Problem Set 1

Ana, Daniela, Rafael

October 25, 2022

Productivity Estimation

Question 1

Table 1: Summary Statistics for the Full Sample

	Mean	SD	Min	Perc. 25	Median	Perc. 75	Max	N
Log of Output	13.49	1.7	5.91	12.42	13.59	14.66	19.16	39,569
Log of Labor	5.00	1.0	0.62	4.33	5.01	5.68	8.86	39,569
Log of Investment	5.03	1.0	1.13	4.37	5.03	5.71	9.34	39,569
Log of Capital	8.99	1.9	2.09	7.99	9.29	10.29	14.57	39,569
Age of the firm	8.54	3.2	1.00	6.00	9.00	11.00	17.00	39,569

Table 2: Summary Statistics for the Balanced Sample

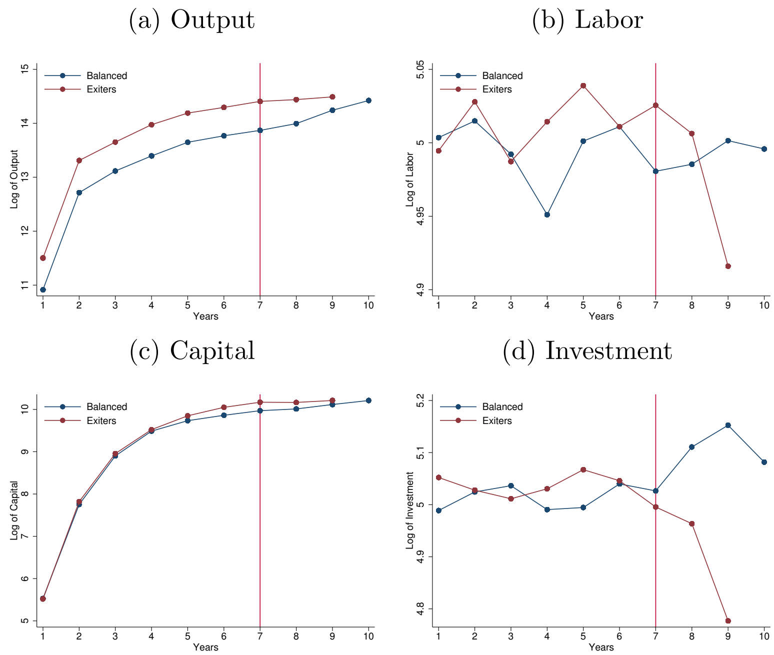
	Mean	SD	Min	Perc. 25	Median	Perc. 75	Max	N
Log of Output	13.41	1.7	5.91	12.36	13.52	14.57	18.87	21,800
Log of Labor	4.99	1.0	1.10	4.32	5.00	5.67	8.86	21,800
Log of Investment	5.04	1.0	1.13	4.37	5.04	5.73	9.34	21,800
Log of Capital	9.16	1.8	2.24	8.26	9.43	10.39	14.34	21,800
Age of the firm	7.32	3.2	1.00	5.00	7.00	10.00	16.00	21,800

Table 3: Summary Statistics for the Exiters Sample

	Mean	SD	Min	Perc. 25	Median	Perc. 75	Max	N
Log of Output	13.59	1.7	6.71	12.51	13.69	14.77	19.16	17,769
Log of Labor	5.01	1.0	0.62	4.34	5.01	5.69	8.60	17,769
Log of Investment	5.02	1.0	1.34	4.37	5.02	5.70	8.87	17,769
Log of Capital	8.78	1.9	2.09	7.66	9.11	10.15	14.57	17,769
Age of the firm	10.03	2.5	1.00	8.00	10.00	12.00	17.00	17,769

Firms that exit the market are, on average, 2.7 years older than the firms that stay in the market. As can be seen in Table 3, the differences in distribution suggests that firms that leave the market have less capital and investment, similar labor, and higher output. Lower capital points toward the fact that exiters had less protection against negative shocks.

Figure 1: Time series by samples



(a) Output (b) Labor (c) Capital (d) Investment

Time series of average by year for both exiteres and firms that stay. Half of the exiters leave the market after year 7.

The fact that exiting firms have higher output is puzzling. However, as can be seen in Figure 1, after year three, exiter firms seem to have tried to compensate for a bad productivity draw by increasing labor and investment. However, this overspending drove them to leave the market.

Question 2

In order to estimate technology from a production function, we need the best possible estimates of the βs in the main regression because everything that we don't accurately account for in terms of labor, capital, and firm age will go in the error which will ultimately be the estimate of technology. We expect the labor coefficient to be positively biased given the simultaneity between a flexible input and output. Regarding capital, we expect it to suffer from both attenuation bias and negative bias. First, because of capital measurement error. Second, because of selection, since we are focused solely on the firms that stay which are the ones that have more capital and therefore have a lower productivity cutoff-level.

The results for the pooled, fixed, random, and between effects are shown in Table 4. As can be seen in column (1), it is the case that the coefficient of labor is larger than the standard 0.3 in the US literature. Likewise, capital seems less that then benchmark of 0.6. However, it is important to note that the labor coefficient doesn't change much across specifications.

In the case of the within estimator (column (3)), we assume that the errors have the form $\varepsilon_{it} = \omega_i + \eta_{it}$. So if, for example, more productive firms hire more people, the fixed effect estimator is a better choice and should alleviate the positive bias for labor. However, it does not look like this is the case. The labor coefficient barely changes from the pooled estimator and the capital coefficient (which should be negatively biased because of selection) even decreases. Finally, only 13% of the estimated variance is due to firm-level fixed effects, which points toward the fact that the within estimator is not addressing our main concerns.

Given that we suspect that the capital coefficient is both negatively biased and attenuated, the between model might be a better choice. The estimator for β_k in the between model (column (2)) is the highest of all the regressions.

Finally, the Hausman test tell us that we can reject the null of the random and fixed effect estimators being statistically equivalent. The necessary asumption for random effects is very implausible in this setting because we precisely expect a non-zero correlation between the observed and unobserved variables. It is then not surprising that random effects estimators are almost the same as the pooled regression.

Table 4: Total, Between, Within and Random Effects Estimators

	(1) Pooled	(2) Between	(3) Within	(4) Random Effects
Age of the firm	0.133***	0.128***	0.188***	0.133***
	(0.005)	(0.006)	(0.006)	(0.006)
Log of Capital	0.431***	0.555***	0.388***	0.421***
	(0.007)	(0.016)	(0.008)	(0.007)
Log of Labor	0.594***	0.613***	0.592***	0.594***
	(0.008)	(0.030)	(0.008)	(0.008)
Observations	21800	21800	21800	21800

Question 3

Taking the difference between time periods and running the regression on the differenced variables would only address our biases if we assume again that $\varepsilon_{it} = \omega_i + \eta_{it}$. In this scenario, the ω_i cancels out when we take the difference. There are two issue here. First, as we discussed above, this assumption on the structure of the errors is probably not right. Second, if the error term in measurement of capital varies in time, it is not canceled out by differencing the variables. Say true capital, k_t^* has measurement error, then $k_{it}^* = k_{it} + \nu_{it}$. As we showed in class, this implies that the error term of the regression is $\eta_{it}^* - \beta_k \nu_{it}$. In this case, when we take first differences we get

$$\begin{aligned}
Cov(\Delta k_{it}, \Delta\varepsilon_{it}) &= Cov\left(\Delta(k_{it}^* + \nu_{it}), \Delta(\eta_{it}^* - \beta_k \nu_{it})\right) \quad \text{where } \Delta k_{it} = k_{it} - k_{ij} \text{ with } j < t \qquad (1)\\
&= -\beta_k Var(\Delta \nu_{it}) = -\beta_k Var(\nu_{it} - \nu_{it-1})\\
&= -2\beta_k Var(\sigma_\nu^2 - \rho_{t,j}) \quad \text{where } \rho_{t,j} \text{ is the serial correlation of } \nu \text{ between time } t \text{ and } j\\
&\Rightarrow \; plim(\hat{\beta}_k) - \beta_k = -\frac{2(\sigma_\nu^2 - \rho_{t,j})}{\sigma_k^2}
\end{aligned}$$

The fact that the coefficient for capital in the pooled regression (Table 4, column 1) is attenuated comes from the fact that the errors include the $-\beta_k \nu_{it}$ term from measurement error, so the estimator is biased towards zero. In column (1) of Table 5 the capital coefficient is even more attenuated now because the measurement error is amplified by the difference estimator. In columns (2)-(4), we can see that the capital coefficient increases as $\rho_{t,j}$ decreases (bigger time difference) because the bias from

Equation 1 is becoming smaller. These results show that running the difference regression, even up to three periods apart, only amplifies the bias from measurement error in capital.

Table 5: Difference Estimators

	(1) First	(2) Second	(3) Third
Age of the firm	0.139***	0.101***	0.183***
	(0.013)	(0.009)	(0.008)
Log of Capital	0.252***	0.372***	0.399***
	(0.011)	(0.009)	(0.009)
Log of Labor	0.593***	0.595***	0.573***
	(0.009)	(0.009)	(0.009)
Observations	15260	15260	15260