HOME ASSIGNMENT

Empirical Industrial Organization — ECONM0013

University of Bristol, Academic Year 2024-2025

This home assignment is made of several questions, which must all be answered (no optional questions). The points associated to each question are reported below in square brackets. The first part of the assignment is a continuation of problem sets 3 and 4 and relates to the methods discussed in topic 2: the estimation of production functions. Perform all the statistical analyses in this first part of the assignment using STATA. The second part of the assignment asks you to discuss an article not covered in class but included in the reading list.

In addition to writing your answers to all questions, you must provide a copy of the STATA codes used to produce the results in the same document (only submit one computer-typed document including both answers and codes). Each step of your STATA procedure should be well explained and your answers should rely on the use of mathematical symbols and derivations (as we have been doing in the course): whenever it is unclear where your results or conclusions come from, no points will be assigned. The document with your answers and codes should be typed with the computer (with Microsoft Word or any TeX editor), including mathematical formulae and tables (no hand writing or drawing).

You must work entirely on your own. The deadline to return the document with your answers and the STATA codes is Tuesday 15/04/2025 at 11:00 (Bristol time).

1 Estimation of Production Functions [80 Points]

Load the data "HA_Data.dta" in STATA. This is a similar dataset to the one used in problem sets 3 and 4. There are 8 variables in the dataset: $firm$ (firm ID, i), $year$ (time period, t), L (log of labour, l_{it}), I (log of investment, i_{it}), K (log of capital, k_{it}), A (age of the firm, a_{it}), X (continuation dummy), and Y (log of output, y_{it}). The production function we wish to estimate is:

$$y_{it} = \alpha_0 + \alpha_A a_{it} + \alpha_L l_{it} + \alpha_K k_{it} + \omega_{it} + e_{it}, \tag{1}$$

where ω_{it} is a productivity shock and e_{it} is an i.i.d. residual that represents measurement error in output (see topic 2 for more details). Perform all computations with STATA and write your do-files from scratch.

[A, 5 Points]. Report sample statistics (number of observations, mean, median, standard deviation, etc.) for the key variables in the data (y_{it}, l_{it}, k_{it}, i_{it}, and a_{it}) for the full sample, the balanced sub-panel (i.e., those firms that are present in all years), and the exiters (i.e., those firms that are not present in all years). Do these statistics seem different? What do these differences tell you about the types of firms that tend to survive versus those that exit?

[B, 10 Points]. Using only the balanced sub-panel compute the total, between, within, and random effects estimators for equation (1). How are they different? Perform a Hausman test of random effects versus fixed effects (i.e., within estimator). What have you learned about firm heterogeneity and about possible measurement error from these results?

[C, 10 Points]. Using the balanced sub-panel, compute difference estimators in which you take differences over t of both sides of equation (1). Report results from estimates of the first (i.e., 1

year) differenced model, second (i.e., 2 years) differenced model and third (i.e., 3 years) differenced model. What do these tell you about measurement error? Base your discussion on Golsbee (2000, NBER).

[D, 10 Points]. The following two questions try to measure the importance of endogenous exit and sample selection:

1. Using the full (unbalanced) panel, compute the pooled and fixed-effect estimators. How do these estimates compare to the pooled and fixed-effect estimates on the balanced panel? What does this tell you about the possible effects of selection in this dataset?

2. Use a Probit model to estimate the probability that a firm exits in period $t+1$ as a function of i_{it}, a_{it}, and k_{it}. (Variable X in the dataset is zero in t if the firm exits in $t+1$.) Compute the implied *inverse mills ratio* (as in a standard endogenous sample selection model) and include it as a regressor in both your pooled and fixed effect regressions above. Does this appear to correct for selection bias?

[E, 20 Points]. By following the procedure detailed in problem sets 3 and 4, implement the control function approach by Olley & Pakes [i.e., OP] (1996) (both ignoring and correcting for endogenous exit) for the estimation of equation (1). Note that in problem sets 3 and 4 there was not any a_{it} variable. Consider both a_{it} and k_{it} as *state variables* and treat them symmetrically: estimate both α_A and α_K only in the second step of the procedure. When correcting for endogenous exit, estimate the probability of survival as a function of i_{it}, a_{it}, and k_{it} (as in question D above).

[F, 10 Points]. Compute the standard errors from the second step of the OP estimation method by *clustered bootstrap*, treating all observations for a single firm as one *cluster*. You can learn how to implement bootstrap methods with STATA from chapter 13 of Cameron & Trivedi (2009). Why

do we need to do this to estimate the standard errors in the second step of the OP estimation procedure?

[G, 15 Points]. How do your OP results compare with the previous ones? Base your discussion on Griliches and Mairesse (1995, NBER).

2 Article Discussion [20 Points]

Discuss the article by Jan De Loecker (Econometrica, 2011): "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity." *Your discussion should be at most 2 pages with font size 12 and at least 2cm of side margins.* In your assessment of the paper, try to be critical: what do you think about it? What is the author actually trying to do? Did he succeed? What are the pros and cons of the article, in your opinion? How does it relate to the existing literature? What do you think is the main contribution of the paper?