Section A

Answer ALL SEVEN questions in section A.

1. Each part of the following is either a true or false statement. If you think Part n is True, your answer should be: "(n) T" and if you think it is false, "(n) F".

   Consider the linear model with k regressors, including a constant

   $$\boldsymbol{y} = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

   where it is known that plim$\frac{1}{n}\boldsymbol{X}'\varepsilon \neq \mathbf{0}$ (n denotes sample size). Suppose the columns of the matrix $\mathbf{Z}$ contain $m > k$ instruments, including a constant, such that plim$\frac{1}{n}\mathbf{Z}'\varepsilon = \mathbf{0}$, plim$\frac{1}{n}\mathbf{Z}'\mathbf{X} = Q_{\mathbf{ZX}}$, a matrix of rank k, and plim$\frac{1}{n}\mathbf{Z}'\mathbf{Z} = Q_{\mathbf{ZZ}}$, a matrix of rank m. Let $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{y}}$ denote the fitted values from regressing the columns of $\boldsymbol{X}$ and $\boldsymbol{y}$ respectively on $\boldsymbol{Z}$. Which of the following will give consistent estimates of $\boldsymbol{\beta}$?

   (a) $(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\boldsymbol{y}$. (1 mark)

   (b) $(\mathbf{X}'\hat{\mathbf{X}})^{-1}\mathbf{X}'\hat{\boldsymbol{y}}$. (1 mark)

   (c) $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$. (1 mark)

   (d) $(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\boldsymbol{y}$. (1 mark)

   (e) $(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$. (1 mark)

2. Each part of the following is either a true or false statement. If you think Part n is True, your answer should be: "(n) T" and if you think it is false, "(n) F".

   A researcher estimates the following dynamic regression using annual data, where y_t is the log demand for a certain good and x_t is its log price, which we take to be exogenous.

   $$\begin{array}{rcccccccccc} y_t & = & 117.7 & + & 0.373y_{t-1} & - & 0.156x_t & - & 0.062x_{t-1} & + & e_t. \\ & & (50.284) & & (0.177) & & (0.019) & & (0.029) & & \\ n & = & 46 & & & & & & & & \end{array}$$

   (Standard errors in parentheses.)

   (a) A t-test that the variable x_t should be omitted from the regression would be rejected at the 5% level. (1 mark)

   (b) If x_t were to fall by one unit we would expect y_t to fall by 0.156. (1 mark)

   (c) A small coefficient for y_{t-1} means that it takes longer for the system to adjust to a new equilibrium value. (1 mark)

(d) The estimated long run elasticity of demand is approximately 0.218. (1 mark)

(e) OLS estimates of this regression are inconsistent if the disturbance term is a first order moving average process. (1 mark)

3. Suppose we are trying to estimate the linear relationship $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\epsilon_i$, where $\epsilon_i \sim N(0,1)$. Our sample has been truncated from below, so that it contains only individuals who report a positive value for the dependent variable, $y_i > 0$. Explain, using diagrams or otherwise, why ordinary least squares estimates of $\boldsymbol{\beta}$ will not be consistent. Outline a method that would produce consistent results. (6 marks)

4. In order to begin to model a quarterly univariate time series, the sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF) are calculated. The following results are obtained.

Lag	SACF	SPACF
1	0.832	0.832
2	0.694	-0.066
3	0.519	-0.019
4	0.372	-0.083
5	0.236	0.017
6	0.176	-0.134
7	0.112	-0.104
8	0.095	-0.030

Briefly explain how the sample partial autocorrelation function (SPACF) is calculated and discuss what type of time series process would fit the table. (6 marks)

5. A researcher is interested in the relationship between the elements of an $m \times 1$ vector, $\mathbf{Y}_t$, using the equivalent expressions

$$\begin{aligned}\mathbf{Y}_t &= \boldsymbol{\alpha} + \boldsymbol{\Phi}_1\mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{Y}_{t-2} + \boldsymbol{\epsilon}_t,\\ \Delta\mathbf{Y}_t &= \boldsymbol{\Pi}(\mathbf{Y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t,\ t = 3, \ldots, n,\end{aligned}$$

where $\Delta\mathbf{Y}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$ and $\boldsymbol{\epsilon}_t \sim IID(\mathbf{0}, \boldsymbol{\Omega})$. Show how the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}_1$ relate to the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$. (6 marks)

6. Define what it means for a time series process, y_t, to be stationary. Under what conditions is an autoregressive moving average (ARMA) process stationary? Show that the following ARMA (2,1) process is stationary: $y_t = y_{t-1} - 0.21y_{t-2} + \varepsilon_t + 0.7\varepsilon_{t-1}$, $t \in (-\infty, +\infty)$ where ε_t is a white noise process with variance σ^2. (6 marks)

TURN OVER

7. Suppose that an econometrician is interested in the relationship between two variables, y_{1t} and y_{2t} using a sample $t = 1, \ldots, n$. They want to estimate the following simultaneous regressions model

$$
\begin{aligned}
y_{1t} &= \gamma_{12} y_{2t} + \beta_1 z_{1t} + \epsilon_{1t}, & (1) \\
y_{2t} &= \gamma_{21} y_{1t} + \beta_2 z_{2t} + \epsilon_{2t}, & (2)
\end{aligned}
$$

where z_{1t} and z_{2t} are exogenous explanatory variables, $\gamma_{12}, \gamma_{21} \neq 0$ and $\boldsymbol{\epsilon} = (\epsilon_{1t}, \epsilon_{2t})'$ contains unobserved disturbances with E$(\boldsymbol{\epsilon}) = \mathbf{0}$, Var$(\boldsymbol{\epsilon}) = \boldsymbol{\Omega}$. Explain why ordinary least squares estimates of Equation (1) and Equation (2) will not be consistent. Propose a method to obtain consistent estimates of γ_{12} and β_1 in Equation (1), taking care to state any conditions and assumptions required. (6 marks)

Section B

Answer ALL THREE questions in section B. To get good marks your answer should contain clear statements that explain the logical structure of your argument.

8. Consider the time series process

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t,$$

where $|\phi| < 1$ and ε_t follows an ARCH(1) process with $\varepsilon_t | Y_{t-1} \sim N(0, \sigma_t^2)$, where $Y_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$ denotes the past history of the process and

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2,$$

with $\alpha_0 > 0$ and $\alpha_1 \geq 0$.

(a) Using the fact that we may write $\varepsilon_t = \eta_t \sigma_t$, where $\eta_t \sim N(0, 1)$ is a sequence of independent standard normal variables, show that ε_t is a white noise process with mean zero and (unconditional) variance $\alpha_0/(1 - \alpha_1)$ as long as ε_t is stationary. (7 marks)

(b) Show that ε_t^2 follows an AR(1) process. Hence or otherwise discuss the conditions under which ε_t is stationary. (7 marks)

(c) Suppose the parameters $\alpha, \phi, \alpha_0, \alpha_1$ are known. Write down the one-step ahead forecast, $\hat{y}_{n+1} = \mathsf{E}\left[y_{n+1} | y_n, y_{n-1}, \ldots\right]$. Derive the variance of the forecast error. (6 marks)

TURN OVER

9. Consider the salary data of a sample of 258 male bank employees. The dependent variable MANAGERIAL takes the value 1 if worker i is in a managerial role, and 0 otherwise. The regressors EDUC measures the number of completed years of education, MINORITY is a dummy taking the value 1 if the employee is a member of an ethnic minority and PREVEXP measures previous work experience, in months. A researcher employs a logit model, $P(y_i = 1) = \Lambda(\mathbf{x}_i'\boldsymbol{\beta})$, where $\mathbf{x}_i$ contains the regressors for worker i and $\Lambda(t) = \frac{1}{1+e^{-t}}$ the logistic function.

   (a) As a first attempt, the researcher estimates the following model.

Table 9.1. Dependent Variable: MANAGERIAL					
Method: ML - Binary Logit					
Variable	Coefficient	std error	Prob.		
C	-26.95253	4.400895	0.0000		
EDUC	1.674803	0.280049	0.0000		
MINORITY	-2.395242	0.847981	0.0047		
PREVEXP	0.003865	0.003078	0.2092		
Mean dependent var		0.286822	Akaike info criterion		0.522420
Log likelihood		-63.39219	Avg. log likelihood		-0.245706

   Comment briefly on the signs and significance of the coefficients in this model. Are they in line with your expectations? (4 marks)

   (b) The researcher re-estimates the model, omitting the variable PREVEXP.

Table 9.2. Dependent Variable: MANAGERIAL					
Method: ML - Binary Logit					
Variable	Coefficient	std error	Prob.		
C	-26.21472	4.311652	0.0000		
EDUC	1.644798	0.276714	0.0000		
MINORITY	-2.119683	0.793999	0.0076		
Mean dependent var		0.286822	Akaike info criterion		0.520543
Log likelihood		-64.15011	Avg. log likelihood		-0.248644

   Using all relevant information from the tables, discuss which of the two models you prefer. Use the second model to estimate the probability that a male employee not from an ethinic minority with 15 years of education is in a managerial position. What might explain the increase in the estimated coefficient MINORITY when the variable PREVEXP is excluded from the regression?

   (10 marks)

   (c) According to the regression in part 9b, the probability that a male employee has a management job depends on his time in education. Explain carefully why that probability does not increase by 1.645 for every additional year of education. Calculate an estimate of the marginal effect of an extra year of education for an ethnic minority male with 17 years of education. (6 marks)

10. A researcher is using a sample of 30 observations taken annually from 1970 to 1999 to examine the relationship between gasoline consumption (GC), gasoline price (PG) and disposable income (RI), all recorded in real terms and then taken in logarithms.

    (a) Using GC as the dependent variable, they obtain the following results using ordinary least squares.

Table 10.1. Dependent Variable: GC
Method: Least Squares

Variable	Coefficient	t-statistic	Prob.
Constant	4.9860	61.4791	0.0000
PG	-0.5276	-20.0457	0.0000
RI	0.5732	23.3864	0.0000
R-squared	0.9872	Adjusted R-squared	0.9862
F-statistic	1037.467	Prob(F-statistic)	0.0000
Log likelihood	71.0958	Durbin Watson	1.1126
Akaike info criterion	-4.5397	Schwarz criterion	-4.3996

Interpret the meaning of the coefficients in this regression in relation to economic theory. Do they have the signs you would expect? (5 marks)

    (b) Briefly explain the problem of spurious regressions using time series data. What evidence is there in the above table to indicate that this regression may be spurious? How could the researcher test this claim? (5 marks)

    (c) The researcher finally estimates the following autoregressive distributed lag model.

Table 10.2. Dependent Variable: GC
Method: Least Squares

Variable	Coefficient	t-statistic	Prob.
Constant	2.7905	3.1086	0.0049
GC(-1)	0.4323	2.4289	0.0234
PG	-0.4159	-7.7150	0.0000
PG(-1)	0.1083	1.0188	0.3189
RI	0.9167	3.5360	0.0018
RI(-1)	-0.5814	-2.1257	0.0445
R-squared	0.9915	Adjusted R-squared	0.9896
F-statistic	535.7889	Prob(F-statistic)	0.0000
Log likelihood	74.7485	Durbin Watson	2.0313
Akaike info criterion	-4.7413	Schwarz criterion	-4.4584

Rewrite this model in error correction form and interpret the results. Compute the long run elasticities of price and income. Discuss the evidence that this model should be preferred to the model in part (a). (10 marks)

[redacted]