University of Gothenburg
Department of Economics, Econometrics III (PhD)
Static and Dynamic Linear Panel Data Models: Empirical Exercise

Yonas Alem*

April 27, 2017

A panel data set contains repeated observations on the same economic units (individuals, households, firms, countries etc.) collected over a number of periods. We learned that panel data are larger than cross-sectional and time-series data with variation over two dimensions (individuals and time) and thus, yield estimators that are more accurate. In panel data models, we are able to control for time-invariant unobserved individual heterogeneity and reduce the problem of omission variables bias. Panel data also allow us to estimate dynamic models - models with a lagged dependent variable - that are of interest in a wide range of economic applications. In this empirical exercise, we will practice estimating a production function using alternative static and dynamic panel data estimators. This empirical exercise draws on the following two references.

- Bond S. 2002. Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. *Portuguese Economic Journal* 1, 141-162. An identical working paper version available at: http://www.cemmap.ac.uk/wps/cwp0209.pdf.

- Blundell, R.W. and Bond, S.R. 2000. GMM Estimation with Persistent Panel Data: an Application to Production Functions. *Econometric Reviews* 19, 321-340.

*Department of Economics, University of Gothenburg; e-mail: yonas.alem@economicss.gu.se, homepage: www.yonasalem.net.

Consider the following Cobb-Douglas production function:

$$y_{it} = \beta_n n_{it} + \beta_k k_{it} + \gamma_t + (\eta_i + \upsilon_{it} + m_{it}) \tag{1}$$

where y, n, k denote output, labour and capital (in logs), respectively. γ_t is a year-specific intercept reflecting, for example, a common technology shock. The term η_i controls for a firm fixed effect which is interpretable as time-invariant total factor productivity, υ_{it} is a possibly autoregressive (productivity) shock, and m_{it} is captures a serially uncorrelated measurement errors.

Using a balanced panel dataset that covers 509 R&D-performing US manufacturing companies observed for 8 years, 1982-89, we shall consider results from four static linear panel data estimators that we discussed in class: Pooled OLS (POLS), Random Effects (RE), Fixed Effects (FE), and First Differences (FD), and two dynamic linear estimators (Difference GMM and System GMM). The dataset has been used previously by Blundell and Bond (2000); the file is called "usbal89".

You can form groups people and collaborate in the group to compare and discuss your solutions for this exercise. Nevertheless, you are each required to submit your own lab report. Please e-mail your reports to **yonas.alem@economics.gu.se** on Monday, **May 8, 2017** at 17:00 at the latest. Late reports will not be accepted.

1. Estimate the parameters in (1) using POLS. Decide on the preferred way of estimating the standard errors and compare these to the standard errors based on the conventional formula. Explain why the two variance estimators differ. Test for constant returns to scale using the preferred variance estimator and interpret the time trend. Briefly explain why the POLS is not suitable to estimate (1).

2. Next, estimate (1) using the random effects (RE) approach. Do the results differ in any interesting way compared to those obtained in (1)? If so, why? Carry out the Breusch-Pagan test for random effects and interpret the result. Comment on the validity of RE to estimate (1).

3. Fixed effects next: i.e., estimate (1) using the "within" (FE) estimator. Check whether alternative methods for computing the variance matrix give similar standard

errors; decide on the preferred method (motivate). Test for constant returns to scale. Are the key assumptions of the FE likely to hold in estimating (1)?.

4. Which of the estimators used in (1), (2), (3) do you think is the preferred one and why? Discuss any remaining econometric problems that have not been addressed by any these estimators.

5. Estimate (1) in first differences (FD) and comment on the results.

6. Suppose the capital stock variable is measured with error. How might this affect the parameter estimates obtained above?

7. Return to the production function setting now, and assume that v_{it} is serially correlated:

$$v_{it} = \alpha v_{i,t-1} + e_{it} \qquad |\alpha| < 0, \qquad e_{it} \sim MA(0) \tag{2}$$

Suppose both η_i and v_{it} are assumed observed to the manager and unobserved to the econometrician. We can therefore expect the factor inputs (capital and labour) to be correlated with both η_i and v_{it}, simply because these unobserved productivity terms affect the marginal product of capital and labour, and so determine factor demand. We also expect measurement errors m_{it} to be correlated with the factor inputs.

Blundell and Bond (2000) propose the following equation as a basis for estimation of the production function:

$$y_{it} = \pi_1 n_{it} + \pi_2 n_{i,t-1} + \pi_3 k_{it} + \pi_4 k_{i,t-1} + \pi_5 y_{i,t-1} + \gamma_t^* + (\eta_i^* + w_{it}) \tag{3}$$

see eq. (2.3) in their paper or eq. (2.3) in Bond (2002).

Explain how (3) is related to (1) and (2). Write down the non-linear (common factor) restrictions in (3) implied by (1) and (2).

8. Blundell and Bond report results for various estimators that control for firm fixed effects. In some of their regressions, they also address the potential problem that capital and labour may be correlated with the time varying part of total factor productivity, by using lags of capital and labour as instruments. Why are lagged factor inputs more convincing instruments if we are estimating (3), compared to (1)?

9. Take first differences of (3) in order to remove the firm fixed effect. Do you expect output, capital and labour lagged 2 periods (t-2) to be valid and informative instruments, for the first-differenced equation? What are the advantages and disadvantages of using these variables dated t-3 as instruments instead?

10. Replicate the dynamic production function specifications in Table 4 of Bond (2002). Interpret the results. Comment on the test results for Sargan's overidentification tests and the tests for serial-correlation in the error terms. What do the results indicate with regards to the returns to scale?

11. Which regression results are most preferred for the production function specified in (3)? Why?