

Problem Set – Production Functions

PHD IO, Spring 2021

Boston College

Prof. Charlie Murry

Due: April 14, 2021

I posted a dataset: `firms-cleaned.dta`. This is data from a survey of firms in Ghana starting in the 1991. An observation is a **firm-year**. I describe the dataset in more detail in a dataset `readme` file.

When using the data, please use real values (real values are indexed to 1991 in the local currency) and use the value added measurement of output (variable already created), which is the value output minus the cost of materials.

Please report point estimates only, unless otherwise asked. Please provide your results in a separate file from your code, but submit your code as well.

OLS

(a) Estimate output elasticities, assuming a Cobb-Douglas functional form for the production function, using OLS.

Note: Do not estimate a constant. Instead, de-mean all of the logged variables before estimation.

(b) Now, include firm fixed effects. Given the nature of the data, do FEs seem appropriate? Why might FEs not take care of the endogeneity issue raised by Olley and Pakes and previous authors?

Olley and Pakes

Estimate the output elasticities using the methodology proposed by Olley and Pakes (1996). Why might this methodology be particularly troublesome in this context (ie, using these particular data from Ghana). *[While it is probably best to estimate the OLS and FE models in STATA/R, you should probably switch to Matlab/Julia for this and the following tasks.]*

LP

Estimate the output elasticities using the methodology proposed by Levinsohn and Petrin (2003) (ie, using *materials* as a proxy) but using the moments suggested by ACF.

ACF

Estimate the output elasticities using the methodology proposed by Levinsohn and Petrin (2003). What assumption about materials are you implicitly making? Is this assumption different than you were treating materials in the previous estimations?

Dynamic Panel

Another way to think about the endogeneity of TFP is to decompose TFP into two shocks like usual, $\omega_{it} + \epsilon_{it}$, and then put structure on ω . [*Note: ACF have a discussion of this. There is also a discussion in Victor A's book chapter.*]

One solution comes from the econometrics literature on dynamic panels: see Arrellano and Bond (1991, Restud). Suppose ω follows an AR(1) process: $\omega_{it} = \rho\omega_{i,t-1} + \xi_{it}$, with the same timing assumptions on ω , k , and ℓ as OP/LP.

We can “ ρ -difference” the production function and arrive at:

$$y_{it} = \rho y_{i,t-1} + \beta_1(k_{it} - \rho k_{i,t-1}) + \beta_2(\ell_{it} - \rho \ell_{i,t-1}) + \xi_{it} + (\epsilon_{it} - \rho \epsilon_{i,t-1})$$

This is exactly the “lagged dependent variable” form of Arrellano and Bond (1991), so we can use their identification results – essentially constructing moments using lagged values (you probably saw this in cross-section, and there is a STATA package for this). [*Note: there is a large literature discussing how well these lagged dependent variable models are identified.*]

Estimate the output elasticities applying the dynamic panel assumptions and using the following moment conditions:

$$\frac{1}{N} \frac{1}{T} \sum_n \sum_t \left(u_{it} \otimes \begin{pmatrix} 1 \\ k_t \\ k_{t-1} \\ \ell_{t-1} \end{pmatrix} \right),$$

where $u_{it} = \xi_{it} + (\epsilon_{it} - \rho \epsilon_{i,t-1})$, i.e. the exogenous econometric error. Explain why those are the moments we should use. Compare the objective function using the dynamic panel method to ACF.

Markups

Plot the average markup as a function of time. Regress markups on export status. Given the procedure you have done so far, why might the results from

regressing markups on export status be suspect?

Selection

What bias might be caused by selection out of the sample? Show me a picture that suggests that selection is or isn't a problem.