Project 1

Linear Panel Data and Production Technology

Econometrics B

Mikkel L. Dollerup ndt926

Emil V. Ramsbæk tdh424

Part 1 xxx

Part 2 xxx

1 Introduction

We are tasked with investigating the assumption that production exhibits constant return to scale by examining the hypothesis on 441 French manufacturing firms using sales, capital, and employment data. To accomplish this, we focus on the first three periods of data and estimate the parameters in a Cobb-Douglas production model. We conduct a Hauseman test to determine whether there are any individual specific effects uncorrelated with the explanatory variables, which will help us identify the most appropriate model for our data.

The conclusion of the Hauseman test in our case indicated that the data did not support constant return to scale. However, this conclusion was reached under several assumptions, and different assumptions could potentially lead to another outcome. To ensure the accuracy of our hypothesis, we will discuss the assumptions that need to be satisfied for our tests to be valid.

In the event that fixed effects (FE) are the more appropriate choice based on the Hauseman test results, we may consider using first differences. On the other hand, if random effects are more suitable, we could also employ pooled OLS. Ultimately, we will select the model that aligns best with the test results and furthers our understanding of the constant return to scale assumption in the context of these French manufacturing firms.

2 Economic model

2.1 Model

$$F(K, L) = y = AK^{\beta_K} L^{\beta_L} \tag{1}$$

In our model y is the total production, A ¿ 0 is the total factor productivity, K is the capital stock, L is labor and β_K, β_L are parameters. For production to show constant return to scale is must satisfy:

$$F(\lambda K, \lambda L) = \lambda F(K, L) \tag{2}$$

Following equation (1):

$$\lambda y = A(\lambda K)^{\beta_K}(\lambda L)^{\beta_L} \tag{3}$$

$$\Leftrightarrow \lambda y = A\lambda^{\beta_K+\beta_L}K^{\beta_K}L^{\beta_L} \tag{4}$$

$$\lambda = \lambda^{\beta_K+\beta_L} \tag{5}$$

This is only true when:

$$\beta_K + \beta_L = 1 \tag{6}$$

So we want to estimate both β_k and β_L and test the hypothesis that they sum to 1, which indicates CRS, to estimate these parameters using linear regression, we want to take the log to get a linear form for the parameters:

$$lny = lnA + \beta_K lnK + \beta_L lnL \tag{7}$$

We have chosen log of the deflated sales to be lny, which means we assume that markets clear and production output is equal to sales. Further more we have chosen the log of employment to be $ln\ L$ and log of adjusted capital stock to be $ln\ K$. Since A is unobserved, we allow for the possibility that it contains time invariant and time varying factors. So our final model is:

$$ldsa_{it} = \beta_K lcap_{it} + \beta_L lemp_{it} + u_{it} + c_i \tag{8}$$

Here $u_{it} + c_i = v_{it}$ is the same as the sum of the unobserved time invariant and time varying production factors, that explain y ($ldsa_{it}$).

2.2 Estimators

We have chosen to use the First Effects (FE) estimator. When selecting an estimator, it is evident that the time-invariant factors are correlated with the outcome variable, which can be empirically observed. Since the error term is correlated, it is, following the assumptions, not possible to use either Pooled OLS (POLS) or Random Effects (RE), as both of these methods would not hold under the assumption $E(v_{it}|x_i) \neq 0$.

The fixed effects (FE) estimator removes the time-invariant effect through time-demeaning and controls for it to ensure that it does not affect the consistency of the estimate. Other estimators used in the analysis include the Fixed Effects (FE), First Differences (FD), and Random Effects (RE), with a Hausmann-test used to test the assumption that $E(c_i|x_i) \neq 0$.

Assumptions required for the FE estimator include:
FE.1: Strict exogeneity in the time-varying error term: $E(u_{it}|x_{i1}, x_{i2}, \ldots, x_{iT}, c_i) = 0$. If not fulfilled, it could be caused by incorrect observations and lead to erroneous models or even data-driven decisions.

FE.2: Rank condition: $rank(\ddot{\mathbf{X}}_\mathbf{i}'\ddot{\mathbf{X}}_\mathbf{i}) = K$. This ensures we have more observations than variables to avoid over-identification. It excludes time-invariant variables since they wouldn't fulfill the rank condition. We also need to have full rank to be able to perform the necessary mathematical operations.

FE.3: Homoskedasticity assumption: $E(\mathbf{u_i}\mathbf{u_i}'|\mathbf{x_i}, c_i) = \sigma_u^2\mathbf{I_T}$. This ensures that all observations across time have the same variance and that there's no serial correlation. Fulfilling this assumption contributes to the consistency and efficiency of the FE estimator.
Although it is impossible to completely rule out endogeneity, the FE estimator is consistent and efficient when these assumptions are met. The variance is estimated, and robust variance estimation is used for all variance estimates except RE to ensure consistent variance estimates in the presence of heteroskedasticity.
In summary, we have opted for the First Effects (FE) estimator as our primary choice due to its ability to control for time-invariant factors and the correlation between the error term and the explanatory variables. By satisfying the necessary assumptions, we can confidently use the FE estimator for our analysis, ensuring consistent and efficient estimates. By comparing the FE estimator with alternatives such as First Differences (FD) and Random Effects (RE), and employing a Hausmann-test to validate our choice, we can increase the robustness of our findings. Consequently, this approach enables us to better understand the constant return to scale assumption in the context of the French manufacturing firms under investigation.

3 Empirical Analysis

The empirical analysis estimates the parameters using the fixed effects (FE) estimation method, resulting in the parameter estimates $\hat{\beta}_K = 0.050$ and $\hat{\beta}_L = 0.600$, with a total of 0.650 (Table 1). The high R^2 values for all the estimation methods used indicate that the linear model fits the data closely.

3.1 Wald Test

To investigate whether the fixed effects estimate of $\hat{\beta}_K + \hat{\beta}_L = 0.650$ significantly departs from the expectation that $\beta_K + \beta_L = 1$, we conduct a Wald test. The Wald statistic formula employed is:

$$w = (R\hat{\beta}_{FE} - R\beta)'(R\text{Avar}(\hat{\beta}_{FE})R')^{-1}(R\hat{\beta}_{FE} - R\beta) \tag{9}$$

Here, $R = [0, 1, 1]$ and $r = R\beta = 1$ represent the restriction and the null hypothesis, respectively. The null hypothesis is $H_0 : \beta_L + \beta_K = 1$, and the alternative hypothesis is $H_A : \beta_L + \beta_K \neq 1$. Since there is only one restriction, the Wald statistic is chi-squared distributed with one degree of freedom. Using a critical value of 3.84 from a chi-squared distribution table, the Wald statistic is calculated for the FE-estimates, resulting in $w = 66.06 > 3.84$. This result leads us to reject the null hypothesis and conclude that $\beta_L + \beta_K \neq 1$ based on our FE-estimates.

3.2 Hausman Test

To determine whether the assumption that $E(c_i|x_i) \neq 0$ is necessary, we compare the efficiency of the fixed effects (FE) estimator to the random effects (RE) estimator using a Hausman test. A significant difference between the estimates would indicate that $E(x_i'c_i) \neq 0$, meaning the RE estimator is more appropriate.
The Hausman test statistic is calculated using the formula:

$$H := (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\hat{\text{Avar}}(d\hat{\beta}_{FE}) - \hat{\text{Avar}}(d\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \tag{10}$$

The null hypothesis is $H_0 : \hat{\beta}_{FE} = \hat{\beta}_{RE}$, and the alternative hypothesis is $H_A : \hat{\beta}_{FE} \neq \hat{\beta}_{RE}$. Since there are $K = 3$ variables, the Hausman statistic is χ^2_3 distributed, with a critical value of 7.815 from a table lookup.
The calculated H value is $50.86 > 7.815$, indicating a significant difference between the estimates. Therefore, we reject the null hypothesis and conclude that $\hat{\beta}_{FE} \neq \hat{\beta}_{RE}$, and $E(c_i|x_i) \neq 0$. This suggests that our decision to use FE estimation was appropriate.

4 Discussion

While the fixed effects (FE) model has its benefits, there are also potential pitfalls to consider. For example, when estimating the separate intercepts for each individual, the FE model subtracts the time-invariant component. This could lead to less precise estimates, particularly when the variation in the independent variables is limited, resulting in fewer degrees of freedom. Additionally, if the model assumes that the time-invariant component is uncorrelated with the independent variables and this assumption is violated, it could lead to biased estimates of the other model parameters.

In contrast, the first-differences (FD) model also has some drawbacks. One of the pitfalls is the loss of information when taking first differences. This occurs because the FD model eliminates time-invariant characteristics of each individual, leading to a loss of information that could reduce the precision of the estimates. Furthermore, the FD model can have problems with endogeneity if the dependent variable and the lagged independent variable are correlated with some of the unobserved factors that change over time. This results in biased estimates of the coefficients. Finally, the non-stationarity assumption in the FD model assumes that the variables are stationary over time, which implies that the mean and variance do not change over time. This assumption may not hold, leading to inaccurate estimates.

Overall, we must carefully consider the assumptions and limitations of each model before deciding on which one to use. While FE and FD models are useful in empirical estimation, they may not always be the best fit for the problem at hand, and other models should also be considered. By carefully considering the pros and cons of each approach, researchers can select the most appropriate model to estimate their parameters and draw meaningful conclusions.

5 Conclusion

Our fixed effects (FE) estimation reveals that production doesn't exhibit constant returns to scale after accounting for time-invariant factors, with results differing from pooled ordinary least squares (POLS) and random effects (RE) estimates (Table 1). A significant portion of capital productivity under RE is due to correlated unobserved factors, leading to FE capital being nearly unproductive. Comparing FE and first-differences (FD) estimation, standard errors are similar, and results don't require reinterpretation. Overall, considering time-invariant unobserved factors is crucial when estimating production functions, as constant returns to scale may not hold. The FE vs. RE comparison offers valuable insights into capital productivity and unobserved factors' impact on productivity variation.

6 Tabels

Table 1: Regression Results

Variable	POLS	FD Estimator	FE Estimator	RE-estimation
Constant	3.84116e-08	-	0.0000 (0.0169)	0.0000 (0.0169)
lemp	0.686033	0.5509 (0.0365)	0.6004 (0.0346)	0.2476 (0.0214)
lcap	0.2768	0.0381 (0.0432)	0.0502 (0.0382)	0.6912 (0.0235)
lcap+lemp	0.96283	0.589	0.6506	0.9388
R-squared	0.217	0.284	0.797	
Sigma-squared	0.013	0.008	0.008	

Table 1

FD regression	Beta	Se	t-values
lcap	0.5509	0.0365	15.0788
lemp	0.0381	0.0432	0.8816
$R^2 = 0.217$			
$\sigma^2 = 0.013$			

Table 2

FE regression	Beta	Se	t-values
lemp	0.6004	0.0282	21.2592
lcap	0.0502	0.0382	1.3143
$R^2 = 0.284$			
$\sigma^2 = 0.005$			

Table 3

RE regression	Beta	Se	t-values
constant	0.0000	0.0196	0.0000
lemp	0.6898	0.0242	28.4772
lcap	0.2377	0.0223	10.6592
$R^2 = 0.797$			
$\sigma^2 = 0.008$			
$\lambda = 0.851$			