Studienzentrum Gerzensee
Doctoral Program in Economics

Feb. 96

NAME/Exam ID: Suggested answers

Midterm Examination in Econometrics

There are 120 points on this 120 minute exam. The number of possible points for each question is shown in parentheses preceding the question. Please answer all questions on the exam sheet. If you need additional space use the back of the exam sheet. Feel free to use a calculator, your notes and any textbooks that you may find useful.

(7) 1. If X is distributed N(1,4), calculate P(1<X^2<9).

Let $Z = (X-1)/2$. Then $Z \sim N(0,1)$.

$$\begin{aligned} P(1 &< X^2 < 9) = P(-3 < X < -1) + P(1 < X < 3) \\ &= P\left(\frac{-3-1}{2} < \frac{X-1}{2} < \frac{-1-1}{2}\right) + P\left(\frac{1-1}{2} < \frac{X-1}{2} < \frac{3-1}{2}\right) \\ &= P(-2 < Z < -1) + P(0 < Z < 2) \\ &= (\Phi(-1) - \Phi(-2)) + (\Phi(2) - \Phi(0)) \\ &= 0.159 - 0.023 + 0.977 - 0.500 \\ &= 0.613 \end{aligned}$$

(20) 2. An economist wanted to investigate the relationship between the price of a computer and its characteristics, and ran a regression of the form:

price = b1 + b2x"multiplication time" + b3x"memory size" + b4x"access time" + error term

The results for the slope coefficients, b2, b3, and b4 were:

	coefficient	standard error	T-Stat	95% Conf. interval
mult. time	-0.04	0.08	-0.5	[-0.214, 0.134]
memory	0.58	0.0184	29.42	[0.54, 0.62]
access time	-0.15	0.1	-1.5	[-0.37, 0.07]

R^2 = 0.877

Number of Observations = 16

(9) (a) Fill in the missing values represented by "___'s" in the table of results.

You need to use the expressions

$$T = \frac{\text{coefficient}}{\text{standard error}} \qquad 95\% \text{ CI} = \text{coefficient} \pm t_{n-k,0.975} \cdot \text{standard error}$$

Note that we need a t-distribution with 12 degrees of freedom and $t_{12,0.975} = 2.179$

(5) (b) Test the hypothesis that the coefficient on access time is 0 (against the alternative that it differs from 0) at the 5% level of significance.

The t-statistics is −1.5. As this is less than 2.179 in absolute value, we do not reject that the coefficient is 0.

(6) (c) Construct an F-test for the hypothesis that all three slope coefficients (b2, b3, and b4) are equal to 0.

The F–test is given by (see Greene section 6.7):

$$F[K-1, n-K] = \frac{R^2/(K-1)}{(1-R^2)/(n-K)} = \frac{0.877/3}{(1-0.877)/12} = 28.52$$

If you test at a 5% level of significance, the critical is 3.49. As large values are critical you reject.

(23) 3. Consider the production function:

$$Y_t = \beta N_t^{\gamma} K_t^{\lambda} u_t, \quad t=1, \ldots, T,$$

where Y_t, N_t and K_t are the measured values of output, labor input, and the stock of capital, respectively, and u_t is the unobserved stock of "technology," which is assumed to follow the (random walk) stochastic process:

$$\mathrm{Log}(u_t) = \alpha + \mathrm{Log}(u_{t-1}) + \epsilon_t$$

where ϵ_t is NIID$(0,\sigma^2)$.

Assume that N_t and K_t are predetermined, in the sense that the firm chooses the values of these inputs before observing ϵ_t, the time "t" shock to technology.

(6) (a) How would you estimate β, γ, λ and σ^2? Be specific.

Taking logs

$$\log Y_t = \log \beta + \gamma \log N_t + \lambda \log K_t + \log u_t \qquad (*)$$

so

$$\begin{aligned} & (\log Y_t - \log Y_{t-1}) \qquad (**) \\ = \; & \gamma (\log N_t - \log N_{t-1}) + \lambda (\log K_t - \log K_{t-1}) + (\log u_t - \log u_{t-1}) \\ = \; & \gamma (\log N_t - \log N_{t-1}) + \lambda (\log K_t - \log K_{t-1}) + \alpha + \epsilon_t \end{aligned}$$

We could estimate γ, λ and α by OLS in that regression. σ^2 can be estimated by the usual estimator of the variance of the errors in a linear regression model.

β cannot be estimated consistently. To see why, write

$$\log u_t = t\alpha + \log u_0 + \sum_{s=1}^{t} \epsilon_s$$

Inserting this in (*), we see that we can never tell $\log u_0$ apart from β.

(6) (b) How would you test the hypothesis of constant-returns-to scale? Be specific.

Constant returns to scale means that $\gamma + \lambda = 1$. We could test this using an F–test in (**). In practice it is easier to define $\kappa = \gamma + \lambda - 1$ and write

$$\begin{aligned}(\log Y_t - \log Y_{t-1}) &= \gamma(\log N_t - \log N_{t-1}) + \lambda(\log K_t - \log K_{t-1}) + \alpha + \epsilon_t \\ &= (1 - \lambda + \kappa)(\log N_t - \log N_{t-1}) \\ &\quad + \lambda(\log K_t - \log K_{t-1}) + \alpha + \epsilon_t\end{aligned}$$

or

$$\begin{aligned}&(\log Y_t - \log Y_{t-1}) - (\log N_t - \log N_{t-1}) \\ = \; &\kappa(\log N_t - \log N_{t-1}) \\ &+\lambda((\log K_t - \log K_{t-1}) - (\log N_t - \log N_{t-1})) + \alpha + \epsilon_t\end{aligned}$$

Constant returns to scale can now be tested by testing κ (using a t–test)

(6) (c) Suppose that $\log(u_t)$ was generated by a stationary AR(1) process instead of a random walk. How would your answer to (a) change?

The answer depends on the exact assumptions made on N_t and K_t. If they are strictly exogeneous, then γ and λ can be estimated by OLS on (*). GLS (which in this case is Cochrane–Orcutt) would be more efficient. The variance of the errors is $\sigma^2/(1-\phi^2)$ where ϕ is the coefficient in the AR(1). The latter can be estimated from the correlation of the residuals, and we can then estimate σ^2 from the variance of the residuals.

If they are just predetermined, then OLS will not be consistent. The reason is that both the errors and the regressors may depend on the lagged errors. This is also the reason why 1–step Cochrane Orcutt is inconsistent (the first step does not consistently estimate the AR-parameter). Instead one must use some kind of (generalized) method of moments approach. One set of useful moments is

$$\begin{aligned}E[(\log u_t - \rho \log u_{t-1}) N_t] &= 0 \\ E[(\log u_t - \rho \log u_{t-1}) K_t] &= 0 \\ Corr(\log u_t, \log u_{t-1}) &= \phi\end{aligned}$$

Cochrane–Orcutt iterated until convergence can be interpreted as method of moments estimation using these moments.

We cannot separate β from α (the latter determines the mean of the errors, the former the intercept in the regression).

(5) (d) Suppose N_t was not predetermined, so that the firm chose labor input input <u>after</u> observed ϵ_t. How would your answer to (a) change.

$(\log N_t - \log N_{t-1})$ is not exogeneous. Depending on the assumptions that are made on the way in which the firm chooses N_t, it is possible that N_{t-1} is a valid instrument. It is also possible that N_{t-s} $(s > 1)$ can be used as instruments.

(30) 5. Let X_1, X_2, ..., X_n denote a random sample from a $N(0,\sigma^2)$ distribution.

(5) (a) Prove that $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} X_i^2$ is an unbiased estimator of σ^2.

Let $Y_i = X_i^2$. Then Y_i is distributed like σ^2 times a $\chi^2(1)$–distributed random variable, and hence $E\left[X_i^2\right] = E\left[Y_i\right] = \sigma^2$ and $V\left[X_i^2\right] = V\left[Y_i\right] = 2\sigma^4$. Also note that $Y_1, Y_2, \ldots, Y_n$ are independent.

$$E\left[\widehat{\sigma}^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[Y_i\right] = \sigma^2$$

(5) (b) Prove that the variance of $\hat{\sigma}^2$ is σ^4/n.

$$V\left[\widehat{\sigma}^2\right] = V\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V\left[Y_i\right] = \frac{2\sigma^4}{n}$$

(15) 4. Let f(x)=2x, 0<x<1, zero elsewhere, be the p.d.f. of X.

(5) (a) Compute E(1/X)

$$E[1/X] = \int_0^1 \frac{1}{x}(2x)\,dx = 2$$

(10) (b) Compute the Distribution function and p.d.f. of Y=1/X.

The support for Y is $(0,1)$. For $0 < y < 1$

$$P(Y < y) = P\left(\frac{1}{X} < y\right) = P\left(\frac{1}{y} < X\right) = \int_{1/y}^{1} (2x)\,dx = 1 - \frac{1}{y^2}$$

So the Cumulative Distribution Function is

$$F_Y(y) = \begin{cases} 0 & \text{if} \quad y \le 0 \\ 1 - \frac{1}{y^2} & \text{if} \quad 0 < y < 1 \\ 1 & \text{if} \quad 1 \le y \end{cases}$$

The density is found by differentiation the CDF

$$f_Y(y) = \begin{cases} 0 & \text{if} \quad y \le 0 \\ \frac{2}{y^3} & \text{if} \quad 0 < y < 1 \\ 0 & \text{if} \quad 1 \le y \end{cases}$$

You could also have used the standard expression for the density of a function of a random variable (Greene equation (3.41) or Hogg and Craig page 169).

(5) (c) Prove that $\hat{\sigma}^2$ converges in probability to σ^2.

Follows from the Law of Large Numbers.

(8) (d) Prove that $\hat{\sigma}^2$ is an efficient estimator of σ^2.

Calculate the Cramér–Rao lower bound:
To make the notation less confusing let $\theta = \sigma^2$

$$I(\theta) = V\left[\frac{\partial \ln f(X_i;\theta)}{\partial\theta}\right] = V\left[-\frac{1}{2\theta} - \frac{X_i^2}{2\theta^2}\right] = \left(\frac{1}{2\theta^2}\right)^2 V\left[X_i^2\right] = \left(\frac{1}{2\theta^2}\right)^2 2\theta^2 = \frac{1}{2\sigma^4}$$

The Cramér–Rao lower bound is (see Hogg and Craig page 377)

$$\frac{1}{nI(\theta)} = \frac{2\sigma^4}{n}$$

As $\hat{\sigma}^2$ achieves the Cramér-Rao lower bound, it is efficient.

(7) (e) How would you test the competing hypotheses $H_o:\sigma^2=1$ vs. $H_a:\sigma^2=1.3$? Is the test that you propose the most powerful? Explain.

The answer is given by the Neyman–Pearson Theorem (Hogg and Craig page 397).
The likelihood function is

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta) = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{x_i^2}{2\theta}\right)$$

The likelihood ratio is

$$\frac{L(1; x_1, x_2, \ldots, x_n)}{L(1.3; x_1, x_2, \ldots, x_n)} = 1.3^{n/2} \exp\left(-\sum_{i=1}^{n} \frac{0.6x_i^2}{2.6}\right)$$

The most powerful test rejects when the likelihood ratio is small. In other words, the most powerful test is one which rejects when $\sum_{i=1}^{n} X_i^2$ is large.

How large?

To determine this, observe that under the null $\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$.

(25) 6. Consider the regression model:

$$y_t = x_t\beta + \epsilon_t, \quad t=1,\ldots,n$$

where ϵ_t is distributed $Niid(0,\sigma_\epsilon^2)$, and x_t is a scalar random variable generated by the AR(1) model

$$x_t = \phi x_{t-1} + e_t$$

where e_t is $Niid(0,\sigma_e^2)$, and $0<|\phi|<1$. Assume that ϵ_t and e_{t-k} are statistically independent for all t and k. Let $\hat{\beta}$ denote the OLS estimator of β.

(7) (a) Is $\hat{\beta}$ an unbiased estimator of β? Explain (provide a proof).

$$\widehat{\beta} = \frac{\sum_t x_t y_t}{\sum_t x_t^2} = \beta + \frac{\sum_t x_t \epsilon_t}{\sum_t x_t^2}$$

$$\begin{aligned} E\left[\widehat{\beta}\middle| x_1,\ldots,x_n\right] &= \beta + E\left[\frac{\sum_t x_t\epsilon_t}{\sum_t x_t^2}\middle| x_1,\ldots,x_n\right] \\ &= \beta + \frac{\sum_t x_t E\left[\epsilon_t | x_1,\ldots,x_n\right]}{\sum_t x_t^2} \\ &= \beta + \frac{\sum_t x_t \cdot 0}{\sum_t x_t^2} \\ &= \beta \end{aligned}$$

By the law of iterated expectations (Greene (3–65))

$$E\left[E\left[\widehat{\beta}\middle| x_1,\ldots,x_n\right]\right] = E\left[\beta\right] = \beta$$

(6) (b) Discuss how the precision of $\hat{\beta}$ is affected by the value of the parameter ϕ.

See next page

Let $\tilde{\beta}$ denote the instrumental estimator for β constructed using x_{t-1} as an instrument for x_t.

(6) (c) Is $\tilde{\beta}$ an unbiased estimator of β? Explain (provide a proof).

$$\widetilde{\beta} = \frac{\sum_t x_{t-1} y_t}{\sum_t x_{t-1} x_t} = \beta + \frac{\sum_t x_{t-1} \epsilon_t}{\sum_t x_{t-1} x_t}$$

$E\left[\widetilde{\beta}\right] = \beta$ then follows exactly as in (a).

First consider the conditional variance of $\widehat{\beta}$:

$$\begin{aligned}
V\left[\widehat{\beta}\middle| x_1,\ldots,x_n\right] &= V\left[\frac{\sum_t x_t\epsilon_t}{\sum_t x_t^2}\middle| x_1,\ldots,x_n\right] \\
&= \frac{\sum_t x_t^2 V\left[\epsilon_t| x_1,\ldots,x_n\right]}{\left(\sum_t x_t^2\right)^2} \\
&= \frac{\sum_t x_t^2\sigma_\epsilon^2}{\left(\sum_t x_t^2\right)^2} \\
&= \frac{\sigma_\epsilon^2}{\sum_t x_t^2}
\end{aligned}$$

By Greene (3–70)

$$\begin{aligned}
V\left[\widehat{\beta}\right] &= V\left[E\left[\widehat{\beta}\middle| x_1,\ldots,x_n\right]\right] + E\left[V\left[\widehat{\beta}\middle| x_1,\ldots,x_n\right]\right] \\
&= V\left[\beta\right] + \sigma_\epsilon^2 E\left[\frac{1}{\sum_t x_t^2}\right] \\
&= \sigma_\epsilon^2 E\left[\frac{1}{\sum_t x_t^2}\right]
\end{aligned}$$

The question then is how ϕ affects $E\left[\frac{1}{\sum_t x_t^2}\right]$. The answer is not obvious. The following suggests that the precision of $\widehat{\beta}$ is high when ϕ is close to -1 or 1.

$$E\left[x_t^2\right] = E\left[\left(\phi x_{t-1}+e_t\right)^2\right] = \phi^2 E\left[x_{t-1}^2\right] + \sigma_e^2$$

Using stationarity $E\left[x_t^2\right] = E\left[x_{t-1}^2\right]$, and hence

$$E\left[x_t^2\right] = \frac{\sigma_e^2}{1-\phi^2}$$

so the closer ϕ^2 is to 1, the bigger x_t^2 will tend to be. Hence $\left(\sum_t x_t^2\right)^{-1}$ will tend to be lower. A more formal answer would require calculating $E\left[\frac{1}{\sum_t x_t^2}\right]$, which is quite difficult.

Alternatively, consider the large sample distribution (see the notes) of $\widehat{\beta}$

$$\sqrt{n}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma_\epsilon^2 E\left[x_t^2\right]^{-1}\right)$$

or

$$\sqrt{n}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \frac{\sigma_\epsilon^2\left(1-\phi^2\right)}{\sigma_e^2}\right)$$

It is clear that the variance of the asymptotic distribution of $\widehat{\beta}$ is smaller when ϕ^2 is close to 1.

(6) (d) Is $\hat{\beta}$ more efficient that $\tilde{\beta}$? Explain (if time allows, provide a proof.)

Conditional on $x_1, \ldots, x_n$ it is certainly no less efficient. This is the conclusion of the Gauss–Markov Theorem. The two estimators are equally efficient only if they are identical with probability 1. This will be the case only if $\sigma_e^2 = 0$