Ec241a, Spring 2023

Professor Bryan Graham

Problem Set 2

Due: April 5, 2023

Problem sets are due in class. You may work in groups (indeed I encourage you to all work together), but each student should turn in their own write-up.

Production function estimation using panel data

Readings: Olley and Pakes (1996), Blundell and Bond (2000), Bloom et al. (2013), de Loecker (2013)

The file `semiconductor_firms.out`, available in the Problem Sets folder on the course GitHub, contains information on 113 semiconductor firms (NAICS 4-digit code 3344) for the years 2010 to 2014 (i.e., $t = 0, 1, \ldots, T$ with $T = 4$ – four periods plus an initial condition). Each row of the dataset corresponds to a firm-year observation. Included in the file are the variables:

Y, output (specifically total sales, in millions of dollars, deflated by an industry-specific output price index)

K, total capital (gross property, plant and equipment in millions of dollars deflated by an industry-by-estimated-capital-age-specific capital price index)

L, total employed labor (thousands of workers)

M, materials (specifically total sales minus operating income before depreciation minus total labor expenses minus R&D expenditures, deflated by an industry-specific output price index)

VA, valued added $(Y - M)$ – note that a few firm-year VA measures are negative. This typically reflects a firm reporting negative operating income before depreciation in a given year). You will need to drop firms with any negative firm-year VA measures from your estimation sample.

$r\&d$, total spending on research and development (deflated by an industry-specific investment price index)

w, industry-specific average annual pay measure (deflated by an industry-specific output price index)

i, total capital expenditure (i.e., investment) in millions of dollars (deflated by an industry-specific investment price index)

You will use this dataset to estimate the parameters of a value added production function of the form

$$VA_{it} = K_{it}^{\alpha} L_{it}^{\beta} \exp\left(A_{it} + U_{it}\right)$$

with U_{it} an unforecastable productivity shock which is realized after all period t input decisions are made and A_{it} a systematic productivity component with

$$A_{it} = h\left(A_{it-1}, X_{it-1}\right) + \epsilon_{it},\ \mathbb{E}\left[\epsilon_{it} \middle| \mathcal{I}_{it}\right] = 0$$

with $\mathcal{I}_{it}$ firm i's beginning-of-period t information set and X_{it} expenditure on research and development. Both capital and labor are costly adjust, such that

$$L_{it} = g_L\left(L_{it-1}, K_{it}, A_{it}; W_t\right)$$
$$I_{it} = g_K\left(L_{it}, K_{it}, A_{it}; W_t\right)$$

with I_{it} equal to period t capital expenditures (which are reflected in the period $t+1$ capital stock) and W_t a vector industry-specific input prices.

1. Let $Z_0^t = (Z_0, \ldots, Z_t)'$. What elements of L_0^T, K_0^T, I_0^T, X_0^T and VA_0^T are part of $\mathcal{I}_t$? In what sense is the estimation problem a sequential moments one?

2. Comment on the restriction that $A_{it} \perp \mathcal{I}_{it} | \, A_{it-1}, X_{it-1}$. Give a few examples of how it could be violated.

3. Implement the following algorithm (e.g., Bloom et al., 2019, p. 1656):

   (a) Compute the OLS fit of $\ln Y_{it}$ onto $\ln K_{it}$, $\ln L_{it}$, $\ln X_{it}$, a vector of firm-specific dummy variables and a vector of time-specific dummy variables.

   (b) Discuss the problems, if any, with this approach in light of the economic model sketched above.

4. Compute estimates of α, β and $h(A_{it-1}, X_{it-1})$ using the "Olley-Pakes" procedure described in lecture (carefully describe any modeling/implementation choices you make).

5. Assume that $h(A_{it-1}, X_{it-1})$ is linear. Now estimate using the "Blundell-Bond" procedure described in lecture.

6. Summarize your empirical findings.

7. Provide measures of statistical precision for as many estimated parameters as you are able to. Justify your measures as well as you are able to.

8. Discuss how you might construct an estimator with greater precision?

Nonlinear panel data

Readings: Chamberlain (1980), Chamberlain (1985), Hahn (1994), Arellano (2003)

Let $t = 1, 2$ index two spells, of lengths Y_1 and Y_2. For example Y_1 might equal time spent of first job, Y_2 time spent on second job. Alternatively Y_1 might be time to first publication post PhD, and Y_2 the time that elapses between the first and second post-PhD publication etc. Let X_t be a vector of spell-varying regressors and A spell-invariant heterogeneity. Assume that the hazard function takes the form

$$\lambda_0(y_t | \, x, a; \theta_0) = \lambda_0(y_t, \alpha_0) \exp(x_t' \beta_0 + a), \; t = 1, 2$$

with $\theta = (\alpha, \beta')'$ and $\lambda_0(y_t, \alpha)$ a known baseline hazard function indexed by α. You may assume that $Y_1$and Y_2 are conditionally independent given $X = (X_1, X_2)'$ and A. Let $Z = (X_1', Y_1, X_2' Y_2)'$ and assume that $\{(Z_i', A_i)\}_{i=1}^{\infty}$ is an iid sequence with A_i unobserved. Let $\rho(z_t, \theta) = \Lambda_0(y_1, \alpha) \exp(x_t' \beta)$ for $t = 1, 2$ with$z_t = (x_t', y_t)'$ and $\bar{\rho}(z, \theta) = \rho(z_1, \theta) + \rho(z_2, \theta)$. Additionally define the random variables $P_1 = \rho(Z_1, \theta_0)$, $P_2 = \rho(Z_2, \theta_0)$ and $\bar{P} = P_1 + P_2$.

[a] Show that the joint density for the two spells given $X = x$ and $A = a$

$$f_{Y_1, Y_2 | X, A}(y_1, y_2 | \, x, a; \theta_0) = \lambda_0(y_1, \alpha_0) \exp(x_1' \beta_0) \lambda_0(y_2, \alpha_0) \exp(x_2' \beta_0) \frac{f_{\bar{P}|A}(\bar{\rho}(z, \theta_0) | \, a)}{\bar{\rho}(z, \theta_0)}$$

where $f_{\bar{P}|A}(v|\, a) = e^{2a} v \exp(-e^a v)$ is the density for the Gamma$(2, e^a)$ random variable $\bar{P}$ at $\bar{P} = v$.

[b] Let $\pi(a|\, x)$ be the unknown density for the conditional distribution of A given X such that

$$f_{\bar{P}|X}(\bar{\rho}|\, x) = \int e^{2a} \bar{\rho} \exp(-e^a \bar{\rho})\, \pi(a|\, x)\, \mathrm{d}a.$$

Show that

$$\frac{\frac{\partial}{\partial \bar{\rho}} f_{\bar{P}|X}(\bar{\rho}|\, x)}{f_{\bar{P}|X}(\bar{\rho}|\, x)} = \frac{1}{\bar{\rho}} - \mathbb{E}\left[e^A \middle|\, \bar{\rho}, x \right].$$

[c] Consider the joint fixed effects estimator which treats $\delta_i = e^{A_i}$ as an additional parameter to be estimated. Show that unit i's contribution to the log-likelihood function in this case is

$$\begin{aligned} l_i^{\mathrm{JFE}}(\theta, \delta_i) =& \ln \lambda_0(Y_{i1}, \alpha) + \ln \lambda_0(Y_{i2}, \alpha) + (X_{i1} + X_{i2})' \beta \\ &+ 2 \ln \delta_i - [\rho(Z_{i1}, \theta) + \rho(Z_{i2}, \theta)]\, \delta_i. \end{aligned} \tag{1}$$

[d] Show that for a fixed value of θ, the MLE of δ_i is

$$\hat{\delta}_i(\theta) = \left[\frac{\rho(Z_{i1}, \theta) + \rho(Z_{i2}, \theta)}{2} \right]^{-1}.$$

Observe that $\hat{\delta}_i(\theta)$ only varies with unit i's data.

[e] Show that unit i's contribution to the concentrated log-likelihood function equals

$$l_i^{\mathrm{JFE}}\left(\theta, \hat{\delta}_i(\theta)\right) = \ln \lambda_0(Y_{i1}, \alpha) + \ln \lambda_0(Y_{i2}, \alpha) + (X_{i1} + X_{i2})' \beta - 2 \ln \left[\frac{\rho(Z_{1i}, \theta) + \rho(Z_{i2}, \theta)}{2} \right] - 2.$$

[f] Show that the the joint fixed effects estimate of β solves the sample analog of the population moment restriction

$$\mathbb{E}\left[\psi^{\mathrm{JFE}}(Z_i, \theta) \right] = \mathbb{E}\left[-(X_{i2} - X_{i1}) \left\{ \frac{\rho(Z_{2i}, \theta) - \rho(Z_{1i}, \theta)}{\rho(Z_{i1}, \theta) + \rho(Z_{2i}, \theta)} \right\} \right] = 0.$$

Show that $\mathbb{E}\left[\psi^{\mathrm{JFE}}(Z_i, \theta) \right]$ is, indeed, mean zero at $\theta = \theta_0$. Discuss your finding in light of what you know about incidental parameters bias. Can you say anything about the JFE estimate of α_0, the parameter indexing the baseline hazard? Discuss.

References

Arellano, M. (2003). Discrete choices with panel data. *Investigaciones Económicas*, 27(3):423 – 458.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., and Van Reenen, J. (2019). What drives differences in management practices? *American Economic Review*, 109(5):1648 – 1683.

Bloom, N., Schankerman, M., and van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347 – 1393.

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

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225 – 238.

Chamberlain, G. (1985). *Longitudinal Analysis of Labor Market Data*, chapter Heterogeneity, omitted variable bias, and duration dependence, pages 3 – 38. Cambridge University Press, Cambridge.

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

Hahn, J. (1994). The efficiency bound of the mixed proportional hazards model. *Review of Economic Studies*, 61(4):607 – 629.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263 – 1297.