Studienzentrum Gerzensee Doctoral Program in Economics
Final Econometrics Exam

Bo Honoré and Mark Watson

March 2000

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

ID Number: Sketch of Answers

There are 180 points on this 180 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 your notes and any textbooks that you may find useful.

1. (15) Suppose that x_t follows an MA process

$$x_t = \varepsilon_t - \theta\varepsilon_{t-1}$$

where $\varepsilon_t \sim NIID(0,1)$. Let $z_t = x_t x_{t-1} + \theta$.

(a) (5) Show that $E(z_t) = 0$.

$$z_t = x_t x_{t-1} + \theta$$

$$= (\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-1} - \theta\varepsilon_{t-2}) + \theta$$

$$= [\varepsilon_t\varepsilon_{t-1} - \theta\varepsilon_t\varepsilon_{t-2} - \theta\varepsilon_{t-1}^2 + \theta^2\varepsilon_{t-1}\varepsilon_{t-2} + \theta]$$

$$E(z_t) = E[\quad]$$

$$0 \qquad 0 \qquad -\theta \qquad 0 \qquad \theta$$

$$= -\theta + \theta = 0$$

(b) (10) Prove that $T^{-1}\sum_{t=1}^{T} x_t x_{t-1} \xrightarrow{p} \theta$

(i) $T^{-1}\sum \varepsilon_t \varepsilon_{t-1} = a_T$

$E(a_T)=0$; $VAR(a_T) = \frac{1}{T^2}$ $\Rightarrow$ $a_T \xrightarrow{p} 0$

(ii) $T^{-1}\sum \varepsilon_t \varepsilon_{t-2}$

$T^{-1}\sum \varepsilon_{t-1} \varepsilon_{t-2}$ $\longrightarrow$ similar calculation

(iii) $\frac{1}{T}\sum \varepsilon_{t-1}^2 = b_T$

$E(b_T) = 1$ $VAR(b_T) = \frac{3}{T}$ $\Rightarrow$ $b_T \xrightarrow{p} 1$

and Thus $T^{-1}\sum X_t X_{t-1} \xrightarrow{p} -\theta$

(Note minus sign)

2. (30) Suppose that y_t follows the AR(1) process

$$y_t = \mu + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where ε_t is $iid(0, \sigma^2)$, and $|\rho| < 1$. Let

$$\widehat{\mu}^{ols} = T^{-1} \sum y_t$$

(a) (8) Show that $\sqrt{T}(\widehat{\mu}^{ols} - \mu) \xrightarrow{L} N(0, V_{ols})$ and derive an expression for V_{ols}.

$$\sqrt{T}(\hat{\mu}^{OLS} - \mu) = \frac{1}{\sqrt{T}} \sum u_t$$

$$u_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$$

and so $\frac{1}{\sqrt{T}} \sum u_t =$

$$(1-\rho)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t + \underbrace{a_T}_{\xrightarrow{p} 0} + \underbrace{b_T}_{\xrightarrow{p} 0}$$

$$\downarrow$$

$$N(0, V_{OLS})$$

$$= (1-\rho)^{-2} \cdot \sigma_\varepsilon^2$$

$$a_T = \varepsilon_0(\rho + \rho^2 + \rho^3 + \cdots) + \varepsilon_{-1}(\rho^2 + \rho^3 + \cdots) + \varepsilon_{-2}(\rho^3 + \rho^4 + \cdots) + \varepsilon_{-3}(\rho^4 + \cdots) \; \vdots$$

Thus

$$V_{OLS} = (1-\rho)^{-2} \sigma_\varepsilon^2$$

$$b_T = -\varepsilon_T(\rho + \rho^2 + \rho^3 + \rho^4 + \cdots) - \varepsilon_{T-1}(\rho^2 + \rho^3 + \rho^4 + \cdots) - \varepsilon_{T-2}(\rho^3 + \rho^4 + \cdots) - \varepsilon_1(\rho^{T+1} + \rho^{T+2} + \cdots)$$

and a straight forward calculation shows $a_T \xrightarrow{p} 0$ and $b_T \xrightarrow{p} 0$

(b) (5) Suppose that $\varepsilon_0 = 0$. Explain how you would construct the GLS estimator of μ.

write the model as

$$y_1 = \mu + \varepsilon_1$$

$$y_t - \rho\, y_{t-1} = (1-\rho)\mu + \varepsilon_t \qquad t = 1, \cdots, T$$

RUN OLS

$$\begin{bmatrix} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} = \begin{bmatrix} 1 \\ (1-\rho) \\ (1-\rho) \\ \vdots \\ (1-\rho) \end{bmatrix} \mu + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

(c) (7) Let $\widehat{\mu}^{gls}$ denote the GLS from part c. Show that $\sqrt{T}(\widehat{\mu}^{gls} - \mu) \xrightarrow{L} N(0, V_{gls})$ and derive an expression for V_{gls}.

$$\hat{\mu}^{GLS} - \mu = \lambda \left[\varepsilon_1 + (1-\rho) \sum_{t=2}^{T} \varepsilon_t \right]$$

where $\lambda = \left[1 + (T-1)(1-\rho)^2 \right]^{-1}$

Thus

$$\sqrt{T}(\hat{\mu}^{GLS} - \mu) = \sqrt{T}\lambda\, \varepsilon_1 + T\lambda\, (1-\rho)^2 \frac{1}{\sqrt{T}} \sum \varepsilon_t$$

$\sqrt{T}\lambda \to 0$

$T\lambda \to (1-\rho)^{-2}$

$$\Rightarrow \quad \sqrt{T}(\hat{\mu}^{GLS} - \mu) = \underbrace{(1-\rho)^{-1} \frac{1}{\sqrt{T}} \sum \varepsilon_t}_{\downarrow d \; N(0, V_{GLS})} + o_p(1)$$

where $V_{GLS} = (1-\rho)^{-2} \cdot \sigma^2$

(d) (5)Is the GLS estimator better that the OLS estimator? Explain.

No -- both have same Limities Variance.

(e) (5) Suppose that $T = 100$, and $\hat{\mu}^{ols} = 2$. Let $\hat{u}_t = y_t - \hat{\mu}^{ols}$. The regression of $\hat{u}_t$ onto $\hat{u}_{t-1}$ yields a regression coefficient of 0.4 and the standard error of the regression is 1.1. Construct a 95% confidence interval for μ.

$\hat{\rho} = 0.4$

$\hat{\sigma} = 1.1$

$SE(\hat{\mu}) = \frac{1}{\sqrt{T}} \frac{\hat{\sigma}}{1-\hat{\rho}} = \frac{1}{10} \frac{1.1}{.6} = .1833$

Thus 95% CI is (approximately)

$\underset{\hat{\mu}}{2} \pm \underset{(\text{or } 1.96)}{2} (.1833) =$

1.64 to 2.36

3. (20) Suppose that two variables y and x are cointegrated with cointegrating vector $(1 \ -\beta)'$, so that $y_t - \beta x_t$ is an $I(0)$ variable. I want to construct a confidence interval for β and a colleague suggests that I can construct a confidence interval using the following intuitive idea:

   For a candidate value of β, say β_o form $z_t = y_t - \beta_o x_t$. Carry out a "unit-root" test on z_t (say, the appropriate Dickey-Fuller test described in Hamilton). If I reject the unit root null (so that z_t appears stationary) then β_o is potentially the correct value of β and should be included in the confidence interval. If I fail to reject the unit root null (so that z_t appears to be non-stationary) then β_o should not be included in the confidence interval.

   I implement this idea as follows. Let $\tau(z)$ denote the Dickey-Fuller test statistic constructed from a time series z_t, and let C denote the (symmetric) 5% critical value, so that $P(|\tau(z)| > C) = .05$ under the unit root null. Then a 95% confidence interval for β is formed as

   $$B = \{\beta | \ |\tau(z)| > C \text{ where } z = y - \beta x\}$$

   Comment on this procedure. Will it produce a valid 95% confidence interval?

<u>NO</u>

There is a confusion here about the Null : Alternative.

If β_o is the true value of β then

$z_t(\beta_o)$ is $I(0)$.

When a DF test is carried out using z_t, the Null is that $z_t \sim I(1)$ (NOT $z_t \sim I(0)$)

Thus this procedure will not produce a valid conf interval.

4. (15) A macroeconomist is interested in estimating a structural vector autoregression involving quarterly values of y (detrended real GDP in logarithms) and R (the central bank's short-term interest rate). The VAR has two equations: (1) A central bank reaction function that relates R to y and lagged values of R and y, and (2) A dynamic "IS" equation that relates y to R and lagged values of R and y. The disturbances in the two equations are assumed to be mutually uncorrelated. From studying the central bank's procedures the economist knows that, other things constant, the bank raises nominal interest rates by 25 basis points ($\Delta R_t = 0.25$) when it sees GDP increase 1% faster than trend gowth ($\Delta y_t = .01$).

(a) (5) How should this information be incorporated in the VAR?

The Reaction Function should be

$$R_t = 25\, Y_t + \sum_{i=1}^{p} \lambda_i R_{t-i} + \sum_{i=1}^{p} \beta_i Y_{t-i} + \varepsilon_t^R$$

To impose the coef of 25 on Y_t write the equation as

$$(R_t - 25\, Y_t) = \sum \lambda_i R_{t-i} + \sum \beta_i Y_{t-i} + \varepsilon_t^R$$

(b) (5) Using this information, is the VAR identified?

Yes in a Bivariate VAR with uncorrelated shocks one additional Restriction is required. Here the Restriction is that the coef. on Y_t in the R_t equation is 25.

(c) (5) How would you estimate the VAR?

① Regress $R_t - 25Y_t$ on lags of R_t & Y_t
this is Reaction Function

② "Regress" Y_t on R_t & lags of R & Y.
This Requires instruments. Valid instruments are the lagged values of R & Y and the Residual from the Reaction Function.

5. (10) Suppose that y_t follows the process

correction should read

$$y_t = h_t^{\frac{1}{2}} \cdot \varepsilon_t$$

$$y_t = h_t^{\frac{1}{2}}\varepsilon_t$$

where ε_t is $iid(0,1)$ and h_t follows one of two process:

$$\text{ARCH(1): } h_t = \alpha_0 + \alpha_1 y_{t-1}^2$$

or

$$\text{GARCH(1,1): } h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 y_{t-1}^2$$

(a) (5) Looking at a time series plot of y_t^2 how would you decide whether the process was best described as an ARCH(1) or a GARCH(1,1)?

Solving the GARCH Model yields

$$h_t = (1-\beta_1)^{-1}\beta_0 + \beta_2 \cdot \sum_{i=1}^{\infty} \beta_1^{i-1} y_{t-i}^2$$

Thus, the persistence in variance will be longer lasting in the GARCH Model than in the ARCH Model

(b) (5) (b) How could you use formal econometric analysis to decide whether the process was best described as an ARCH(1) or a GARCH(1,1)?

The Models are Nested with the ARCH model a special case of GARCH with $\beta_1 = 0$

Thus, Estimate GARCH model: test

H_0: $\beta_1 = 0$

6. (6) A positive random variable (duration), T, has hazard given by $\lambda(t) = 1 + t$. Find the density and CDF of T.

$$F(t) = 1 - S(t) = 1 - \exp\left[-\int_0^t \lambda(s)\,ds\right] = 1 - \exp\left[-t - \frac{1}{2}t^2\right]$$

$$f(t) = \lambda(t) \cdot S(t) = (1+t)\exp\left(-\int_0^t \lambda(s)\,ds\right)$$

7. (15) You have a sample of n independent observations. For each observation, you have a discrete variable, y, which takes values 0 and 1. You also have a regressor, x. Suppose you estimate a logit model in order to characterize the relationship between y and x (you use x and a constant as explanatory variables). Let α be the constant and let β be the coefficient on x. Suppose you estimate α to be 1 and β to be -1. Suppose further that you have estimated the covariance matrix of $\widehat{\alpha}$ and $\widehat{\beta}$ to be $\begin{pmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{pmatrix}$

(a) (3) What would be your estimate of $P(y = 1|x = 1)$?

In general

$$\hat{P} = \frac{\exp(\hat{\alpha} + \hat{\beta})}{1 + \exp(\hat{\alpha} + \hat{\beta})}$$

Here

$$\hat{P} = \frac{\exp(1-1)}{1+\exp(1-1)} = \frac{1}{2}$$

(b) (6) Test whether $P(y = 1|x = 1) = 0.4$ at a 5% level of significance..

Method 1

$$P(y=1|x=1) = 0.4 \Leftrightarrow \frac{\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)} = .4 \Leftrightarrow \alpha+\beta = \log\frac{2}{3} \approx -0.4$$

$$\hat{\alpha} + \hat{\beta} = 0 \qquad \hat{V}[\hat{\alpha} + \hat{\beta}] = 0.3$$

Test stat : $\dfrac{0-(-0.4)}{\sqrt{0.3}} = 0.73$ Do not reject !

Method 2

You could also use the δ-method to find

$$\hat{V}[\hat{P}] = \hat{V}[\exp(\hat{\alpha}+\hat{\beta})/(1+\exp(\hat{\alpha}+\hat{\beta}))] = \left(\tfrac{1}{4}\right)^2 \cdot 0.3$$

Test stat : $\dfrac{0.5-0.4}{\sqrt{\left(\frac{1}{4}\right)^2 \cdot 0.3}} \approx 0.73$

(c) (6) Construct a 95% confidence interval for $P(y = 1|x = 1) - P(y = 1|x = -1)$.

Use the δ-method

Let $$g(\alpha, \beta) = \frac{\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)} - \frac{\exp(\alpha-\beta)}{1+\exp(\alpha-\beta)}$$

Then

$$\sqrt{n}\left(g(\hat\alpha, \hat\beta) - g(\alpha,\beta)\right) \xrightarrow{d} N\left(0, \frac{\partial g}{\partial(\alpha,\beta)} V \frac{\partial g}{\partial(\alpha,\beta)}'\right)$$

provided that

$$\sqrt{n}\left(\begin{pmatrix}\hat\alpha \\ \hat\beta\end{pmatrix} - \begin{pmatrix}\alpha \\ \beta\end{pmatrix}\right) \xrightarrow{d} N(0, V)$$

The resulting CI is $(-0.68, -0\ 10)$

8. (21) Suppose that you have n independent and identically distributed observation of (y, x, z) from the nonlinear regression model

$$y = x\beta + (x\beta)^3 + \varepsilon, \qquad E[\varepsilon | x, z] = 0 \quad V[\varepsilon | x, z] = \exp(z)$$

where y, x and z are all one-dimensional and β is the parameter of interest. Assume that all relevant moments exits.

(a) (7) Find the asymptotic distribution of the non–linear least squares estimator, $\widehat{\beta}_1$, defined by minimizing

$$\sum_{i=1}^{n} \left(y_i - \left(x_i b + (x_i b)^3\right)\right)^2$$

over b.

$$\sqrt{n}\,(\hat{\beta}_1 - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1})$$

$$A = E\left[\left(\frac{\partial f}{\partial \beta}\right)^2\right] \qquad B = E\left[\varepsilon^2 \left(\frac{\partial f}{\partial \beta}\right)^2\right]$$

(note β is a scalar)

Here

$$A = E\left[\left(X_i + 3X_i^3 \beta^2\right)^2\right]$$

$$B = E\left[\left\{\exp(z)\,(X_i + 3X_i^3 \beta^2)\right\}^2\right]$$

(b) (7) Find the asymptotic distribution of the estimator, $\widehat{\beta}_2$, defined by minimizing

$$\sum_{i=1}^{n} \frac{1}{\exp(z_i)} \left(y_i - \left(x_i b + (x_i b)^3\right)\right)^2$$

over b.

Let $\tilde{y}_i = \frac{y_i}{\sqrt{\exp(z_i)}}$ $\quad \tilde{f} = \frac{X_i b + (X_i b)^3}{\sqrt{\exp(z_i)}}$ $\quad \tilde{\varepsilon}_i = \frac{\varepsilon_i}{\sqrt{\exp(z_i)}}$

then

$$\tilde{y} = \tilde{f}(X_i, z_i, \beta) + \tilde{\varepsilon}_i \quad \text{where } V[\varepsilon_i] = 1$$

Therefore

$$\sqrt{n}\,(\hat{\beta}_2 - \beta) \approx N(0,\ A^{-1} B A^{-1})$$

$$A = E\left[\exp(-z_i)\,(X_i + 3X_i^3 \beta^2)^2\right]$$

$$B = E\left[\exp(-z_i)\,(X_i + 3X_i^2 \beta^2)^2\right] = A$$

(c) (7) Find the joint asymptotic distribution of $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

Set it up as GMM. $\hat{\beta}_1$ and $\hat{\beta}_2$ are both determined by a moment condition.

9. (18) Consider observations (y_{it}, x_{it}) from the linear panel data model

$$y_{i,t} = x_{it}'\beta + y_{i,t-1}\gamma + \alpha_i + \varepsilon_{it}, \quad t = 1, ..., T, \quad i = 1, ..., N$$

where α_i is an unobserved individual–specific effect. You may also assume that you observe y_{i0}. (In answering this question, assume that N is much bigger than T, so that the relevant asymptotics is for $N \to \infty$ and T fixed)

(a) (12) How would you estimate β (without making additional assumptions about α_i) if ε_{it} is independent of $(y_{i0}, y_{i1}, y_{i2} ..., y_{i,t-1}, x_{i1}, x_{i2}, ..., x_{it})$?

The most obvious way to proceed is to difference out α_i:

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta y_{it-1}\gamma + \Delta\varepsilon_{it} \qquad t = 2 .. T$$

Estimate by IV using $x_1 \cdots x_{t-1}$ as instruments.

($x_t \cdots x_T$ are not valid instruments)

$y_0 \cdots y_{t-2}$ are also valid instruments.

(b) (6) How would your answer change, if you were willing to assume that ε_{it} is independent of $(y_{i0}, y_{i1}, y_{i2} ..., y_{i,t-1}, x_{i1}, x_{i2}, ..., x_{iT})$

now $x_1 \cdots x_T$ are all valid instruments.

10. (12) Consider the model

$$\begin{aligned} y_{1i}^* &= x_{1i}'\beta_1 + \varepsilon_{1i} \\ y_{2i}^* &= x_{2i}'\beta_2 + \varepsilon_{2i} \end{aligned}$$

where $\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix}$ is normally distributed (with mean 0 and covariance matrix Σ) and independent of (x_{1i}, x_{2i}). Suppose that you observe $(y_{1i}, y_{2i}, x_{1i}, x_{2i})$ where $y_{i1} = 1\{y_{i1}^* > 0\}$ and $y_{i2} = \max\{0, y_{i2}^*\}$. Find the likelihood function.

There are four cases. The contribution to the likelihood function for the four cases are:

1) $y_1 = 1,\ y_2 = 0$: $P\left(\frac{\varepsilon_1}{\sigma_1} > -\frac{X_1\beta_1}{\sigma_1},\ \frac{\varepsilon_2}{\sigma_2} < -\frac{X_2\beta_2}{\sigma_2}\right)$

(this can be calcated from a bivariate normal CDF, - it depends on $\mathrm{corr}(\varepsilon_1, \varepsilon_2)$)

2) $y_1 = 0,\ y_2 = 0$: $P\left(\frac{\varepsilon_1}{\sigma_1} < -\frac{X_1\beta_1}{\sigma_1},\ \frac{\varepsilon_2}{\sigma_2} < -\frac{X_2\beta_2}{\sigma_2}\right)$

3) $y_1 = 1\ \ y_2 > 0$: $f(y_2 \mid y_1^* > 0)\, P(y_1^* > 0) = \int_0^\infty f(y_1^*, y_2)\, dy_1^*$

$f(y_2) \int_0^\infty f(y_1^* \mid y_2)\, dy_1 =$

$\Phi\left(\left\{X_1\beta_1\sigma_1^{-1} + \sigma_{12}\sigma_1^{-1}\sigma_2^{-2}(y_2 - X_2\beta_2)\right\} \cdot \left[1 - \sigma_{12}^2\sigma_1^{-2}\sigma_2^{-2}\right]^{-1/2}\right) \sigma_2^{-1}\, \phi\left(\sigma_2^{-1}(y_2 - X_2\beta_2)\right)$

where Φ is the normal CDF, & ϕ the normal density.

Note that this exactly like the contribution to the likelihood function for a type 2 Tobit model.

4) $y_1 = 0\ \ y_2 > 0$: This case is obtained in the same way

$f(y_2 \mid y_1^* < 0) \cdot P(y_1^* < 0) = \int_{-\infty}^0 f(y_1^*, y_2)\, dy_1^*$

11. (18) Consider a random sample of size n from a discrete choice model with y equal to 0 or 1, and

$$P(y = 1|x^*) = p(x^*)$$

As you know, a linear regression of y on x^* will produce an estimator of the cofficients in the best linear approximation of $p(x^*)$ (in a mean squared error sense). Now imagine that x^* is measured with error, so what you observe is

$$x = x^* + v_1.$$

You also have a second measurement of x^*:

$$z = x^* + v_2$$

Assume that v_1 and v_2 have mean 0 and are independent, and that (v_1, v_2) is independent of (y, x^*). Suppose that you estimate the model

$$y = x\beta + \varepsilon$$

by instrumental variables using z as instruments (i.e., you calculate $\left(\sum_{i=1}^n z_i'x_i\right)^{-1}\left(\sum_{i=1}^n z_i'y_i\right)$). Will this result in a consistent estimator of the cofficients in the best linear approximation of $p(x^*)$?

$$\left(\sum z_i' x_i\right)^{-1} \sum z_i' y_i = \left[\tfrac{1}{n}\sum (x_i^* + v_{2i})'(x_i^* + v_{1i})\right]^{-1} \cdot \tfrac{1}{n}\sum (x_i^* + v_{1i})\, y_i$$

$$\xrightarrow{a.s.} \left[E[x^{*\prime} x^*] + 0 + 0 + 0\right]^{-1}\left(E[x^* y] + 0\right)$$

by the Law of Large Numbers and Slutsky.

Now $E[x_i^{*\prime} x^*]^{-1} E[x^* y] =$

$$E[x^{*\prime} x^*]^{-1} E[x^* p(x^*)]$$

Hence the answer is "yes"