Problem Set 3 -Estimation of Production Function-

Instructor: Yuta Toyama

Last updated: October 27, 2018
Due date: November 19, 2018

Remarks

1. Use A4 report papers, staple them in the upper left corner.
2. Type your answer.
3. You are allowed (and encouraged) to form a study group up to 3, and hand in one solution per group. If you do this, please put names and student IDs of all members.
4. Attach the print of your programming code for empirical exercise as an appendix.
5. When asked to report results present the answer in a table. Nothing fancy but don't simply attach a printout of the statistical program you used. You should attach the code you used to generate the results as an appendix.

Question 1

Download from the course home page the file "GMdata.csv". It contains the data from the Griliches-Mairesse paper (Griliches and Mairesse 1995, NBER WP). There are 9 variables: $index$ (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.

(a) Report sample statistics (number of observations, mean, median, standard deviation, etc.) for the variables for both the all sample and the balanced sub-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?

(b) Using only the balanced sub-panel compute (and report) the total (OLS using all the data), within and random effects estimators, for:

$$ldsal_{it} = \beta_0 + \beta_1 lemp_{it} + \beta_2 ldnpt_{it} + \beta_3 ldrst_{it} + \gamma_t + \delta_t d357_i + \alpha_i + \epsilon_{it}$$

where γ_t is the year fixed effect and $d357_i$ is a dummy variable for computers (SIC 357) (note that the coefficient δ_t varies by time period).

(c) Perform a Hausman test of random effects versus fixed effects.

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

Question 2 (Selection)

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

(b) Use a Probit model to estimate the probability that a firm exists in $t+1$ as a function of $ldnpt_{it}$, $ldrst_{it}$ and $ldinv_{it}$. Compute the implied inverse mills ratio and include it in the above first difference regression and the OLS regression that used the firms that were present in at least 2 periods.

Question 3 (Olley-Pakes)

Consider the model:

$$ldsal_{it} = \beta_0 + \beta_1 lemp_{it} + \beta_2 ldnpt_{it} + \beta_3 ldrst_{it} + d_t + d_t d357_i + \omega_{it} + \epsilon_{it},$$

where $\omega_{it} = g(\omega_{i,t-1}) + \xi_{it}$ and ξ_{it} is an innovation term.

Let's compute an Olley-Pakes like estimator by the following steps:

(a) Regress $ldsal_{it}$ on $lemp_{it}$, the dummy variables and a second order polynomial in $ldnpt_{it}$, $ldrst_{it}$ and $ldinv_{it}$. Report the coefficients on $lemp_{it}$ and the dummy variables.

Hint: Remember the equation

$$ldsal_{it} = \beta_0 + \beta_1 lemp_{it} + \beta_2 ldnpt_{it} + \beta_3 ldrst_{it} + d_t + d_t d357_i + \omega_{it} + \epsilon_{it}.$$

Olley-Pakes inversion implies that $\omega_{it} = f(ldnpt_{it}, ldrst_{it}, ldinv_{it})$. Note that for simplicity, the function $f(\cdot)$ is common across periods. Since we cannot separately identify this function and the constant term in the production function, the number of time dummies we can estimate in the 1st stage is 3. Denote time dummies as $(y73_t, y78_t, y83_t, y88_t)$. Then, the estimating equation in the 1st stage is

$$\begin{aligned} ldsal_{it} &= \beta_1 lemp_{it} + y73_t + y78_t + y83_t + d357_i \cdot y73_t + d357_i \cdot y78_t + d357_i \cdot y83_t + d357_i \cdot y88_t \\ &\quad + \Phi(ldnpt_{it}, ldrst_{it}, ldinv_{it}) + \epsilon_{it} \end{aligned}$$

where $\Phi(ldnpt_{it}, ldrst_{it}, ldinv_{it}) = \beta_0 + \beta_2 ldnpt_{it} + \beta_3 ldrst_{it} + f(ldnpt_{it}, ldrst_{it}, ldinv_{it})$. We approximate the function Φ by the 2nd order polynomials (including the interaction terms).

(b) Estimate the remaining coefficients $(\beta_0, \beta_2, \beta_3)$ and the productivity process $g(\cdot)$.

Hint: Let $\Phi_{it} \equiv \Phi(ldnpt_{it}, ldrst_{it}, ldinv_{it})$. Then,

$$\Phi_{it} = \beta_0 + \beta_2 ldnpt_{it} + \beta_3 ldrst_{it} + \omega_{it}.$$

Using the fact that $\omega_{it} = g(\omega_{i,t-1}) + \xi_{it}$, we have

$$\begin{aligned} &\Phi_{it} = \beta_0 + \beta_2 ldnpt_{it} + \beta_3 ldrst_{it} + g(\omega_{i,t-1}) + \xi_{it}. \\ \iff &\Phi_{it} = \beta_0 + \beta_2 ldnpt_{it} + \beta_3 ldrst_{it} + g\left(\Phi_{i,t-1} - \beta_0 - \beta_2 ldnpt_{it-1} - \beta_3 ldrst_{it-1}\right) + \xi_{it}. \end{aligned}$$

Now, use the nonlinear least squares to minimize the sum of squares of the following residuals:

$$\Phi_{it} - \beta_0 - \beta_2 ldnpt_{it} - \beta_3 ldrst_{it} - \beta_h h_{it-1}(\beta_0, \beta_2, \beta_3) - \beta_{h2}\left\{h_{it-1}(\beta_0, \beta_2, \beta_3)\right\}^2,$$

where $h_{it-1}(\beta_0, \beta_2, \beta_3) = \Phi_{i,t-1} - \beta_0 - \beta_2 ldnpt_{it-1} - \beta_3 ldrst_{it-1}$. Notice that Φ_{it} can be obtained by the first stage estimates.

(c) Remember that you estimate the probability that a firm exists in $t+1$ as a (linear) function of $ldnpt_{it}$, $ldrst_{it}$ and $ldinv_{it}$ in Q2 (b). We call this probability $P_{j,t+1}$. Use P_{jt} (the probability that firm i exists in period t as a linear function of the three variables $ldnpt_{it}, ldrst_{it}$, and $ldinv_{it}$). Repeat part (ii), but now include P_{it} and P_{it}^2 instead of h_{it} and h_{it}^2.

(d) Repeat part (ii), but include a second order polynomial in both h_{it} and P_{it}.

(e) Based on the results (b)–(d), discuss about the importance of selection and firm effects.