

OLLSCOIL NA hEIREANN, GAILLIMH
THE NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II EXAMINATIONS 2004/2005
M.A./M.Econ.Sc. Economic Policy and Planning

EC506 Econometrics

Professor Vincent Munley
Professor M.P. Cuddy
Dr. Cathal O'Donoghue

5 Questions, Attempt 3.

TIME ALLOWED: 3 HOURS

1. Applied Econometrics

- (i) Describe the modelling steps you would use to test the hypothesis that an individual's potential wage rate influences their decision (a) to participate in the labour market and (b) the amount of hours they work. Justify the data and methods you would use.
- (ii) Describe the difference between censoring and truncation. How do these issues affect your regression estimates? How would you correct for them?

2. Classical Assumptions

- (i) List the Assumptions underlying the disturbance term of the Simple Linear Regression Model.
- (ii) What do we mean by autocorrelation? What does it arise from? How does auto correlation affect our parameter estimates?

- (iii) The Durbin Watson statistic is defined as
$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2},$$

where e_j represents the residual for observation j . Derive an expression for the statistic as a function of $\hat{\rho}$, an estimator of the autocorrelation coefficient ρ . If we have 7 explanatory variables excluding the constant term and 40 observations, the Durbin Watson Statistic produces values for d_L and d_U as 1.120 and 1.924 respectively. How do we interpret the following values of the DW statistic:

- a. 1.3
- b. 0.95
- c. 3.9
- d. 2.06
- e. 2.16
- f. 2.8

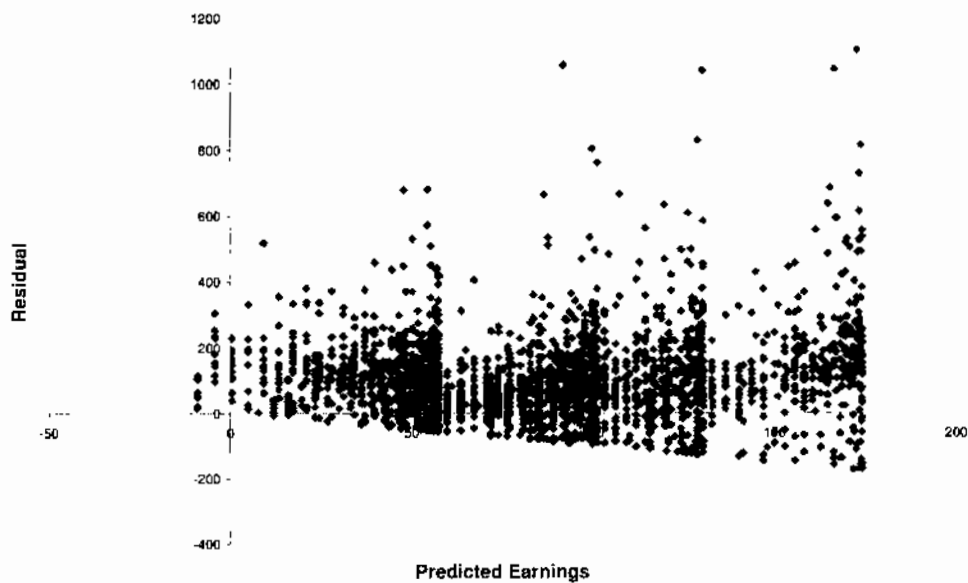
If one of the explanatory variables was a lagged dependent variable, should we use the existing method to test for autocorrelation? Discuss your answer.

- (iv) The following plot (figure 1) of the residual against predicted earnings has been generated by the econometric model (model 1) of earnings against years of schooling, years of experience and years of experience². Which assumption does the disturbance term fail? What is the effect on the parameter estimates of this assumption failure? Describe a statistical test could one use in addition to a visual test? Suggest a solution to this problem.

Model 1.

Earnings	Coefficient	Std. Err.	T	P>t	[95% Conf. Interval]
Years of Education	19.59923	0.795266	24.64	0	18.03991 21.15855
Years of Experience	12.32431	0.526292	23.42	0	11.29238 13.35624
Years of Experience ²	-0.19268	0.011438	-16.85	0	-0.2151 -0.17025
Constant	-163.576	11.3058	-14.47	0	-185.744 -141.408

Figure 1



3. Endogeneity

- (i) What do we mean by an endogenous variable? Describe mathematically how it affects our regression model estimates and why?
- (ii) Describe what we mean by measurement error. In a simple linear regression model, where measurement errors are normally distributed with zero mean and constant variances and are uncorrelated with each other, Show that the regression estimates are biased.
- (iii) What is omitted variables bias? Where Y is log earnings and we have a model
$$Y = b_0 + b_1 \text{experience} + b_2 \text{experience}^2 + b_3 \text{education} + v,$$
comment on why we have omitted variables bias. Where education is assumed to be endogenous, describe mathematically the direction of the bias. What factors influence this bias?
- (iv) What is an instrument and instrumental variable estimation? In the model in (iii), why is mother's education likely to be a poor instrument for education? Justify how Angrist and Krueger were able to use the quarter of birth as an instrument for education in the USA.

4. Maximum Likelihood Estimation

- (i) What do we mean by maximum likelihood estimation?
- (ii) For the regression model $y_i = x_i'\beta + \varepsilon_i$, derive an expression for the log-likelihood function in terms of the density function $f(\varepsilon/\sigma^2)$. Justify your derivation.
- (iii) Is MLE for a multiple regression model more efficient than OLS? Justify your answer. What extra assumptions are made?
- (iv) Where the following describes a latent relationship underlying the sequential decisions to (a) work (y_0) and conditional on working to (b) work full-time (y_1):

$$y_0^* = x\beta_0 + u_0$$

$$y_1^* = z\beta_1 + u_1$$

and

$$y_0 = 1[y_0^* > 0]$$

$$y_1 = 1[y_1^* > 0]$$

and where the disturbance terms, $\{u_0, u_1\}$ are jointly distributed with a bivariate normal distribution with correlation ρ , $\Phi_2(x\beta_0, z\beta_1, \rho)$, derive the contribution to the log-likelihood for each sample case.

5. Panel Data Models

- (i) Describe what we mean by a Panel Data Model. What advantages and problems do they have?
- (ii) Consider a dataset where n (large) individuals are interviewed over m periods. Discuss the nature of the endogeneity bias if we use a pooled OLS estimator in a regression model.
- (iii) Describe the difference between a random effects model and a fixed effects model. When should we use one and not the other? What assumptions are required to use the random effects model? Describe a method used to test for which method to use in a panel data model. Describe a method to test whether the random effects are significantly different from zero.