Ec142, Spring 2018

Professor Bryan Graham

Problem Set 2

Due: March 9th, 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). Please also e-mail a copy of any such notebooks to the GSI (if applicable). Please use markdown boxes within your Jupyter notebook for narrative answers to the questions that appear below.

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

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. 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 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, 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 OLS production function estimates

For this part use only 2014 observations.

Compute the least squares fit of log sales onto a constant, log capital, log labor, and log materials. Report heteroscedastic robust standard errors. Briefly comment on your results. Can you reject the null hypothesis of constant returns to scale? Do you believe your coefficient estimates are consistent for the underlying output elasticities? Why?

Create a dummy variable for each four digit NAICS sectors. Compute a second OLS fit with these dummy variables also included (remember you will either need to exclude the constant or one of the dummy variables). Briefly comment on your results. Can you reject the null hypothesis of constant returns to scale? Do you believe your coefficient estimates are consistent for the underlying output elasticities? Why?

1.3 Panel production function estimates

For this part use all firms with complete data for both 2013 and 2014. How many firms are dropped due to a lack of information for 2013?

Compute the least squares fit of the *change* in log output onto a constant, the change in log capital, the change in log labor, and the change in log materials. Report heteroscedastic robust standard errors. Briefly comment on your results. Can you reject the null hypothesis of constant returns to scale? Do you believe your coefficient estimates are consistent for the underlying output elasticities? Why? How does this analysis differ from the cross-sectional one based on the 2014 data alone? What type of assumptions would justify this "first differenced" approach?

Repeat your analysis now additionally adding the sector-specific dummy variables. Comment on these results.

1.4 Proxy variable production function estimates

Add the following additional variables to your dataframe: the square of log capital, the square of investment, and the interaction of log capital and investment. For this part of the problem set you will work with valued added, which equals total sales minus materials expenditures. Let

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

for $t = 2013, 2014$. Taking logs yields:

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

Let $W = \left(1, \ln \mathrm{K}, \mathrm{i}, (\ln \mathrm{K})^2, \mathrm{i}^2, (\ln \mathrm{K}) \times \mathrm{i}\right)'$ and consider the linear predictor

$$\mathbb{E}^* \left[\ln \mathrm{VA_t} \middle| \ln \mathrm{L}_t, W_t\right] = b_0 \ln \mathrm{L_t} + \sum_{j=1}^{6} \pi_{tj} W_{jt}. \tag{1}$$

Under what conditions does b_0 coincide with the elasticity of output with respect to labor, β_0? Please provide a precise set of conditions as well as an informal discussion of the underlying economics. You might refer to the course readings in order to improve your answer.

Define $\varphi_t(K_t, i_t) = \sum_{j=1}^{6} \pi_{tj} W_{jt}$ and assume that log TFP is such that

$$\mathbb{E}\left[\ln A_t \middle| \mathcal{I}_{t-1}\right] = \lambda_0 + \rho \ln A_{t-1},$$

where $\mathcal{I}_{t-1}$ denotes the firm's period $t-1$ information set. Let $V_t = \ln A_t - \lambda_0 - \rho \ln A_{t-1}$ be the unforecastable innovation to firm log TFP. Describe this TFP process in words.

Next show that

$$\ln \mathrm{VA_t} - \beta_0 \ln \mathrm{L}_t = \lambda_0 + \alpha_0 \ln \mathrm{K}_t + \rho\left(\varphi_{t-1}(K_{t-1}, i_{t-1}) - \alpha_0 \ln \mathrm{K}_{t-1}\right) + V_t. \tag{2}$$

Complete the following computations

1. Using the 2013 data compute the sample analog of equation (1). Discuss your estimate of b_0.
2. Construct, for each firm, the estimate $\hat{\varphi}_{2013}(K_{2013}, i_{2013}) = \sum_{j=1}^{6} \hat{\pi}_{2013j} W_{j2013}$.
3. Compute the OLS fit of $\ln \mathrm{VA}_{2014} - \hat{b} \ln \mathrm{L}_{2014}$ onto a constant $\ln K_{2014}$, $\hat{\varphi}_{2013}(K_{2013}, i_{2013})$, and $\ln \mathrm{K}_{2013}$.

What are your implied estimates of α_0, β_0 and ρ_0? Discuss your estimation procedure and the assumptions justifying it.

1.5 Total factor productivity estimates

For this part you need to select one of your 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

$$\frac{Y}{K^{\alpha}L^{\beta}M^{\gamma}}$$

using your preferred input elasticity estimates. Interpret this ratio. Summarize its distribution (use a table and/or figure). Note if your prefer the estimates constructed in part 2.4 compute the ratio

$$\frac{VA}{K^{\alpha}L^{\beta}}$$

instead.

1.6 Productivity decomposition

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

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

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 2014.

1.7 Wrap up

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