# Exercise Chapter 9

The firm based data which is used in this exercise and also for the exercises in Chapters 10 and 24 can be found in the Stata data file ‘Ghana\_Firms\_JDE04’. This is the data used in Söderbom and Teal (2004). There is a longer panel of these firms available at:

http://www.csae.ox.ac.uk/datasets/ghana-rped/Ghmain.html.

1. Explain clearly how a panel data set differs from pooled cross sections of independently sampled observations.

The key point is that in a panel the same units whether they be individuals, households or firms are surveyed over time. This means that if the cross section is randomly sampled it will be representative of the population. While this may be true for a panel when it starts, assuming it too was taken as a random sample, this will not, in general, be true over time. It is important not to think panel data sets are to be preferred to cross-section ones. They are complementary sources of information.

2. What are the advantages and disadvantages of having a panel compared to repeated cross sections?

As we have explained in this chapter such data allow us to control for some elements of the unobservable characteristics of the individuals, households or firms. While such data has clear advantages, it has some disadvantages. These are discussed in Angus Deaton’s The Analysis of Household Surveys (pages 105ff). One is the decrease in the precision of the estimates that comes with the decrease in sample size that comes from moving from the pooled OLS to the FE or differenced estimator. As we see below this matters a lot. Another, and Deaton argues possibly the most important, is when the regressors are measured with error. In this case it is possible that the bias in the FE or differenced estimator exceeds the bias in the cross section. As Deaton also stresses it is important to compare results across different estimators and not assume that one is automatically superior to another.

3. Why might the OLS and Fixed Effects estimates differ?

One reason why OLS and fixed effects will differ is that the regressors are correlated with the time invariant unobservables and such a correlation is biasing our point estimates of the regressors. It is for this reason that it is always a good idea to compare OLS results with those of FE, if they differ little that is information that either there are no time invariant unobservables or, rather more likely, that they are not correlated with the regressors.

However the point we have already made regarding measurement error is relevant here. It is possible OLS and FE differ greatly not due to any correlation between time invariant unobservables and the regressors but the fact that the variance of the independent variables in the FE estimates is much lower than in the cross section and that measurement error is greater. Recall that using the FE estimator ensures that only the time variation in the data is used to estimate the parameters. If there is very little change then it will be difficult to identify the effect (this point is discussed more formally in Chapter 11, Section 11.2.3). In the limit, of course, where the regressor does not change we cannot identify the effect at all from the FE estimate as the fixed regressor is perfectly collinear with the fixed effect.

4. What is the economic meaning of the fixed effect estimate in both the macro and micro production functions?

We provide an interpretation of the fixed effect in Section 9.4 and caution that in short panel these estimates are not consistent. However it is important to understand that the fixed effect is important and in both macro and micro production functions can be given an interpretation as a measure of total factor productivity (TFP).

5. Using the data from ‘Ghana\_Firms\_JDE04’ for waves 2 and 7 of the data set replicate the findings in Tables 9.5 and 9.6 and then drop the time dummy. In doing so what are you assuming about the rates of growth of TFP in the data? Does the evidence support your assumption?

This is done for you in the Stata do file

‘Exercise\_Chapter\_9.do’.

The first point to notice is that the time dummy is never significant. Indeed in the pooled cross section the point estimate is negative. What this time dummy is telling you is that the rate of growth of total factor productivity (TFP) for the firms over this period was not significantly different from zero. As you will see if you look at labour productivity this differs greatly across the firms with larger firms having a much higher level of labour productivity. However labour productivity does not change over this period.

Distinguishing between labour productivity which differs across firms and TFP which, in this specification, is common across firms is important. It is differences in labour productivity which are related to differences in wages and the cost of capital across firms. Larger firms pay higher wages and this may be due to their higher labour productivity or may be the cause of that higher productivity. Indeed causality may run both ways. TFP refers to changes over time and while in macro data analysis it is often argued that growth is dominated by changes in TFP that appears not be true in this micro firm data set. Dropping the time dummy, as you are asked to do in the question, simply imposes on the data what the regression suggest in the case namely that there is no growth in TFP over this period.