Ec141, Spring 2020

Professor Bryan Graham

Problem Set 4

Due: April 17th, 2020

Problem sets are due at 5PM via bCourses. You may work in groups, but each student should turn in their own write-up (including a narrated/commented and executed Jupyter Notebook). Please use markdown boxes within your Jupyter notebook for narrative answers to the questions below. As always, I encourage you to work together with your classmates (you will learn more and enjoy it more). Make use of Zoom and other collaborative, but social distancing compliant, technologies.

Useful reading: Griliches & Mairesse (1998), Blundell & Bond (2000), de Loecker & Warzynski (2012) and de Loecker (2013).

1 Production function: identification

Let Y_t, K_t, L_t respectively denote a firm's period t output, capital stock and labor. The econometrician observes (Y_{it}, K_{it}, L_{it}) for a random sample of $i = 1, \ldots, N$ firms in each of $t = 0, \ldots, T$ periods.

Output is produced according to

$$Y_t = \exp\left(A + U_t\right) K_t^{\alpha} L_t^{\beta},$$

with Aa time-invariant, firm-specific, component of productivity and U_t a time-varying productivity component. We assume that any period t decisions made by the firm occur *after* observing U_t. Further assume that U_t evolves according to the linear Markov process

$$U_t = \lambda U_{t-1} + \epsilon_t$$

with $\mathbb{E}\left[\left.\epsilon_t\right| U_0^{t-1}\right] = 0$ for $U_0^{t-1} = (U_0, U_1, \ldots, U_{t-1})'$.

Capital evolves according to

$$K_t = (1-\delta) K_{t-1} + I_{t-1}$$

with δ the rate of depreciation and I_t capital expenditure. An implication of the capital process is that a firm's capital stock is determined one period in advance. The firm's beginning of period t information set equals $\mathcal{I}_t = \left(K_0^t, L_0^{t-1}, Y_0^{t-1}, U_0^{t-1}\right)$. Firms choose their investment level, I_t, and employment level, L_t, *after* additionally observing U_t.

[a] Assume that $A \equiv \kappa$, with κ equal to some constant. Discuss this assumption; what does it imply about the distribution of productivity across firms when (i) $0 \leq \rho < 1$ and (ii) $\rho = 1$?

[b] Let $\eta(Z_t, \gamma) = \ln Y_t - \alpha \ln K_t - \beta \ln L_t$ for $Z_t = (K_t, L_t, Y_t)'$, $\gamma = (\alpha, \beta)'$ and, for $\theta = (\zeta, \lambda, \alpha, \beta)'$, let

$$\rho\left(Z_{t-1}^{t}, \theta\right) = \eta(Z_t, \gamma) - \zeta - \lambda \eta(Z_{t-1}, \gamma)$$

with $\zeta = (1 - \rho)\kappa$. Show that, for θ_0 the population parameter,

$$\mathbb{E}\left[\rho\left(Z_{t-1}^{t}, \theta_0\right) \middle| \mathcal{I}_t\right] = \mathbb{E}\left[\epsilon_t \middle| \mathcal{I}_t\right] = 0 \tag{1}$$

for $t = 1, \ldots, T$.

[c] For $T = 1$ show that (1) implies the unconditional moment restriction

$$\mathbb{E}\left[\rho\left(Z_0^1, \theta_0\right) \begin{pmatrix} 1 \\ \ln K_0 \\ \ln K_1 \\ \ln L_0 \\ \ln Y_0 \end{pmatrix}\right] = 0.$$

[d] Now assume that A varies across firms. Let

$$\rho\left(Z_{t-2}^{t}, \theta\right) = \eta(Z_t, \gamma) - \lambda \eta(Z_{t-1}, \gamma) - \left[\eta(Z_{t-1}, \gamma) - \lambda \eta(Z_{t-2}, \gamma)\right]$$

Show that, for $\theta_0 = (\lambda_0, \alpha_0, \beta_0)'$ the population parameter,

$$\mathbb{E}\left[\rho\left(Z_{t-2}^{t}, \theta_0\right) \middle| \mathcal{I}_{t-2}\right] = \mathbb{E}\left[\epsilon_t + \epsilon_{t-1} \middle| \mathcal{I}_{t-1}\right] = 0 \tag{2}$$

for $t = 2, \ldots, T$. Compare the assumptions about productivity made in part [a] with those made here. Which do you find more plausible? Why?

[e] For $T = 2$ show that (2) implies the unconditional moment restriction

$$\mathbb{E}\left[\rho\left(Z_0^2, \theta_0\right) \begin{pmatrix} 1 \\ \ln K_0 \\ \ln K_1 \\ \ln L_0 \\ \ln Y_0 \end{pmatrix}\right] = 0.$$

2 Production function: estimation

The file `semiconductor_firms.out` contains several thousand firm-by-year observations for a sample of publicly traded semiconductor firms (NAICS 4-digit code 3344) drawn from the *S&P Capital IQ - Compustat* database. The following firm attributes, measured from 1998 to 2014 inclusive, are included:

`gvkey` – Compustat firm identification code

`conm` – firm name

`year` – calendar year

`Y` – total real sales by the firm (in millions of 2009 US$)

`K` – capital stock (in millions of 2009 US$)

`L` – employees (in thousands)

`M` – materials expenditures (in millions of 2009 US$)

`VA` - total real valued added by the firm (in millions of 2009 US$)

`w` - annual wage rate (in 2009 US$)

`i` – real investment (in millions of 2009 US$)

`naics_4digits` – NAICS four digit sector code for the firm

This is the same dataset you used in Problem Set 1. For this assignment, keep only those observations corresponding to 2012 ($t = 0$), 2013 ($t = 1$) and 2014 ($t = 2$). Further only retain "complete cases"; that is firms with information on `VA,K,L` in all three periods. This will constitute our estimation sample. In what follows you may treated value added as output.

[a] Construct a table of summary statistics for the estimation sample. How many firms are in the sample?

[b] Using the first two periods of data (i.e., $t = 0, 1$). Related changes in log output to changes in log inputs using OLS. Under what assumptions does this approach provide consistent capital and labor coefficients?

[c] Consider the model outlined in part [a] to [c] of question 1 above. Write a computer program that implements Algorithm 1.

[d] Explain why (3) should be small when $\theta^{(s)} \approx \theta_0$. More generally give a verbal justification for the estimation procedure with reference to your theoretical analysis in the first part of the problem set. Can you think of any modifications you might like to make to your procedure? Speculate on any advantages or disadvantages of these modifications.

Algorithm 1 "Olley-Pakes" Type Estimation Procedure

1. Let $\alpha^{(s)}$ and $\beta^{(s)}$ be the current values of $\alpha$and β. Create the "variables" $\eta\left(Z_{i0}, \gamma^{(s)}\right) = \ln Y_{i0} - \alpha^{(s)} \ln K_{i0} - \beta^{(s)} \ln L_{i0}$ and $\eta\left(Z_{i1}, \gamma^{(s)}\right) = \ln Y_{i1} - \alpha^{(s)} \ln K_{i1} - \beta^{(s)} \ln L_{i1}$ for $i = 1, \ldots, N$. Note you will only use years 2012 and 2013 for this part of the problem set.

2. Compute the OLS fit of $\eta\left(Z_{i1}, \gamma^{(s)}\right)$ onto a constant and $\eta\left(Z_{i0}, \gamma^{(s)}\right)$. Let the intercept and slope coefficient from this fit be $\zeta^{(s)}$ and $\lambda^{(s)}$ respectively.

3. Form the sample moment 5×1 vector

$$\begin{aligned} \psi_N\left(\theta^{(s)}\right) = &\frac{1}{N} \sum_{i=1}^{N} \Big(\eta\left(Z_{i1}, \gamma^{(s)}\right) - \zeta^{(s)} \\ &- \lambda^{(s)} \eta\left(Z_{i0}, \gamma^{(s)}\right) \Big) \begin{pmatrix} 1 \\ \ln K_{i0} \\ \ln K_{i1} \\ \ln L_{i0} \\ \ln Y_{i0} \end{pmatrix} \end{aligned}$$

   and the associated quadratic form

$$\psi_N\left(\theta^{(s)}\right)' \psi_N\left(\theta^{(s)}\right). \tag{3}$$

4. Repeat steps 1 to 3 for all pairs $\alpha^{(s)} \in \{0.05, 0.06, \ldots, 0.95\}$ and $\beta^{(s)} \in \{0.05, 0.06, \ldots, 0.95\}$. Let $\hat{\alpha}, \hat{\beta}$ be the $\alpha^{(s)}, \beta^{(s)}$ pair which minimizes (3). Let $\hat{\zeta}$ and $\hat{\lambda}$ be the associated $\zeta^{(s)}$ and $\lambda^{(s)}$ estimates from step 3.

Algorithm 2 Bootstrap Standard Errors

1. For each of $b = 1 \ldots B$ bootstrap replications:
   (a) Draw N firms *with replacement* from you estimation sample. Sample firms, not firm-years. That is, if firm i is sampled, all periods of firm i's data should be included. Since you are sampling with replacement it is possible that the same firm may appear multiple times in a given bootstrap sample.
   (b) Using the bootstrap sample constructed in part (a) form an estimate of θ using Algorithm 1. Call this estimate $\hat{\theta}^{(b)}$.
2. Compute the standard deviations of the components of $\hat{\theta}^{(b)}$ across your B boostrap replications. Use these standard deviations as standard error estimates for your point estimates of θconstructed using the actual sample.

---

[e] Compute $\widehat{A+U_{i1}} = Y_{i1} - \hat{\alpha}\ln K_{i1} - \hat{\beta}_1 \ln L_{i1}$. Plot a histogram of $\widehat{A+U_{i1}}$. Compare your analysis with the productivity analysis you undertook in Problem Set 1. Compute the average, standard deviation and 5th, 25th, 50th, 75th and 95th percentiles of the sample distribution of $\widehat{A+U_{i1}}$.

[f] To construct standard errors for your estimate of θ_0 you will use the the bootstrap procedure described in Algorithm 2. Set $B = 1000$ (or more!). Report your estimation results (with bootstrap standard errors) in an easy-to-read table.

[g] Try to construct an estimation and inference procedure along the lines of the one outlined above, but this time appropriate for the model outlined in parts [d] and [e] of part 1 of the problem set. Carefully implement and describe your procedure. Repeat parts [e[and [f] above with your new procedure's coefficient and productivity estimates.

[h] What have you learned about the distribution of productivity across large U.S. semiconductor firms? What else are you interested in learning? What data/methods might help you do so?

References

Blundell, R. & Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), 321 – 340.

de Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, 5(3), 1 – 21.

de Loecker, J. & Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6), 2437 – 2471.

Griliches, Z. & Mairesse, J. (1998). *Econometrics and Economic Theory in the 20th Century*, chapter Production functions: the search for identification, (pp. 169 – 203). Cambridge University Press: Cambridge.