Homework 1

Econ 8180: Empirical IO.

Due on February 2, 2021

Each group hands in only one homework.

Each member in the group has to sign this statement:
Pledge: On my honor, I pledge that I have neither given nor received help on this assignment. I also pledge that I have not used graphying calculators during the exam. I also pledge that I have not reviewed or used or consulted the solutions of homeworks from previous years in which this class was taught.
Name and Signature of Student (*if this is a group homework, all students must sign*):

1 Question 1 (Production Function Estimation)

Download from Collab the file GMdata. It contains the data from the Griliches-Mairesse paper.

There are 9 variables: index (firm ID), sic3 (3 digit SIC), yr (year = 73, 78, 83, 88), ldsal (log of deflated sales), lemp (log of employment), ldnpt (log of deflated capital), ldrst (log of deflated R&D capital), ldrnd (log of deflated R&D), ldinv (log of deflated investment). They are deflated to make them comparable over years. See the Griliches-Mairess paper for more details on how the data set was collected.

The equation you want to estimate is:

$$ldsal_{it} = \beta_1 \cdot lemp_{it} + \beta_2 \cdot ldnpt_{it} + \beta_3 \cdot ldrst_{it} + d_t + d_t \cdot d357 + a_i + \epsilon_{it}$$

where d_t are year dummy variables and $d357$ is a dummy variable for computers (SIC 357). Notice that in this example $lemp_{it}$ is what I called L in class and what Griliches and Mairess call x in their paper. $ldnpt_{it}$ and $ldrst_{it}$ is what I called K_{jt} in class and Griliches and Mairesse call z in their paper.

To open the dataset in STATA write the following commands:

cd c:/yourdirectory

set memory 50m (this allocates 50m of the computer memory to Stata)

set matsize 100 (this tells the computer that you are working with a dataset that has an horizontal size less than 100 columns)

use GMdata

Once you have open the dataset, do the following:

1) Report sample statistics (number of observations, mean, median, standard deviation, etc..) for the variables. *(Commands: summarize, tabulate, centile).* Report the statistics for the firms that existed at least 2 periods *(Hint: first create a variable, e.g. flag, that is equal to 1 if a firm is present in at least two periods. Then do the analysis for those firms for which flag==1).* Do these statistics seem different? If so, what does this suggest? Report the statistics for the firms that were present in all years (the so called balanced panel). Do these statistics seem different? If so, what does this suggest?

2) Compute and report the total (OLS), between, within, and random effects estimators for the above equation for the original unbalanced panel and for the balanced panel. (*Commands: regress, xtreg*). What can you say about the coefficient estimates across the two specifications? Comment on their magnitudes and standard errors. Compare your results with those presented in Table 3 of the Griliches-Mairess paper. Do you get the same results? (you should!).

3) Perform a Hausman test of random effects versus fixed effects. (*Command: hausman*). What is the main intuition behing the Hausman test? Do you reject it? What do you conclude?

4) Compute an Olley-Pakes like estimator by computing the following steps.

i) Regress $ldsal_{it}$ on $lemp_{it}$, the dummy variables and a 2^{nd} order polynomial ϕ in $ldnpt_{it}$, $ldrst_{it}$, and $ldinv_{it}$:

$$ldsal_{it} = \beta_1 \cdot lemp_{it} + \theta_1 d_t + \theta_2 d_t \cdot d357 + \phi(ldnpt_{it}, ldrst_{it}, andldinv_{it}) + e$$

Notice that you were not using $ldinv_{it}$ in the previous estimation. Report the coefficients on $lemp_{it}$ and the dummy variables.

ii) Regress the following regression:

$$\begin{aligned} & ldsal_{i,t+1} - \left(\hat{\beta}_1 lemp_{i,t+1} + \hat{\theta}_1 d_{t+1} + \hat{\theta}_2 d_{t+1} \cdot d357\right) - \hat{\phi}_t(ldnpt_{it}, ldrst_{it}, andldinv_{it}) \\ = \; & \alpha_1 \left(ldnpt_{it+1} - ldnpt_{it}\right) + \alpha_2 \left(ldrst_{it+1} - ldrst_{it}\right) + A_{it+1} - A_{it} + \zeta_{it+1} \end{aligned}$$

to estimate α_1 and α_2.

iii) Use a probit model *(Command; probit)* to estimate the probability that the firm exists at time $t+1$ as a function of $ldnpt_{it}$, $ldrst_{it}$, and $ldinv_{it}$. Denote by $\hat{P}_{it}$ the predicted probability from this model *(Command: predict)*. Run the following regression:

$$\begin{aligned} & ldsal_{i,t+1} - \left(\hat{\beta}_1 lemp_{i,t+1} + \hat{\theta}_1 d_{t+1} + \hat{\theta}_2 d_{t+1} \cdot d357\right) = \alpha_1 ldnpt_{it+1} + \alpha_2 ldrst_{it+1} \\ & g\left[\hat{\phi}_t(ldnpt_{it}, ldrst_{it}, andldinv_{it}) - \alpha_1 ldnpt_{it} - \alpha_2 ldrst_{it}, \hat{P}_{it}\right] + A_{it+1} - A_{it} + \zeta_{it+1}, \end{aligned}$$

where we approximate the function g with a 2-nd order polynomial as

well:

$$
\begin{aligned}
& g\left[\hat{\phi}_t(ldnpt_{it}, ldrst_{it}, andldinv_{it}) - \alpha_1 ldnpt_{it} - \alpha_2 ldrst_{it}, \hat{P}_{it}\right] = \\
& \hat{\phi}_t(ldnpt_{it}, ldrst_{it}, andldinv_{it}) - \alpha_1 ldnpt_{it} - \alpha_2 ldrst_{it} + \hat{P}_{it} \\
& + \left[\hat{\phi}_t(ldnpt_{it}, ldrst_{it}, andldinv_{it}) - \alpha_1 ldnpt_{it} - \alpha_2 ldrst_{it}\right]^2 + \hat{P}_{it}^2 \\
& + \left[\hat{\phi}_t(ldnpt_{it}, ldrst_{it}, andldinv_{it}) - \alpha_1 ldnpt_{it} - \alpha_2 ldrst_{it}\right] \hat{P}_{it}.
\end{aligned}
$$

To estimate this function you will run non-linear-least-squares *(Command: nl - use the help command!!).*

Compare the results from parte 2 with these ones. What are the potential benefits from using an additional investment decision equation?