KØBENHAVNS UNIVERSITET
DET SAMFUNDSVIDENSKABELIGE FAKULTET

AME, E23

Assignment 1

Linear Panel Data and Production Technology

October 2023

Character count: 12.042 (5 pages)

Division of responsibilities:

xx:

xx:

xx:

Contents

1 Introduction

In this paper, our aim is to assess whether a set of French manufacturing firms operate under constant returns to scale in their production processes. A production function is deemed to exhibit constant returns to scale if it satisfies the condition in (1).

$$F(\lambda K, \lambda L) = \lambda F(K, L) \tag{1}$$

for any $\lambda \geq 0$. We assume that firms produce according to a Cobb-Douglas production function as presented in (2).

$$F(K, L) = AK^{\beta_k} L^{\beta_l} \tag{2}$$

where $A > 0$ is the Total Factor Productivity (TFP), and the inputs are respectively capital (K) and labour (L). β_k and β_l are positive parameters.

Our results provide some evidence against the assumption of constant return to scale. As we discuss, it is essential to consider the limitations of the estimators employed in this study, when considering this conclusion.

2 Data

The provided panel data is on 441 French manufacturing firms in the period 1968-79. Our analysis will be based on a subset of the data which covers the years 1968-1970. The panel is balanced.

The data set contains three variables:

y_{it}: Log of deflated sales

l_{it}: Log of employment

k_{it}: Log of adjusted capital stock

Each variable contains one yearly measurement for each firm. The i:th firm thereby has three measurements of each variable, i.e,. $T = 3$. The deflated sales will be used as the output of firm i's production.

Figure 1 shows the relationship between sales and the inputs of capital and labour. As illustrated, the relationship of the inputs of production, capital and labour, and the deflated sales is overall positive.

3 Econometric theory

By taking the logarithm of the productions function in (2) and including an intercept[1], we set up the relationship in (3) consisting of the variables presented in the data section above.

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + v_{it} \tag{3}$$

where small letters refer to the logarithm of the variable and $v_{it} := \ln A_{it}$. We consider, that the error term consists of possible firm specific time invariant unobservables like geographical placement of the manufacturing firms, industry etc., affecting the deflated sales of the firms. Thus, we rewrite $v_{it} = c_i + u_{it}$ and refer to c_i as unobserved individual heterogeneity and to u_{it} as time varying unobservables. In consistently estimating the model in (3), it is important whether the unobserved individual heterogeneity, c_i, is correlated with the regressors, $x_{it} = (k_{it}, l_{it})$, or not. In the following, we present two different estimators. The Random Effects (RE) estimator, which assumes $E[x'_{it}c_i] = 0$ for consistency, and the Fixed Effects (FE) estimator, that does not require this assumption.[2]

3.1 Consistency

To make best use of the panel structure of the data, we employ the Random Effects (RE) estimator rather than pooled OLS. This involves transforming the data as described in (4).

$$\begin{aligned} y_{it} - \hat{\lambda}\bar{y}_i &= (\boldsymbol{x}_{it} - \hat{\lambda}\bar{\boldsymbol{x}}_i)\boldsymbol{\beta} + (c_i - \hat{\lambda}\bar{c}_i) + (u_{it} - \hat{\lambda}\bar{u}_i) \\ \Rightarrow \breve{y}_{it} &= \breve{\boldsymbol{x}}_{it}\boldsymbol{\beta} + \breve{c}_i + \breve{u}_{it} \end{aligned} \tag{4}$$

[1]The intercept is included to account for the firms having some production even without applying capital and labour, i.e. from fixed costs.

[2]The First Difference estimator (FD) does not require $E[x'_{it}c_i] = 0$ either. However, we have refrained from employing this estimator, as it would entail the loss of an entire time period — a choice we do not find favorable when dealing with a relatively small T value of 3.

where $\hat{\lambda} = 1 - \sqrt{\frac{1}{1+T(\hat{\sigma}_c^2/\hat{\sigma}_u^2)}}$. Thus the RE estimator of $\hat{\beta}$ is defined as in (5).

$$\hat{\boldsymbol{\beta}}_{RE} = \boldsymbol{\beta} + (\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{\breve{x}'_{it}\breve{x}_{it}})^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{\breve{x}'_{it}}\breve{v}_{it}) \tag{5}$$

For this estimator to be consistent, the second addend needs to equal 0 as $N \to \infty$. This requires two assumptions. First, the regressors of the model need to be uncorrelated with the error term, $\breve{v}_{it}$. This imply the following assumption, RE.1a/b:

$$E[u_{it}|x_i, c_i], \quad t = 1, 2, \ldots, T \tag{RE.1a/FE.1}$$

$$E[c_i, x_i] = E[c_i] = 0 \tag{RE.1b}$$

RE.1a does not only imply, that the the time-varying part of the error term, u_{it} is unocorrelated with the unobserved heterogeneity, c_i and regressors, x_{it}, but also implies *strict exogeneity.* This means, that the time varying error term, u_{it}, is unrelated to the explanatory variables, both in the current period and in any past or future periods. RE.1b implies, that also the unobserved heterogeneity, c_i, and the regressors, x_{it} are uncorrelated. Together the assumptions secure that $\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{\breve{x}'_{it}}\breve{v}_{it} \xrightarrow{p} 0$.

Further, the first part of the second addends need to be invertable, implied by the following rank condition:

$$E(\breve{\boldsymbol{X}}_i{}'\breve{\boldsymbol{X}}_i) = rank(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{\breve{x}'_{it}\breve{x}_{it}}) = K \tag{RE.2}$$

RE.2 implies, that the rank of the regressors are full. If this condition is fulfilled, together with RE.1, the estimator is consistent.

However, when estimating the parameters of (3), we must consider whether the unobserved individual heterogeneity, c_i, correlates with the regressors, x_{it}. Such a correlation can emerge from factors like firms employing distinct technologies that influence the utilization of capital and labor inputs in production. In this situation $E[c_i, x_i] \neq 0$ (RE.1b) is not satisfied and RE is inconsistent. The FE estimator does not require RE.1b for consistency.

When estimating with FE, the data is transformed as in (6).

$$
\begin{aligned}
y_{it} - \bar{y}_i &= (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)\boldsymbol{\beta} + (c_i - \bar{c}_i) + (u_{it} - \bar{u}_i) \\
\Leftrightarrow \ddot{y}_{it} &= \ddot{\boldsymbol{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}
\end{aligned}
\tag{6}
$$

As seen from the transformed equation, the time-invariant unobservable is not present anymore. Thus the FE estimator of $\hat{\beta}$ is defined as in (7) and is no longer dependent of c_i.

$$
\hat{\boldsymbol{\beta}}_{FE} = \boldsymbol{\beta} + (\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}'_{\boldsymbol{it}}\ddot{\boldsymbol{x}}_{\boldsymbol{it}})^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}'_{\boldsymbol{it}}\ddot{u}_{it}) \tag{7}
$$

For this estimator to be consistent, we also need FE.1 (similar to RE.1a above) and the rank condition, $E(\ddot{\boldsymbol{X}}_i{}'\ddot{\boldsymbol{X}}_i) = rank(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}'_{\boldsymbol{it}}\ddot{\boldsymbol{x}}_{\boldsymbol{it}}) = K$ (FE.2), to be fulfilled. Thus, when estimating with FE, the assumptions needed are less restrictive as RE.1b is not required because c_i differences out.

3.2 Efficiency

Given the first order assumptions of each estimator stated above it can be shown, that the difference between the estimator and the true value converges to a normal distribution, as shown i (8) and (9).

$$
\sqrt{N}(\hat{\boldsymbol{\beta}}_{RE} - \boldsymbol{\beta}_{RE}) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1}) \tag{8}
$$

where $A := E(\breve{\boldsymbol{X}}'_i\breve{\boldsymbol{X}}_i)$ and $B := E(\breve{\boldsymbol{X}}'_i\boldsymbol{v_i}\boldsymbol{v'_i}\breve{\boldsymbol{X}}_i)$.

$$
\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_{FE}) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1}) \tag{9}
$$

where $A := E(\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i)$ and $B := E(\ddot{\boldsymbol{X}}'_i\boldsymbol{u_i}\boldsymbol{u'_i}\ddot{\boldsymbol{X}}_i)$. The variances of these distributions are said to be asymptotic efficient under assumption $E(\boldsymbol{u_i}\boldsymbol{u'_i}|\boldsymbol{x_i}, c_i) = \sigma_u^2\boldsymbol{I_T}$ (RE.3a/FE.3) and for the RE estimator also $E(c_i^2|\boldsymbol{x}_i) = \sigma_c^2$. These assumptions establish conditions of homoscedasticity and the absence of serial correlation within the error terms. Nonetheless, our investigation led us to reject a test for serial correlation within the errors, denoted as u_{it} (see Appendix A1). Consequently, we employ robust variances, utilizing estimators as outlined in equations (10) and (11).

$$\widehat{Avar\boldsymbol{\beta}}_{RE} = (\breve{\boldsymbol{X}}'\breve{\boldsymbol{X}})^{-1}(\sum_{i=1}^{N}\breve{\boldsymbol{X}}_i'\hat{\breve{\boldsymbol{v}}}_i\hat{\breve{\boldsymbol{v}}}_i'\breve{\boldsymbol{X}}_i)(\breve{\boldsymbol{X}}'\breve{\boldsymbol{X}})^{-1} \tag{10}$$

$$\widehat{Avar\boldsymbol{\beta}}_{FE} = (\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}})^{-1}(\sum_{i=1}^{N}\ddot{\boldsymbol{X}}_i'\hat{\ddot{\boldsymbol{u}}}_i\hat{\ddot{\boldsymbol{u}}}_i'\ddot{\boldsymbol{X}}_i)(\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}})^{-1} \tag{11}$$

where $\hat{\breve{v}}_i$ and $\hat{\ddot{u}}_i$ are the residuals of the respective estimators. These variances are robust to both the serial correlation, and takes potential heteroskedasticity into account.

3.3 Hyphotesis and test statistics

If we have constant returns to scale ($CTRS$), it must hold that $\beta_k + \beta_l = 1$. Therefore, to test whether the firms exhibit $CTRS$, we set up the following hypotheses:

$$\mathcal{H}_0 : \beta_k + \beta_l = 1$$
$$\mathcal{H}_A : \beta_k + \beta_l \neq 1$$

These will be tested for the coefficients of each of our estimators using Wald tests. The Wald test statistic is defined as in (12).

$$W = (R\hat{\beta} - r)'[\widehat{Avar(\hat{\beta})}]^{-1}(R\hat{\beta} - r) \tag{12}$$

Where R is a $Q \times K$ matrix and r is a $Q \times 1$ vector. In this case, the RE estimator has $R_{re} = (0, 1, 1)$ and $r_{re} = (1)$, and the FE estimator has $R_{fe} = (1, 1)$ and $r_{fe} = (1)$. Under the null $R\beta = r$, and the Wald statistic is drawn from a χ^2_1-distribution.

We will compare the FE and RE estimators using a Hausman test. This is done by testing whether the results significantly differ. It has a test statistic defined as in (13).

$$H = (\widehat{\beta}_{\text{FE}} - \widehat{\beta}_{\text{RE}})'[\widehat{Avar}(\widehat{\beta}_{\text{FE}}) - \widehat{Avar}(\widehat{\beta}_{\text{RE}})]^{-1}(\widehat{\beta}_{\text{FE}} - \widehat{\beta}_{\text{RE}}) \tag{13}$$

The null of the Hausman test is that the assumptions for the RE and FE estimators are valid. Since the RE estimator relies on the stricter assumption of $E(x'c_i) = 0$ for consistency, this assumption may not hold if the null is rejected. The test statistic is drawn from a χ^2_K-distribution under the null.

4 Analysis

Table 1 presents the coefficients obtained using both the FE and the RE estimator. As seen in the table, the two estimators yield different results. The FE estimator suggests that a one percent increase in the input of capital (k_{it}) leads to a 0.05 percent increase in sales (y_{it}).[3] The coefficient for labor is significantly larger, indicating that a one percent increase in employment (l_{it}) results in a 0.6 percent increase in sales (y_{it}). In contrast, the RE estimator provides different coefficients: 0.25 percent for capital (k_{it}) and 0.69 percent for employment (l_{it}). Both coefficients are larger than those suggested by the FE model. Additionally, the RE model includes a constant term of 0.00, implying that the model predicts no sales for firms that do not use either of the two input factors.

To determine if the estimates of the two models predict $CRTS$, $\mathcal{H}_0 : \beta_k + \beta_l = 1$, we apply this restriction on both estimators. The Wald test statistic for the FE estimator is $\chi^2 = 38.643$ and for the RE estimator, $\chi^2 = 22.514$. Both with p<0.001. Therefore, when using both estimators, we reject the hypothesis of $CRTS$ in the production function of the French manufacturing firms.

4.1 Tests of consistency

In order to have confidence in the results presented above, it is essential to formally assess the assumption of strict exogeneity, as stated in (RE.1a/FE.1). We estimate the models with leads and lags of x_{it} - i.e. the RE the models are $\breve{y}_{it} = \breve{\boldsymbol{x}}_{it}\boldsymbol{\gamma} + \breve{\boldsymbol{x}}_{it-1}\boldsymbol{\alpha} + \breve{c}_i + \breve{u}_{it}$ and $\breve{y}_{it} = \breve{\boldsymbol{x}}_{it}\boldsymbol{\gamma} + \breve{\boldsymbol{x}}_{it+1}\boldsymbol{\alpha} + \breve{c}_i + \breve{u}_{it}$. The assumption of strict exogeneity would imply that $\boldsymbol{\alpha} = 0$. The results of the estimations are presented in table 2. Upon examining the results, it becomes evident that $\boldsymbol{\alpha} \neq 0$ and we can reject strict exogenity in both the RE and FE model. Consequently, neither models produce consistent estimates of $\boldsymbol{\beta}$.

To assess the quality of the estimators further, we perform the Hausman test. The test statistic

[3] This coefficient is *not* significant on a 5 pct. level.

is $H = 30.980$, with p<0.001. This indicates a significant difference in the results obtained from the two estimators. Consequently, it implies that the assumption of $E[c_i, x_i] = 0$ (RE1.b) does not hold.

5 Discussion and concluding remarks

We conducted estimation using both FE and RE models, and the consistency of both estimators hinges on the assumption of strict exogeneity. Our primary limitation surfaces when, in section 4.1, we empirically reject this assumption. This rejection poses a significant challenge to the validity of all our results. One possible approach to address this issue is through the utilization of instrumental variable (IV) estimation. However, to employ IV estimation, an external instrument is required. This necessity arises from our observation of the significance of x_{it-1}, which in itself violates an assumption of sequential exogeneity. A potential way to deal with the issue would be through the use of IV estimation. An external instrument would be required, since we found x_{it-1} to be significant, which would also violate an assumption of sequential exogeneity.

Despite this issue, we still deem the FE estimator to be somewhat more reliable than the RE estimator. Comparing the two, we need an additional assumption (RE.1b) to be met in order to ensure the consistency of the RE estimator. As mentioned in section 3.1, i.e. firm specific technology can lead to a violation of this assumption. Also, one could consider other examples resulting in $E[c_i, x_i] = E[c_i] = 0$ being violated, like firm work ethic, a firms reputation among clients and talent of workers. This reasoning is supported by the Hausman test, suggesting that using an estimator that eliminates c_i is preferable.

In conclusion, our analysis provides some evidence against the assumption of $CRTS$ in the production function of French manufacturing firms using two different types of estimators. Due to the the endogeneity issues, our results are by no means conclusive since we have not been able to ensure consistent estimates of $\boldsymbol{\beta}$. Due to the reasons outlined above, the FE estimator offers the most potential, but an external instrument is recommended to ensure consistency. Our findings underscore the importance of careful model selection when conducting empirical research.

Tables and Graphs

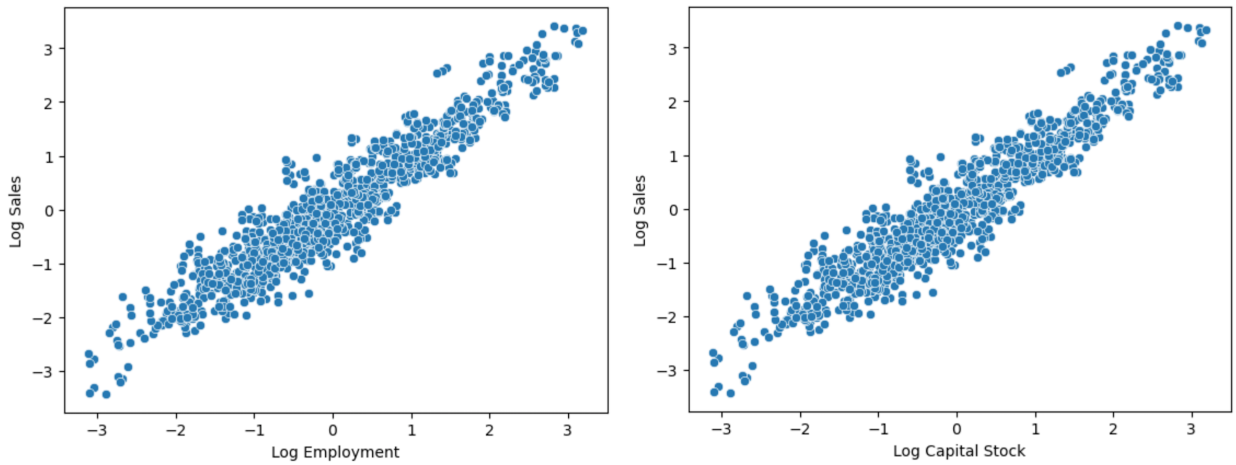

Figure 1: Scatterplots of variables

Table 1: Main results

	log Deflated Sales	
	FE	RE
	(1)	(2)
Constant		0.000
		0.017
log Adjusted Capital	0.050	0.248***
	(0.048)	(0.026)
log Employment	0.600***	0.691***
	(0.050)	(0.030)
Observations	1,323	1,323
R^2	0.284	0.797

*p<0.1; **p<0.05; ***p<0.01. SE are robust.

Table 2: Test of strict exogeneity (RE.1a/FE.1)

	log Deflated Sales (RE)		*log Deflated Sales (FE)*	
	(1)	(2)	(3)	(4)
Constant	0.000	0.000		
	(0.058)	(0.057)		
log Adjusted Capital$_{t-1}$	-0.234**		0.125*	
	(0.123)		(0.089)	
log Employment$_{t-1}$	-0.289***		0.106**	
	(0.088)		(0.057)	
log Adjusted Capital	0.540***	-0.225**	-0.127**	0.058
	(0.126)	(0.122)	(0.076)	(0.075)
log Employment	0.954***	0.217***	0.592***	0.460***
	(0.088)	(0.086)	(0.075)	(0.066)
log Adjusted Capital$_{t+1}$		0.515***		0.066
		(0.124)		(0.084)
log Employment$_{t+1}$		0.469***		0.155***
		0.086		(0.047)
Observations	882	882	882	882
R^2	0.911	0.915	0.248	0.221

*p<0.1; **p<0.05; ***p<0.01. SE are robust.

Appendix

A1. Test of serial correlation (RE.3a/FE.3)

	Dependent variable
	e_{it}
e_{it-1}	-0.185***
	(0.048)
Observations	441
R^2	0.032
*p<0.1; **p<0.05; ***p<0.01.	

The test is cunducted on the erros from a First Difference estimator. Thus, $e_{it} = \Delta u_{it}$. No serial correlation of u_{it} thus implies, that $corr(e_{it}, e_{it-1}) = 0.5$, which we reject, as the significant coefficient on e_{it-1} is -0.36.