

Comprehensive Exam in Econometrics
Suffolk University
August, 2005

Answer each of the following four questions. You have three hours to take this test.

1. Consider $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$, where \mathbf{X} represents a non-stochastic $T \times k$ data matrix, including a constant.
 - a) State the assumptions of the classical linear regression model. Use these assumptions to show that $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\sigma}^2 = e'e/T - k$ are unbiased estimators for the regression coefficients (β) and the error term variance (σ^2).
 - b) Briefly explain the principle of maximum likelihood (ML) estimation. Given the above assumptions along with the normality of the error term $\boldsymbol{\varepsilon}$, derive the maximum likelihood estimators of β_{MLE} and σ_{MLE}^2 .
 - c) Although economic theory provides some information on the choice of \mathbf{y} and \mathbf{X} , it is generally silent on its functional form. Briefly compare a linear model with an analogous log-linear model. How do you choose between the two?

2. Consider the classical linear regression model, $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$, where \mathbf{X} is non-stochastic. Let $E(\boldsymbol{\varepsilon}) = 0$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\Omega$ where Ω is symmetric positive definite matrix.
 - a) If Ω is known, derive the GLS estimates and briefly explain why they are more efficient than the OLS estimates
 - b) What is serial correlation? Briefly discuss the implications of ignoring serial correlation. Consider $y_t = x_t'\beta + \varepsilon_t$. Further, $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ where $-1 < \rho < 1$ and v is i.i.d. Explain the Durbin-Watson and the LM tests for the 1st order serial correlation and the estimated GLS method for correcting it.
 - c) Consider $y_t = \beta_1 + \beta_2x_{2t} + \beta_3x_{3t} + \gamma y_{t-1} + \varepsilon_t$ where $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$. Why is Durbin-Watson not valid anymore? Describe the Durbin's h and the LM tests for testing for serial correlation in this dynamic model. How would you estimate the parameters efficiently if the 1st order serial correlation is detected?

3. The following panel data model is used to estimate the effect of foreign direct investment on output across a sample of N companies over a period of T years:

$$\log(Y_{it}) = \beta_1 FDI_{it} + \gamma z_{it} + \eta_t + c_i + \varepsilon_{it},$$

where:

Y_{it} is output of firm i in year t ,
 FDI_{it} is a share of foreign equity,
 z_{it} is a vector of production inputs,
 η_t are time effects,
 c_i are unobserved firm effects,
and ε_{it} are idiosyncratic errors.

- a) Would you expect the OLS estimator of β_1 to be accurate? Justify your answer.
- b) State the assumptions and explain the role of demeaning in the fixed effects and random effects estimation approaches.
- c) How would you estimate this model? Explain your choice of the estimation method. Derive your chosen estimator and its variance.
- d) Suppose the number of years varied across the firms in your data. How would this change your estimation procedure? Give an explanation.
- e) How would your estimation procedure change if the model included a lag of the dependent variable? Give an explanation.

4. The following model jointly determines child support payments and visitation rights for divorced couples with children (assuming children live with their mothers and fathers pay child support):

$$\text{fathers' equation:} \quad \text{support} = \beta_{11} + \gamma_{21} \text{visits} + \beta_{21} \text{fincome} + \beta_{31} \text{fremarr} + \beta_{41} \text{distance} + \varepsilon_1,$$

$$\text{mothers' equation:} \quad \text{visits} = \beta_{12} + \gamma_{12} \text{support} + \beta_{22} \text{mremarr} + \beta_{32} \text{distance} + \varepsilon_2,$$

where exogenous variables are: *fincome* (father's income), *fremarr/mremarr* (dummies for whether the father/mother is remarried), and *distance* (miles currently between the father and mother).

- a) What are the conditions required for identification? Are these equations identified?
- b) Evaluate advantages and disadvantages of single-equation versus system estimation methods. Under what conditions does the GMM estimator reduce to 2SLS or 3SLS?
- c) How would you estimate this system? Justify your choice of the estimation method. Derive your chosen estimator and its variance.
- d) Explain how you would test the (i) strength and (ii) validity of instruments in this model.
- e) How would your answer to part (i) change if *fincome* did not appear in the first equation? How would this simplify your estimation procedure? Justify your answer.