Studienzentrum Gerzensee Doctoral Program in Economics
Midterm Econometrics Exam

Bo Honore and Mark Watson

July 27 - 28, 1998

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

ID Number: ____

There are 120 points on this 120 minute exam. The number of possible points for each problem is shown in parentheses. 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.

ID Nr.

Problem 1 (20 points)
Suppose the $X_i \sim iidN(0, \sigma^2)$, $i = 1, ..., n$.

(a) Construct the MLE of σ

(b) Is the MLE of σ unbiased? Explain.

(c) Prove that $\sqrt{n}(\widehat{\sigma}_{MLE} - \sigma) \xrightarrow{d} N(0, W)$. Derive an expression for W.

(d) Suppose $\widehat{\sigma}_{MLE} = 4$ and $n = 100$. Construct a 95% confidence interval for σ

Problem 2 (10 points)

Suppose that X and Y are independent random variables, g and h are two functions, and $\mathbf{E}(g(X))$ and $\mathbf{E}(h(Y))$ exist. Prove

(a) $\mathbf{E}(g(X)h(Y)) = \mathbf{E}(g(X))\mathbf{E}(h(Y))$

(b) $\mathbf{E}(g(X)|Y = y) = \mathbf{E}(g(X))$

Problem 3. (15 points)
Suppose X_i, $i = 1, ..., n$, are a sequence of independent Bernoulli random variables with $Prob(X = 1) = p$. Based on a sample of size n, a researcher decides to test the null hypothesis

$$H_o : p = 0.45$$

using the following rule:

- Reject H_o if $\sum_{i=1}^{n} X_i \geq \frac{n}{2}$. Otherwise, do not reject H_o

(a) When $n = 2$ compute the exact size of this test.

(b) When $n = 50$ compute the size of the test. (Since n is large, you can use the Central Limit Theorem).

(c) Show that the test is consistent for the alternative $H_a : p = 0.55$. (Hint: Use Chebyshev's inequality.)

Problem 4. (15 points)
Suppose that X_t is generated by:

$$X_t = Y_t + Z_t$$

where Y and Z are given by

$$Y_t = 0.5Y_{t-1} + \varepsilon_t$$

and

$$Z_t = a_t - 0.3a_{t-1}$$

and where ε_t and a_t are mutually independent *iid* mean zero processes, each with unit variance. Assume that initial conditions Y_0 and a_0 are mutually independent, independent of future values of ε_t and a_t, and are chosen so that the Y and Z processes are covariance stationary.
Let $S_x(\omega)$, $S_Y(\omega)$ and $S_z(\omega)$ denote the spectra of X, Y, and Z.

(a) Show that $S_x(\omega) = S_Y(\omega) + S_z(\omega)$

(b) Sketch graphs of $S_x(\omega)$, $S_Y(\omega)$ and $S_z(\omega)$

(c) Prove that $X_t \sim ARMA(1,2)$, determine the value of the AR coefficient, and explain the calculations that you would perform to find the MA coefficients and error variance.

Problem 5. (12 points)

(a) "The residuals $e \equiv y - x\hat{\beta}_{\text{OLS}}$ is based on estimated quantities whereas $\varepsilon \equiv y - x\beta$ is based on the corresponding true values. It therefore follows that $E[\varepsilon'\varepsilon] < E[e'e]$." Is that a true or a false statement? Explain!

(b) "The OLS estimator (in a regression of Y on X) gives the best predictor of Y given X." Is that a true or a false statement? Explain!

Problem 6. (12 points)

Gasoline consumption is regressed on a constant, a price index for gasoline, disposable income, and a price index for public transportation. The data covers the period 1960-86 (so there are 27 observations).

We are interested in knowing whether there was a structural change in gasoline consumption after 1973. OLS regression using the whole period and before and after 73 yields the following residual sums of squares

1960–86: 25025

1960–73: 313.26

1974–86: 996.86

(a) Calculate an F–test–statistic for the hypothesis that the coefficient vectors are the same for the two periods.

(b) Discuss the assumptions that are needed for the F-test in (a). Which assumptions would you worry about?

Problem 7. (16 points)
A Cobb-Douglas production function

$$Y_i = AK_i^{\alpha} L_i^{\beta} u_i$$

was estimated by OLS using the model

$$y_i = \alpha k_i + \beta \ell_i + \varepsilon_i \qquad i = 1, \dots, n$$

where $y_i = y_i^* - \bar{y}^*$, $k_i = k_i^* - \bar{k}^*$, $\ell_i = \ell_i^* - \bar{\ell}^*$, $y_i^* = \ln(Y_i)$, $k_i^* = \ln(K_i)$, $\ell_i^* = \ln(L_i)$, and $\varepsilon_i = \ln(u_i)$.

You are given the following data:

$$n = 303; \qquad \sum_{i=1}^{n} y_i^2 = 50.4;$$

$$\sum_{i=1}^{n} k_i^2 = 15; \qquad \sum_{i=1}^{n} k_i \ell_i = 10; \qquad \sum_{i=1}^{n} l_i^2 = 20;$$

$$\sum_{i=1}^{n} y_i k_i = 12; \qquad \sum_{i=1}^{n} y_i \ell_i = 20.$$

(a) Calculate the OLS estimates for α and β.

(b) Assume that the vector $(\varepsilon_1, \ldots, \varepsilon_n)'$ has a normal distribution with zero mean and variance-covariance matrix equal to $\sigma^2 I$. Using a 5% significance level test that $\alpha = 0.5$ against $\alpha \neq 0.5$.

Problem 8. (20 points)

A certain economic variable, y is measured daily. In order to describe the change in y over time, an economist runs the regressions

$$y = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 \mathrm{SA} + \beta_5 \mathrm{SU} + \beta_6 \mathrm{MO} + \varepsilon \quad (1)$$
$$y = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_5 \mathrm{SU} + \beta_6 \mathrm{MO} + \varepsilon \quad (2)$$
$$y = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_5 \mathrm{SU} + \varepsilon \quad (3)$$

where t is a time trend, and SA, SU and MO are dummy variables taking the value 1 on Saturdays, Sundays and Mondays, respectively (and 0 otherwise). The output from the three regressions is attached (at the end of the problem)

(a) In the model (1), test the three hypotheses

i. $\beta_4 = 0$

ii. $\beta_5 = 0$

iii. $\beta_6 = 0$

(b) In the model (1), test the (joint) hypothesis

$$\beta_4 = 0 \text{ and } \beta_6 = 0$$

(c) In equation (2), explain in detail how you would test for seventh order serial correlation.

(d) In equation (2), explain in detail how you would test whether the variance of the error–term is different on Sundays.

Equation 1.

Included observations: 103

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.539775	0.002251	2461.335	0.0000
T	-0.012399	0.000389	-31.87205	0.0000
T2	0.000262	1.51E-05	17.33858	0.0000
SU	0.003772	0.001142	3.302199	0.0013

R-squared	0.983111	Mean dependent var	5.436379
Adjusted R-squared	0.982599	S.D. dependent var	0.030936
S.E. of regression	0.004081	Akaike info criterion	-10.96479
Sum squared resid	0.001649	Schwarz criterion	-10.86247
Log likelihood	422.5358	F-statistic	1920.877
Durbin-Watson stat	1.816276	Prob(F-statistic)	0.000000

Equation 2.

Included observations: 103

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.539775	0.002251	2461.335	0.0000
T	-0.012399	0.000389	-31.87205	0.0000
T2	0.000262	1.51E-05	17.33858	0.0000
SU	0.003772	0.001142	3.302199	0.0013

R-squared	0.983111	Mean dependent var	5.436379
Adjusted R-squared	0.982599	S.D. dependent var	0.030936
S.E. of regression	0.004081	Akaike info criterion	-10.96479
Sum squared resid	0.001649	Schwarz criterion	-10.86247
Log likelihood	422.5358	F-statistic	1920.877
Durbin-Watson stat	1.816276	Prob(F-statistic)	0.000000

Equation 3.

Included observations: 103

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.539775	0.002251	2461.335	0.0000
T	-0.012399	0.000389	-31.87205	0.0000
T2	0.000262	1.51E-05	17.33858	0.0000
SU	0.003772	0.001142	3.302199	0.0013

R-squared	0.983111	Mean dependent var	5.436379
Adjusted R-squared	0.982599	S.D. dependent var	0.030936
S.E. of regression	0.004081	Akaike info criterion	-10.96479
Sum squared resid	0.001649	Schwarz criterion	-10.86247
Log likelihood	422.5358	F-statistic	1920.877
Durbin-Watson stat	1.816276	Prob(F-statistic)	0.000000