Ec240a, Fall 2018

Professor Bryan Graham

Problem Set 5

Due: December 7th, 2018

Problem sets are due at 5PM in the GSIs mailbox. You may work in groups, but each student should turn in their own write-up (including a printout of a narrated/commented and executed Jupyter Notebook if applicable). Please also e-mail a copy of any Jupyter Notebook to the GSI (if applicable).

1 Estimating productivity

The file `mf_firms.out` contains 29,836 firm-by-year observations for a sample of publicly traded manufacturing firms 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

`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$)

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

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

You may find the Griliches and Mairesse (1998) chapter on the course syllabus useful background reading for this problem set.

1.1 Preparing the dataset

In this problem set we will focus on manufacturing firms. These correspond to firms with `naics_4digits` codes beginning with a 31, 32 or 33. Delete all other records. We will further focus on the calendar years from 2010 to 2014 inclusive. Delete observations associated with all other calendar years. Further restrict your sample to include only those firms with complete data for all five years (i.e., construct a balanced panel with $t = 1, 2, 3, 4, 5$

observations per firm). Finally drop any firms with negative measured valued added in any year.

How many firm-year observations are left? How many distinct firms? In 2014 what was the aggregate total sales across all manufacturing firms in the final dataset? How many workers did these firms employ in total?

In 2014 compute the average, standard deviation and 5th, 25th, 50th, 75th and 95th percentiles of total sales, valued added, capital stock, employees, materials, and investment across all firms in your dataset. Display this information in a table.

Write a few sentences summarizing your dataset.

1.2 Panel data production function estimation

For this part of the problem set we will work with a valued added (total sales minus materials expenditures) production function. Let

$$VA_t = A_t K_t^{\alpha_0} L_t^{\beta_0}$$

Taking logs yields:

$$\ln \mathrm{VA_t} = \alpha_0 \ln \mathrm{K}_t + \beta_0 \ln \mathrm{L}_t + \ln A_t.$$

Let $\mathcal{I}_t$ denote a firm's period t information set. Assume that firm's costlessly adjust their capital and labor input levels each period so as to maximize expected current profits. Assume firm's are price-takers.

1. Assume that log TFP is such that

$$\ln A_t = \lambda_t^* + A + U_t.$$

   Argue that if (i) $U_t | \mathcal{I}_t \overset{D}{=} U_t$ and (ii) U_t is i.i.d. over time that

$$\mathbb{E}\left[U_t | X_1, \ldots, X_T \right] = 0$$

   where $X_t = (\ln K_t, \ln L_t)'$. Interpret these assumptions.

2. Let $\mathbb{E}^*\left[A | X_1, \ldots, X_T \right] = \kappa + X_1' \pi_1 + \cdots + X_T' \pi_T$. What sign do you expect the components of π_t to have? Why? Show that

$$\mathbb{E}^*\left[\ln A_t | X_1, \ldots, X_T \right] = \lambda_t + X_1' \pi_1 + \cdots + X_T' \pi_T$$

for $\lambda_t = \lambda_t^* + \kappa$ and hence that

$$\mathbb{E}^* \left[\ln \mathrm{VA_t} \middle| X_1, \ldots, X_T \right] = \lambda_t + X_t'\theta_0 + X_1'\pi_1 + \cdots + X_T'\pi_T$$

for $\theta_0 = (\alpha_0, \beta_0)'$.

3. Let N equal the number of firms in your dataset. Form a dataframe with NT rows where rows $iT+1$ to $(i+1)T$ take the form

t	$\ln \mathrm{VA_t}$	X_t	X_1	X_2	$\cdots$	X_T	D_1	D_2	$\cdots$	D_T
1	$\ln \mathrm{VA_{i1}}$	X_{i1}	X_{i1}	X_{i2}	$\cdots$	X_{iT1}	1	0	$\cdots$	0
2	$\ln \mathrm{VA_{i2}}$	X_{i2}	X_{i1}	X_{i2}	$\cdots$	X_{iT}	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
T	$\ln \mathrm{VA_{iT}}$	X_{iT}	X_{i1}	X_{i2}	$\cdots$	X_{iT}	0	0	$\cdots$	1

Note that D_t is a time dummy variable. Compute the least squares fit of $\ln \mathrm{VA_t}$ onto X_t, $X_1, X_2, \ldots, X_T$ and the T time dummies. No constant is required since all time dummies are included. Why? Be sure to report "cluster-robust" standard errors for your point estimates. You should add this inference option to you existing OLS Python function. Present you results in a table. Discuss them, do they accord with your priors.

4. Test the null hypothesis of constant returns to scale. Construct a Wald statistic to test the null hypothesis that $\pi_1 = \cdots = \pi_T = 0$. Interpret your findings.

5. Compute the least squares fit of $\ln \mathrm{VA_t}$ onto X_t and the T time dummies. How do your results change relative to those in the previous questions? Which set of results do you prefer? Why?

1.3 Total factor productivity estimates

For this part you need to select one of your two production function estimates as your preferred one. State which one you are using and why. This is also a good time to just write a few paragraphs summarizing your analysis so far.

After doing this, then for each firm, using the 2014 data, compute the ratio

$$\frac{VA_t}{K_t^{\alpha} L_t^{\beta}}$$

using your preferred input elasticity estimates. Interpret this ratio and summarize its distribution in a table or figure. Discuss your display. Compute TFP measures for the other years as well.

1.4 Productivity decomposition

Let $S_t = Y_t / \mathbb{E}[Y_t]$ and show that

$$\begin{aligned} \mathbb{E}[S_t A_t] &= \mathbb{E}[S_t]\,\mathbb{E}[A_t] + \mathbb{C}(S_t, A_t) \\ &= \mathbb{E}[A_t] + \beta_{A_t,S_t}\mathbb{V}(A_t), \end{aligned}$$

where β_{A_t,S_t} is the coefficient on A_t in the best linear predictor of S_t onto A_t. Interpret this decomposition and calculate estimates of its two terms for 2010 to 2014. Summarize your results in a table.

1.5 Wrap up

Write three or four paragraphs summarizing your analysis of this dataset.