STUDY CENTER
GERZENSEE

Swiss Program for Beginning Doctoral Students in Economics 2006

Midterm Exam in Econometrics

Friday, July 28, 2006, 14.00h - 16.00h

1. You are allowed to use all material that you want (lecture notes, books, etc.) with the exception of PC's.
2. Please do **not** mention your name on top of the pages, but only your identification number. (The exams will be graded anonymously.) If you use the back side of a page, put your ID number there as well.
3. Please answer questions on the same page or the back side of the page. (The exams will be separated in order to be sent to the different professors.)
4. Please use **a pen** to guarantee that your answers can be read without problems.
5. Please **write legibly**. Your exams will be photocopied for grading.
6. Answers should be **concise and precise!** The space provided should be sufficient to answer each question.
7. No cell phones!
8. Please have your ID card ready on the desk. The TA will check your ID cards and have you sign a list.
9. Good luck!

ID-Number:____

(20) 1. X and Y are random variables. The marginal distribution of X is U[0,1], that is X is uniformly distributed on [0,1]. The conditional distribution of $Y|X=x$ is U[x, 1+x] for $0 \leq x \leq 1$.

(5) (a) Derive the mean of X.

(5) (b) Derive the variance of X.

(5) (c) Derive the joint probability density function of X and Y.

(5) (d) Derive the mean of Y.

(15) 2. Let l denote a $p \times 1$ vector of 1's, I_p denote the identity matrix of order I, suppose the $p \times 1$ vector Y is distributed as $Y \sim \mathrm{N}(\mathbf{0}, \sigma^2 I_p)$, with $\sigma^2 = 4$, and let $M = I_p - (1/p)ll'$.

(5) (a) Show that M is idempotent.

(5) (b) Compute $E(Y'MY)$.

(5) (c) Compute Variance(YMY).

(25) 3. $X_i \sim$ iid U$[0,\theta]$, for $i = 1, \ldots, n$, and let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(5) (a) Show that $2\bar{X} \xrightarrow{p} \theta$.

(5) (b) Show that $\sqrt{n}(2\bar{X} - \theta) \xrightarrow{d} N(0,V)$ and derive an expression for V.

(15) (c) I am interested in the competing hypotheses H_0: $\theta = 1$ versus H_a: $\theta = 0.8$, and I decide to reject H_o if $\bar{X} \leq 0.45$.

(5) (c1) Using an appropriate approximation, compute the size of the test.

(5) (c2) Using an appropriate approximation, compute the power of the test.

(5) (c3) Is this the best test to use?

Problem 4 (15 points)

Suppose that you estimate a linear regression model using individuals from a (large) sample of married couples,

$$y_{ij} = x'_{ij}\beta + \varepsilon_{ij}$$

where $i = 1, \ldots, n$ denotes a couple, and $j = 1$ for the wives and $j = 2$ for the husbands. Assume that observations are independent and identically distributed across couples (so $(x_{i1}, \varepsilon_{i1}, x_{i2}, \varepsilon_{i2})$ has the same distribution for all i, and $(x_{i1}, \varepsilon_{i1}, x_{i2}, \varepsilon_{i2})$ is independent of $(x_{\ell 1}, \varepsilon_{\ell 1}, x_{\ell 2}, \varepsilon_{\ell 2})$ provided that $i \neq \ell$). Also assume that the model is correctly specified in the sense that $E\left[\varepsilon_{ij} \left| x_{i1}, x_{i2}\right.\right] = 0$ for $j = 1, 2$.

It is often unreasonable to assume that $(x_{i1}, \varepsilon_{i1})$ is independent of $(x_{i2}, \varepsilon_{i2})$. This implies that the standard theory for OLS is invalid and one cannot rely on the output from standard OLS packages for inference.

Explain in detail how you would test a hypothesis of the form

$$R'\beta = r$$

in this situation. Your test should be based on the asymptotic distribution of $R'\widehat{\beta}_{OLS} - r$.

State any additional assumptions that you make.

Problem 5 (15 points)

A recent paper augments a standard earning function with a variable for whether or not an individual uses a computer at work:

$$\ln y_i = \alpha + \delta C_i + X_i'\beta + \varepsilon_i$$

where y refers to real gross hourly wages, C is a dummy variable, which is equal to 1 if the individual uses a computer at work and zero otherwise, X is a vector of control variables, and ε is the error term. The parameters, α, δ and β, are estimated by OLS using data from 1991 and 2000. The results for δ are

	1991	2000
Computer Use	0.278 (0.021)	0.312 (0.025)

Robust standard errors in parentheses

(a) (5 points) Assume that the model above is correctly specified, and consider an individual who does not use a computer and who has an hourly wage of \$10 in 2000. Construct a 95% confidence interval for that person's wage-rate if she/he starts using a computer (everything else being equal).

(b) (10 points) Test whether the effect of computer use is the same in the two periods (assuming that the two samples are independent of each other).

Problem 6 (20 points)

Consider the linear regression model

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

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

$$x_i = x_i^* + u_i$$

where ε_i, x_i^* and u_i are independent of each other, $E[\varepsilon_i] = E[u_i] = 0$, and that all relevant moments exist.

(a) (6 points) Suppose you regress y on a constant and x. Write the expression for the ordinary least squares estimator of the slope parameter . What is its probability limit if you have a random sample of (y_i, x_i)?

(b) (6 points) Suppose you regress x on a constant and y. Write the expression for the ordinary least squares estimator of the slope parameter . What is its probability limit if you have a random sample of (y_i, x_i)?

(c) (8 points) Explain how the results in (a) and (b) can be used to estimate a bound on β. (Hint: Consider the three cases $\beta > 0$, $\beta = 1$ and $\beta < 0$ separately).

Problem 7 (10 points)

Consider the model from Chapter 3 of Hayashi.

- Assumption 3.1:

$$y_i = z_i'\delta + \varepsilon_i \qquad \text{but} \qquad E\left[z_i\varepsilon_i\right] \neq 0$$

- Assumption 3.2: (y_i, z_i, x_i) ergodic and stationary.
- Assumption 3.3: $E\left[g_i\right] = E\left[x_i\varepsilon_i\right] = E\left[x_i\left(y_i - z_i'\delta\right)\right] = 0$
- Assumption 3.4: $E\left[x_iz_i'\right]$ has rank L where $\dim\left(z_i\right) = L \leq K = \dim\left(x_i\right)$
- Assumption 3.5: $g_i = x_i\varepsilon_i$ is a m.d.s. with $S = E\left[g_ig_i'\right]$ finite.

You know from class that the two stage least squares estimator is the efficient GMM estimator if $E\left[\varepsilon_i^2 \middle| x_i\right] = \sigma^2$ (i.e. a constant). Find the asymptotic distribution of the two stage least squares estimator if $E\left[\varepsilon_i^2 \middle| x_i\right] = \sigma^2\left(x_i\right)$.