Problem Set 3: Productivity

Allan Collard-Wexler (starting with a problem set written by Aviv Nevo)

Due Date: Friday April 13, 2012

The file GMdata.csv contains the data from the Griliches-Mairesse (1995) NBER working paper. There are 9 variables (which appear in this order in the ASCII file): index (firm ID), sic3 (3 digit SIC), yr (year $\in$ {73, 78, 83,88}), ldsal (log of deflated sales), lemp (log of employment), ldnpt (log of deflated capital), ldrst (log of deflated R&D capital), ldrnd (log of deflated R&D), ldinv (log of deflated investment). See the Griliches-Mairesse paper for more details on how the data set was collected. The equation you want to estimate is:

$$ldsal_i^t = \beta_1 lemp_i^t + \beta_2 ldnpt_i^t + \beta_3 ldrst_i^t + d^t + d^t d357 + a_i + \omega^t \quad (1)$$

where d^t are time dummy variables and $d357$ is a dummy variable for computers(SIC 357).

1. Report sample statistics (number of observations, mean, median, standard deviation, etc.) for the variables for both the all sample and the balanced sun-panel (i.e., those firms that are present in all years). Also report these statistics for the firms that existed at least 2 periods. Do these statistics seem different? If so what does this suggest?

2. (a) Using only the balanced sub-panel compute (and report) the total, fixed-effects and random effects estimators, for the above equation.

   (b) Generate measured productivity $\hat{\omega}_i^t$ as the residual from the above regression using the OLS estimator. Describe the autocorrelation with 5 and 10 year lags of productivity. What does this tell you about the nature of productivity dynamics.

   (c) What have you learned about firm heterogeneity from these results?

(d) Using the full (unbalanced) panel compute the total and first difference estimators. Also compute an OLS estimator using only the firms that were present at least 2 periods. How do these estimates compare to the balanced panel estimates? What does this tell you?

3. Selection

   (a) Use a Probit model to estimate the probability that the firm exists in $t+1$ as a function of $ldnpt$, $ldrst$ and $ldinv$. Compute the implied inverse mills ratio (the ratio $\frac{\phi(x)}{\Phi(x)}$ from the Heckman selection model) and include it in the above 1st difference regression and the OLS regression which used the firms that were present in at least 2 periods.

   (b) What does this Heckit procedure (and in particular the coefficient on the propensity score) tell you?

4. Bootstraps (use 200 bootstrap replications)

   (a) Compute bootstrapped standard errors for the OLS estimator of the production function in equation 1 for the entire sample.

   (b) Compute bootstrapped standard errors for the OLS estimator of the production function in equation 1 for the entire sample correcting for correlation across sectors (i.e. clustering by sic codes).

   (c) Compute bootstrapped standard errors for the OLS estimator of the production function in equation 1 for only the sample which is fully balanced, and correct for serial correlation (i.e. clustering by plant).

   (d) Compute standard errors for the Heckman selection model using the bootstrap that account for i) error in the probit regression in the first stage and ii) the correlation by sector.[1]

5. Control Functions

   (a) Do the first-stage ACF Regression with investment controls. What other state variables should be included? What is the effect of the investment proxy.

[1] You may need to investigate the stata code snippet in the lecture notes on the bootstrap.

(b) Do the second-stage moment conditional to estimate the β_1, β_2 and β_3 with $\omega_{it} = \alpha\omega_{it-1} + \mu + \xi_{it}$ as the evolution regression. (I've attached a set of MATA code to help you out with doing GMM in STATA).

(c) Do the second-stage moment conditional to estimate the β_1, β_2 and β_3 with a survival control in the evolution of ω, where the survival depends on firm capital and year.

(d) Interpret the production function coefficients you've estimated and how they differ from the OLS results.

(e) Estimate Standard errors for your estimates in a and b using bootstrap and clustering on the firm.

6. Reallocation

(a) Compute the Olley-Pakes Aggregate Productivity Ω, average productivity $\bar{\omega}$ and Covariance by year for the entire dataset.

(b) Compute the Dynamic Decomposition (for entrants, exitors, cross, between and within) for the entire sample.

(c) Interpret what you've found in a and b.