School of Economics

College of Business and Economics

FINAL ASSESSMENT: NOVEMBER 2021

Course	:	Applied Econometric Techniques (AET9X01)
Moderator	:	Dr A Pholo
Date	:	11 November 2021
Time	:	180 minutes
Marks	:	100 points

Question 1 (15)

Describe succinctly the estimation procedure implied by the control function approach of endogeneity.

Question 2 (15)

Describe succinctly the estimation procedure implied by the correlated random coefficient model.

Question 3 (15)

Use the data in *LOWBIRTH.DTA* for this question

1. For 1987 and 1990, consider the state-level equation

$$\begin{aligned}\text{lowbrth}_{it} &= \theta_1 + \theta_2 \text{d90}_t + \beta_1 \text{afdcprc}_{it} + \beta_2 \log\left(\text{phypc}_{it}\right) \\ &\quad + \beta_3 \log\left(\text{bedspc}_{it}\right) + \beta_4 \log\left(\text{pcinc}_{it}\right) + \beta_5 \log\left(\text{popul}_{it}\right) + c_i + u_{it}\end{aligned}$$

   where the dependent variable is percentage of births that are classified as low birth weight and the key explanatory variable is afdcprc, the percentage of the population in the welfare program, Aid to Families with Dependent Children (AFDC). The other variables, which act as controls for quality of health care and income levels, are physicians per capita, hospitals beds per capita, per capita income, and population. Interpreting the equation causally, what sign should each β_j have? (Note: Participation in AFDC makes poor women eligible for nutritional programs and prenatal care.)

2. Estimate the preceding equation by pooled OLS, and discuss the results. You should report the usual standard errors and serial correlation-robust standard errors.

3. Difference the equation to eliminate the state fixed effects, c_i, and re-estimate the equation. Interpret the estimate of β_1 and compare it to the estimate from part 2. What do you make of $\hat{\beta}_2$?

4. Estimate the model with the within group method? Compare the results with the first differencing estimates?

Question 4 (15)

Describe succinctly the Chamberlain's approach to UE linear panel data model. Explain how this model may be used to test what is the most appropriate model between the RE and the FE models.

Question 5 (20)

The data set AIRFARE.RAW contains information on passengers, airfare, and route concentration ratios for 1,149 routes within the United States for the years 1997 through 2000. We estimate a simple demand model

$$\log(passen_{it}) = \theta_{t1} + \alpha_1 \log(fare_{it}) + \delta_1 \log(dist_i) + \delta_2 [\log(dist_i)]^2 + c_{i1} + u_{it1}, \tag{1}$$

where we allow for separate year intercepts. The variable $dist_i$ is the route distance, in miles; naturally, it does not change over time.

1. Estimate the reduced forms underlying the REIV and FEIV analyses. Using fully robust t statistics, is *concen* sufficiently (partially) correlated with $\log(fare)$?

2. Redo the REIV estimation, but drop the route distance variables. What happens to the estimated elasticity of passenger demand with respect to fare?

3. Now consider a model where the elasticity can depend on route distance:

$$\begin{aligned}\log(passen_{it}) &= \theta_{t1} + \alpha_1 \log(fare_{it}) + \delta_1 \log(dist_i) + \delta_2 [\log(dist_i)]^2 + \gamma_1 [\log(dist_i) - \mu_1] \log(fare_{it}) \\ &+ \gamma_2 \left[(\log(dist_i))^2 - \mu_2\right] \log(fare_{it}) + c_{i1} + u_{it1},\end{aligned} \tag{2}$$

where $\mu_1 = E(\log(dist_i))$ and $\mu_2 = E[\log(dist_i)]^2$.The means are subtracted before forming the interactions so that 1 is the average partial effect. In using REIV or FEIV to estimate this model, what should be the IVs for the interaction terms?

4. Use the data in AIRFARE.RAW to estimate the model in part 3, replacing μ_1 and μ_2 with their sample averages.

5. Obtain fully robust standard errors for the FEIV estimation, and obtain a fully robust test of joint significance of the interaction terms. (Ignore the estimation of μ_1 and μ_2.) What is the robust 95 percent confidence interval for α_1?

Question 6 (20)

Use information in Stata file Application 7 and estimate a Cobb-Douglas production function. After log transformation we have

$$\begin{aligned} y_{it} &= \beta_n n_{it} + \beta_k k_{it} + \gamma_t + (\eta_i + v_{it} + m_{it}) \\ v_{it} &= \alpha v_{i,t-1} + e_{it} \qquad |\alpha| < 1 \\ e_{it}, m_{it} &= \text{MA}\,(0) \end{aligned}$$

where y_{it} is log sales of firm i in year t, n_{it} is log employment, k_{it} is log capital stock and t is a year-specific intercept reflecting, for example, a common technology shock. η_i is an unobserved time-invariant firm-specific effect, v_{it} is a possibly autoregressive (productivity) shock and m_{it} reflects serially uncorrelated (measurement) errors.

Blundell and Bond (2000) consider the time series properties of these series and report estimates of this production function using a balanced panel of 509 R&D-performing US manufacturing companies observed for 8 years, 1982-89, similar to that used in Mairesse and Hall (1996).

Capital stock and employment are measured at the end of the firm's accounting year, and sales is used as a proxy for output.

1. Prove that the above model has the following dynamic representation

$$y_{it} = \pi_1 n_{it} + \pi_2 n_{i,t-1} + \pi_3 k_{it} + \pi_4 k_{i,t-1} + \pi_5 y_{i,t-1} + \gamma_t^* + (\eta_i^* + w_{it})$$

   Run OLS regression and interpret your results using time trend and a correct model that account for possible heteroscedasticity and cross-section errors correlation (4).

2. Run Within-Group regression and interpret your results using time trend and a correct model that account for possible heteroscedasticity and cross-section correlation (4).

3. Compare the results obtained using OLS to those of WG (4).

4. Run a GMM model and compare the results to those of OLS and WG (4).

5. Assess the validity of instruments (4).