sigma}\right)^2 \sim \chi^2_1$

$\Rightarrow D \xrightarrow{p} E(a_i) = 1$

Thus denominator of $Y_n \xrightarrow{p} \sigma^2$

Numerator is $\sigma^2 \cdot [N(0,1)]^2$

Thus $Y_n \xrightarrow{d} [N(0,1)]^2 \sim \chi^2_1$

3. (15 points) A researcher is interested in computing the likelihood ratio statistic

$$LR = \frac{f(y,\theta_1)}{f(y,\theta_2)}$$

but finds the calculation difficult because the density function $f(.)$ is very complicated. However, it easy to compute the joint density $g(y,x,\theta)$ where $x$ denotes another random variable. Unfortunately, the researcher has data on $y$, but not on $x$. Show that the researcher can compute $LR$ using the formula

$$LR = \frac{f(y,\theta_1)}{f(y,\theta_2)} = E_{\theta_2}[\frac{g(y,x,\theta_1)}{g(y,x,\theta_2)}|y]$$

where the notation $E_{\theta_2}[\bullet|y]$ means that the expectation should be taken using the distribution conditional on $y$ and using the parameter value $\theta_2$.

$$E_{\theta_2}\left[\frac{g(y,x,\theta_1)}{g(y,x,\theta_2)}\Big|\, y\right] = \int_x \underbrace{\frac{g(y,x,\theta_1)}{g(y,x,\theta_2)}} f(x,\theta_2|y)\,dx$$

$$= \int_x \qquad \frac{g(y,x,\theta_2)}{f(y,\theta_2)}\,dx$$

$$= \int_x \frac{g(y,x,\theta_1)}{f(y,\theta_2)}\,dx$$

$$= \frac{1}{f(y,\theta_2)}\int_x g(y,x,\theta_1)\,dx$$

$$= \frac{f(y,\theta_1)}{f(y,\theta_2)}$$

4. (15 Points) Let $x_t$ follow a stationary $ARMA(p,q)$ process and let $y_t = \sum_{i=-k}^{k} a_i x_{t+i}$. A researcher has data on $x_t$ for $t = 1, ..., T$ and wants to use these sample values of $x$ to construct an estimate of $y_T$. Since $y_T$ depends on values of $x_t$ outside the sample period, she constructs the estimated value of $y_T$ using a two-step process: first she uses the observed $x$ data to construct forecasts of $x_{T+1}, x_{T+2}, ..., x_{T+k}$; second, she estimates $y_T$ by

$$\widehat{y}_T = \sum_{i=-k}^{0} a_i x_{T+i} + \sum_{i=1}^{k} a_i x_{T+i|T}$$

where $x_{T+i|T} = E(x_{T+i}|\{x_t\}_{t=1}^{T})$ denotes the forecast of $x_{T+i}$ constructed from the observed values of $x$.

Show that $\widehat{y}_T$ is the minimum mean squared error (mmse) estimator of $y_T$. (Hint: remember that the mmse estimator is given by the regression function.)

The mmse estimator is

$$\tilde{y}_T = E\left[y_T \mid \{x_t\}_{t=1}^{T}\right] = E\left[\sum_{i=-k}^{K} a_i x_{T+i} \mid \{x_t\}_{t=1}^{T}\right]$$

$$= \sum_{i=-k}^{K} a_i E\left[X_{T+i} \mid \{X_t\}_{t=1}^{T}\right] = \hat{y}_T$$

5. (12 points) Suppose that you have i.i.d. observations from

$$y_i = x_i^* \beta + \varepsilon_i$$

Unfortunately, you do not observe $x_i^*$. Instead, you observe

$$x_i = x_i^* u_i + v_i$$

where $\varepsilon_i$, $x_i^*$, $v_i$ and $u_i$ are independent of each other and $E[\varepsilon_i] = E[v_i] = 0$, $E[u_i] = 1$, and that all relevant moments exist. Suppose you regress $y_i$ on $x_i$ (that is, you run OLS *without* a constant).

(a) (5 points) Write the expression for the ordinary least squares estimator. What is it's probability limit? Is it consistent?

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum y_i x_i}{\sum x_i^2} = \frac{\frac{1}{n}\sum y_i x_i}{\frac{1}{n}\sum x_i^2} = \frac{\frac{1}{n}\sum (x_i^*\beta + \varepsilon_i)(x_i^* u_i + v_i)}{\frac{1}{n}\sum (x_i^* u_i + v_i)^2} \\
&= \frac{\frac{1}{n}\sum_i \left\{ (x_i^*)^2 \beta u_i + \varepsilon_i x_i^* u_i + x_i^* \beta v_i + \varepsilon_i v_i \right\}}{\frac{1}{n}\sum \left\{ (x_i^* u_i)^2 + v_i^2 + 2 v_i x_i^* u_i \right\}}
\end{aligned}$$

so

$$p\lim \widehat{\beta} = \frac{E\left[(x_i^*)^2 \beta u_i\right] + E[\varepsilon_i x_i^* u_i] + E[x_i^* \beta v_i] + E[\varepsilon_i v_i]}{E\left[(x_i^* u_i)^2\right] + E[v_i^2] + 2E[v_i x_i^* u_i]} = \frac{E\left[(x_i^*)^2\right]\beta}{E\left[(x_i^*)^2\right] E[u_i^2] + E[v_i^2]}$$

So the OLS estimator is consistent only if $E[u_i^2] = 1$ and $E[v_i^2] = 0$ (in which case $x_i^* = x$)

(b) (7 points) Suppose that you have an instrument $z_i$. Under what conditions will the instrumental variables estimator be consistent for $\beta$? Does $z_i = 1$ satisfy those conditions?

$$\widehat{\beta}_{IV} = \frac{\sum y_i z_i}{\sum x_i z_i} = \frac{\frac{1}{n}\sum (x_i^* z_i \beta + z_i \varepsilon_i)}{\frac{1}{n}\sum (x_i^* z_i u_i + z_i v_i)}$$

so

$$p\lim \widehat{\beta}_{IV} = \frac{E[x_i^* z_i]\beta + E[z_i \varepsilon_i]}{E[x_i^* z_i u_i] + E[z_i v_i]}$$

A sufficient set of conditions for consistency is that $E[x_i^* z_i] \neq 0$, $u_i$ is independent of $(x_i^*, z_i)$ and $z_i$ is uncorrelated with $\varepsilon_i$ and $v_i$.

6. (24 points) This problem uses the attached regression output (seven regressions performed in EVIEWS attached to the end of the exam). Throughout this problem, consider the model

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + \varepsilon_i.$$

In answering questions (a)–(f), you may assume that all the usual regularity conditions for OLS are satisfied. (Hint: not all of the attached regressions are relevant)

(a) (2 points) Construct a 90% confidence interval for $\beta_1$.

$1.257 \pm 1.645 \times 0.368$

(b) (2 points) Test $\beta_3 = 2$ (against $\beta_3 \neq 2$) at the 5% level of significance. What do you conclude?

The statistic is

$$\frac{0.8425 - 2}{0.9296} = -1.24$$

The critical value is 1.96, so you do not reject.

(c) (2 points) Test $\beta_2 = -0.5$ (against $\beta_2 \neq -0.5$) at the 5% level of significance. What do you conclude?

The statistic is

$$\frac{-0.3738 - (-0.5)}{0.3739} = 0.33$$

(you do not reject)

(d) (6 points) Test the following three hypotheses (all at the 5% level of significance)

$$\beta_3 = 0$$

$$\beta_4 = 0$$

$$\beta_3 = \beta_4 = 0$$

Discuss the results. How does the third of the hypotheses differ from first considering the first hypothesis and then the second hypothesis?

The T-tests for the first two are 0.906 and 0.679. So you cannot reject either.

For the joint test compare the $R^2$ for the original regression to the $R^2$ for the restricted model

$$F = \frac{(R^2 - R_*^2)/J}{(1-R^2)(n-k)} = \frac{(0.278 - 0.143)/2}{(1-0.278)/95} = 8.88$$

(see Greene p 344). Under the null, F comes from an $F_{2,95}$ distribution, so you reject the null.

The conclusion is that even though you cannot reject that $\beta_3 = 0$ or $\beta_4 = 0$, you can reject that both are true.

(e) (6 points) Calculate the T–test–statistic for the hypothesis that $\beta_4 = 0$. Also calculate the F–test–statistic for the same hypothesis. How do the two tests–statistics compare? How do the critical values for the two tests compare? Explain their relationship.

$T = 0.679$ (from printout)

$$F = \frac{0.2779 - 0.2744}{(1 - 0.2779)/95} = .460 = T^2$$

But the square of a T-distributed random variable has an F-distribution. Hence inference based on T and on F will lead to the same conclusion

(see also discussion on page 343 of Greene)

(f) (6 points) Suppose you are confident that the variance of $\varepsilon_i$ does not depend on $x_{i1}$, $x_{i2}$ or $x_{i3}$, but that you worry that it might depend on $x_{i4}$. Explain how you would test whether the variance of $\varepsilon_i$ is unrelated to $x_{i4}$.

The LM-test could be calculated as $n \cdot R^2$ in the regression of $e_i^2$ on 1 and $X_{i4}$. Under the null it will have an asymptotic $\chi^2_{(1)}$ distribution

Large values are evidence against the null.

7. (12 points) Consider the regression

$$y_t = \beta_1 + y_{t-1}\beta_2 + \varepsilon_t$$

where

$$\varepsilon_t = \alpha\varepsilon_{t-1} + u_t$$

where the $u_t$'s are a sequence of independent and identically distributed random variables with $E[u_t] = 0$ and $V[u_t] = 1$. Assume that $-1 < \alpha < 1$.

(a) Suppose you regress $y_t$ on $y_{t-1}$ and a constant. Discuss the properties of the OLS estimator in this case. Be as explicit as you can.

It is inconsistent because $y_{t-1}$ is correlated with $\varepsilon_t$

To find the probability limit you must work out

$$\frac{\text{cov}(y_t, y_{t-1})}{V[y_t]}$$

(b) Suppose you know that $\alpha = 0.5$. Explain whether you can use this to estimate $\beta_1$ and $\beta_2$. Be as explicit as you can (but do not provide the relevant EVIEWS commands).

$$(*) \quad y_t - \tfrac{1}{2} y_{t-1} = \tfrac{1}{2}\beta_1 + \left(y_{t-1} - \tfrac{1}{2} y_{t-2}\right)\beta_2 + u_t$$

where $u_t$ is uncorrelated with $y_{t-1} - \frac{1}{2} y_{t-2}$

Hence you can estimate $(*)$ by OLS

(c) Suppose you know that $\beta_2 = 0.5$. Explain whether you can use this to estimate $\beta_1$ and $\alpha$. Be as explicit as you can (but do not provide the relevant EVIEWS commands).

$$y_t - \tfrac{1}{2} y_{t-1} = \beta_1 + \varepsilon_t$$

So estimate $\beta_1$ by sample average of $y_t - \frac{1}{2} y_{t-1}$

$\alpha$ can be estimated by the sample correlation between $y_t - \frac{1}{2} y_{t-1}$ and $y_{t-1} - \frac{1}{2} y_{t-2}$

8. (12 points) Define and briefly explain when and why the following are useful (we are looking for very short answers):

(a) The Eicker-White standard errors.

Standard error based on

$$\hat{V}[\hat{\beta}_{OLS}] = \left(\Sigma x_i x_i'\right)^{-1}\left(\Sigma e_i^2 x_i x_i'\right)\left(\Sigma x_i x_i'\right)^{-1}$$

It will result in asymptotically valid inference even if the errors in a regression model are heteroskedastic. It assumes random sampling (no serial correlation)

(b) Feasible GLS.

To do GLS, one must know the variance matrix of the errors (except for a constant). When this is not known, but replaced by an estimator, we obtain the FGLS. Subject to regularity conditions, it is asymptotically equivalent to GLS.

(c) Strict exogeneity.

In the context of a regression model it is the assumption that the error is uncorrelated with future explanatory variables ("no feedback").

(d) LM test for serial correlation.

A test for serial correlation in a regression model. There are many versions of it. It is derived from assuming normal errors, but it can be applied much more generally (depending on exactly what version is used). One version is calculated by $n \cdot R^2$ where $R^2$ comes from a regression of

$$e_t \text{ on } e_{t-1} \dots e_{t-p}, X_t$$

It can be applied even if $X_t$ is not strictly exogenous (unlike Durbin Watson)