Ec2147, Spring 2021

Professor Bryan Graham

Problem Set 2

Due: April 27th, 2021

Problem sets are due by 5PM EST. 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

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

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

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 $h\left(A_{it-1}, X_{it-1}\right)$ using the "Olley-Pakes" procedure described in lecture (carefully describe any modelling/implementation choices you make).

5. Assume that $h\left(A_{it-1}, X_{it-1}\right)$ 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?

Econometrics field exam question

Let $\{(W_i, X_i, Y_i)\}_{i=1}^N$ be a simple random sample drawn from a population characterized by (unknown) distribution F_0; Y is a scalar outcome variable of interest, X a vector of policy or treatment variables, and W a vector of additional control variables or confounders. Assume that the conditional distribution of Y_i given W_i and X_i satisfies:

$$Y = X'\beta_0 + h_0\left(W\right) + U, \ \mathbb{E}\left[U |\, W, X\right] = 0. \tag{1}$$

Here $\beta_0 \subset \mathbb{R}^K$ is the finite dimensional parameter of interest and $h_0\left(W\right)$ an unknown function mapping from a subset of $W \in \mathbb{W} \subset \mathbb{R}^{\dim(W)}$ into $\mathcal{H} \subset \mathbb{R}$. You may assume that all the expressions which appear below are well-defined (i.e., don't worry about regularity conditions in what follows).

(a) Assume that the researcher knows that $h_0\left(W\right) = k\left(W\right)'\delta_0$ for some known $J \times 1$ vector of basis functions $k\left(W\right)$ (which includes a constant) and unknown parameter δ_0. Show that the coefficient on X in the least squares fit of Y onto X and $k\left(W\right)$ has an asymptotic limiting distribution of

$$\sqrt{N}\left(\hat{\beta}_{\text{OLS}} - \beta_0\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2 \mathbb{E}\left[\left(X - \Pi_0 k\left(W\right)\right)\left(X - \Pi_0 k\left(W\right)\right)'\right]^{-1}\right), \tag{2}$$

with $\Pi_0 = \mathbb{E}\left[Xk(W)'\right]\mathbb{E}\left[k(W)k(W)'\right]^{-1}$ the $K \times J$ matrix of projection coefficients associated with the multivariate regression of X onto $k(W)$. *Note this result maintains the additional assumption of homoscedasticity.* For your reference the partitioned matrix formula is:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B\left(D - CA^{-1}B\right)CA^{-1} & -A^{-1}B\left(D - CA^{-1}B\right) \\ -\left(D - CA^{-1}B\right)CA^{-1} & \left(D - CA^{-1}B\right) \end{bmatrix}.$$

(b) Let

$$e_0(w) = \mathbb{E}\left[X \mid W = w\right]. \tag{3}$$

be a vector of conditional expectation functions (CEFs) of the K policy variables given the confounders. Assume that $e_0(W)$ is known. Show that following unconditional moment identifies β_0:

$$\mathbb{E}\left[m\left(Z, \beta_0, e_0(W)\right)\right] = 0. \tag{4}$$

with $m\left(Z, \beta_0, e_0(W)\right) = \left(Y - X'\beta_0\right)\left(X - e_0(W)\right)$. Let $\hat{\beta}_E$ be the method-of-moments estimate associated with this restriction. Is this estimator more or less efficient than the OLS estimator described in part (a)?

(c) Augment the identifying "E-moment" (4) from part (b) with the auxiliary conditional moment

$$\mathbb{E}\left[\left(X - e_0(W)\right) \mid W\right] = 0. \tag{5}$$

Argue that the efficient estimator based on (4) and (5) coincides with the just-identified moment based upon the unconditional moment restriction

$$\psi\left(Z, \beta_0, e_0(W), h_0(W)\right) = m\left(Z, \beta_0, e_0(W)\right) - \mathbb{E}^*\left[m\left(Z, \beta_0, e_0(W)\right) \mid \left(X - e_0(W)\right); W\right].$$

Here $\mathbb{E}^*\left[Y \mid X; W\right]$ denotes the linear regression of Y onto X within a subpopulation homogenous in W (i.e., a conditional linear predictor). You may assume that all "unknowns" except β_0 are known. You may not assume that "unknown unknowns are known" only that "known unknowns are known".

Further show that

$$\psi\left(Z, \beta_0, e_0(W), h_0(W)\right) = \left(Y - X'\beta_0 - h_0(W)\right)\left(X - e_0(W)\right). \tag{6}$$

(d) Replace $h_0(w)$ in (6) with some arbitrary function of w. Is (6) still mean zero? Now replace $e_0(w)$ in (6) with some arbitrary function of w. Is (6) still mean zero? Comment on the implications of your answer for estimation.

(e) Write, in the language of your choice, a short Haiku inspired by questions (a) to (d) above.

References

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.