Econometrics B: exam Part I

Magnus Goltermann (xzb187),

Thomas Busk-Jepsen (tnr653),

Lasse Strandbygaard (nhs881)

Characters Part I: 8194

Characters Part II: 7201

All parts of the exam are a result of an equal workload

Marts 2023

1 Introduction

This study aims to investigate whether French manufacturing firms exhibit constant returns to scale (CRS), using the Cobb-Douglas production function as a model. We will analyze panel data gathered between 1968 and 1979 to achieve this, however, we will only look at the first three years in this paper. Understanding whether French manufacturing firms exhibit CRS can provide important insights into the efficiency of their production processes. To analyze this, we will use the first-difference estimator. Our finding indicates that the firms in question do not demonstrate CRS; instead, they exhibit decreasing returns to scale. This outcome is reasonable given that constant returns to scale represent a theoretical construct that is seldom realized in empirical reality.

2 Econometric theory

2.1 Model

To analyse the production output, we use the Cobb-Douglas production function as our model (1).

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

Where K denotes the amount of capital, L is the amount of labour and A represents the total factor productivity (TFP), $A > 0$. The pair (β_K, β_L) are the parameters we are interested in analyzing. Since our data has the dependent variable as log-deflated sales and the independent variables as the log of adjusted capital stock and log of employment, we can easily manipulate the equation (1) by taking the log.

$$log(F(K, L)) = log(AK^{\beta_K} L^{\beta_L}) = log(A) + \beta_K log(K) + \beta_L log(L) \tag{2}$$

From the equation (2) we can derive it into a linear econometric model as follows below.

$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + v_{it} \tag{3}$$

Here v_{it} contains TFP, where TFP contains both time-varying and time-invariant unobservable productivity factors, therefore $v_{it} = A_{it} = u_{it} + c_i$. To investigate if the function exhibits CRS, the function has to have the characteristics $F(\lambda K, \lambda L) = \lambda F(K, L)$. This is only the case for a Cobb-Douglas function when $\beta_K + \beta_L = 1$, since

$$F(\lambda K, \lambda L) = A(\lambda K)^{\beta_K}(\lambda L)^{\beta_L} = A\lambda^{\beta_K} K^{\beta_K} \lambda^{\beta_L} L^{\beta_L} = A\lambda^{\beta_K + \beta_L} K^{\beta_K} L^{\beta_L} \tag{4}$$

As a result, our main hypothesis is stated as

$$H_0 : \beta_K + \beta_L = 1 \qquad\qquad H_A : \beta_K + \beta_L \neq 1$$

where we will be testing this using the model given in (3).

2.2 Estimator

To estimate β_K and β_L we will use the first difference estimator (FD). The FD estimator is given by

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\Delta \boldsymbol{X}' \Delta \boldsymbol{X}\right)^{-1} \left(\Delta \boldsymbol{X}' \Delta \mathbf{y}\right) \tag{5}$$

Here $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_K \\ \hat{\beta}_L \end{bmatrix}$ where $\Delta \boldsymbol{X}$ is the matrix representation of $\Delta \boldsymbol{x}$ stacked, that together with $\Delta \mathbf{y}$ can be expressed as the difference between two time periods shown below in (6):

$$\begin{gathered} y_{it} - y_{it-1} = \beta_K \left(k_{it} - k_{it-1}\right) + \beta_L \left(l_{it} - l_{it-1}\right) + \left(u_{it} - u_{it-1}\right) \\ \Longleftrightarrow \\ \Delta y_{it} = \Delta \boldsymbol{x_{it}} \beta + \Delta \boldsymbol{u}_{it} \end{gathered} \tag{6}$$

Note that $\Delta \boldsymbol{x_{it}} = \begin{bmatrix} \Delta k_{it} & \Delta l_{it} \end{bmatrix}$. Therefore we lose the first period due to differencing, and the dimensions of $\Delta \boldsymbol{X}$ is, therefore, $[N(T-1) \times K] = [882 \times 2]$ and $\Delta \boldsymbol{y}$ is $[N(T-1) \times 1] = [882 \times 1]$.

2.2.1 Consistency assumptions

For FD to give us the correct parameter estimate, we need it to be consistent, meaning that we get the true value of β_K and β_L asymptotically. If this is not fulfilled our estimations might deviate from the true values. To ensure consistency of the estimator, we need to satisfy FD.1 and FD.2. According to FD.1, the data should demonstrate strict exogeneity.

$$\text{FD.1: } E(u_{it}|\boldsymbol{x_i}, c_i) = 0,\ t = 1, 2, ..., T \tag{7}$$

The strict exogeneity assumption, FD.1 (7), states that the error term from all time periods must be uncorrelated not only with all independent variables but also with their time-invariant components.

$$\text{FD.2 : } \operatorname{rank} E\left(\Delta\mathbf{X}_i'\Delta\mathbf{X}_i\right) = \operatorname{rank}\left(\sum_{t=2}^{T} E\left(\Delta\mathbf{x}_{it}'\Delta\mathbf{x}_{it}\right)\right) = K \tag{8}$$

FD.2 (8), which is the rank condition, requires that the rank of the expected value of $\Delta\mathbf{X}_i'\Delta\mathbf{X}_i$ must be K to attain full rank. Here, $\Delta\mathbf{X}_i$ has the dimensions $[T-1 \times K]$, and the resulting matrix product has the dimensions $[K \times K]$.

2.2.2 Effenciency assumptions

For the FD estimator to be efficient, the assumption FD.3 must hold.

$$\text{FD.3: } E(\boldsymbol{e}_i\boldsymbol{e}_i'|\boldsymbol{x_i}, c_i) = \sigma_e^2\boldsymbol{I}_{T-1} \tag{9}$$

Where $e_{it} = \Delta u_{it}$ for $t = 2, 3, \ldots, T$. The assumptions require that the error term is uncorrelated with any of the dependent variables or their time-invariant components. Consequently, the error term is homoskedastic. Consistency is a crucial aspect of regression analysis. Therefore, while FD.3 is an important consideration, it should be given secondary priority relative to FD.1 and FD.2.

3 Emperical analysis

To estimate the values of β_K and β_L we have the choice between three different methods namely, fixed effects (FE), random effects (RE), and first differences (FD). We have already excluded Pooled OLS since it would not be logical to assume $E(x_{it}'c_i) = 0$. First, we test our strict exogeneity (FD.1) assumption by including lead variables in the regression and then using a Wald test to see whether the coefficients of the leads are 0. With a Wald test statistic of 11.19 and a p-value of 0.0037, we can see that we do not have strict exogeneity (see table 2). This will be addressed in our discussion.

We can then make a Hausman test to examine if $E[c_i x_{it}] \neq 0$ if this is the case then the RE estimator would be inconsistent. If this is not the case, then the RE estimator is both consistent and more efficient than FE and FD. The Hausman test makes the null and alternative hypotheses of

$$H_0 : E(x_i'c_i) = 0 \qquad H_A : E(x_i'c_i) \neq 0$$

The results of the Hausman test can be seen in table 3. We see that there is strong evidence against the null hypothesis at a 95% confidence level. Hence we reject the null hypothesis ($0.05 > 0.00$), therefore the RE estimator would be inconsistent and we will continue with FE or FD.

To compare FE and FD, which are both consistent under the same assumptions, the efficiency however can differ. We know that FE is more efficient if $cov(u_i, u_{it-1}) = 0$. Our assumption is however that $cov(u_i, u_{it-1}) \neq 0$ as we believe that there is a correlation in the error terms. We have tested this in a serial correlation test this can be seen in table 4. We can also see that the standard error of the FE and FD are almost the same, meaning that they are almost equally efficient and therefore we will choose FD as our estimator, for reasons we will discuss in our discussion. See table 1 for FE and FD regression results. We can now test our main null hypothesis on whether the production function exhibits CRS with a Wald test

$$H_0 : \beta_K + \beta_L = 1 \qquad H_A : \beta_K + \beta_L \neq 1$$

We conducted a Wald test which yielded a test statistic of 95.86 and a corresponding p-value of 0.00. Therefore, if we disregard the violation of the strict exogeneit assumption causing the estimator to not be consistent, we are able to say that French manufacturing firms do not exhibit CRS in their production.

4 Discussion and conclusion

The results that we have derived in our empirical analysis depend on our three assumptions for the fixed effect estimator are valid. From the previous section, we have shown there is no strict exogoneity, which violates our first assumption and causes our estimate to be inconsistent. One could have used other estimator methods such as an FD-IV or GMM for the estimator to be consistent. The advantage of using first differences over fixed effects, when strict exogoneity is not present, is that this estimator only creates correlation within one lag, where FE is more problematic since $\ddot{\mathbf{x}}_{it}$ involves all time periods.

Our objective was to investigate whether French manufacturing firms exhibited CRS, based on an empirical analysis using econometric theory. In order to ensure the validity of our findings, we made the assumption of strict exogeneity. We tested this assumption and found it to be unlikely, however, we continued to use the FD estimator.
If the firms were to exhibit constant return to scale then $\beta_k + \beta_L = 1$. Using FD, we got $\beta_k = 0.550946$ and $\beta_L = 0.0381129$. Using the Wald test from our FD results, we can determine whether the null hypothesis is true. The Wald test statistic (95.86, see table 1) indicated a substantial deviation from CRS. Therefore we can conclude that the firms in question do not exhibit CRS. However keeping in mind that our estimator is not consistent, there is a chance that our findings might not depict the true parameters.

5 Appendix

Tabel 1: FE/FD Regression results

FE regression			FD regression	
Parameter names	Theta_hat	t-values	Theta_hat	t-values
lcap	0.6004 (0.0497)	12.0916	0.5509 (0.0501)	10.9944
lemp	0.0502 (0.0477)	1.0533	0.0381 (0.0437)	0.8728
R-Squared	0.284		0.217	
Sigma-squard	0.008		0.013	
Wald test statistics	133.96		95.86	
p-value	0.00		0.00	
No. of observations	441		441	

Note: Regression results from using the FE and FD estimator with 3 time periods and 441 obersvations. Robust standard errors are given below each estimations in paranthesis.

Tabel 2: Exogeneity test

	β	Se	t-values
lcap	0.4599	0.0550	8.3615
lemp	0.0580	0.0674	0.8617
Labor lead	0.0656	0.0702	0.9342
Capital lead	0.1552	0.0504	3.0768
$R^2 = 0.221$	$\sigma^2 = 0.013$		

$H_0 : \beta_{Laborlead} = \beta_{Capitallead} = 0$
$H_A : \beta_{Laborlead} \neq \beta_{Capitallead} \neq 0$
Wald test statistic: 11.19, p-value: 0.0037

Note: Strict exogeneity test using a lead variable, and using a Wald test to test the null hypothesis.

Tabel 3: Hausman Test

β_{fe}	β_{re}	β_{diff}
0.6004	0.6912	-0.0909
0.0502	0.2476	-0.1974
Hausman test statistic	30.82	
p-value	0.00.	

Note: Fixed effect regression, random effects regression and their difference. Hausman test statistic and its corresponding p-value. 3 time periods and 441 firms where regressed upon.

Tabel 4: Serial Correletation test

	β	Se	t-values
e_{it-1}	-0.1849	0.0483	-3.8295
$R^2 = 0.032$	$\sigma^2 = 0.013$		

Note: Serial correlation test to test the correlation of the error term

Econometrics B: Exam, Part II

Magnus Goltermann (xzb187),

Thomas Busk-Jepsen (tnr653),

Lasse Strandbygaard (nhs881)

Characters Part I: 8194

Characters Part II: 7201

All parts of the exam are a result of an equal workload

Marts 2023

1

To express the success probability $P(y_i|x_i)$ in terms of model parameters, we first substitute y_i with the latent outcome variable $y_i^* > 0$ and its definition

$$P(y_i = 1|x_i) = P(y_i^* > 0|x_i) = P(\beta_0 + \beta_1 x_i + \varepsilon_i > 0|x_i) = P(\varepsilon_i > -\beta_0 - \beta_1 x_i|x_i) \tag{1}$$

Since we know the distribution of $\varepsilon_i|x_i$ is Cauchy distributed, (1) can be rewritten further to include it

$$1 - P(\varepsilon_i < -\beta_0 - \beta_1 x_i|x_i) = 1 - F(-\beta_0 - \beta_1 x_i, \mu) \tag{2}$$

Where $F(z; \mu)$ is the CDF of the Cauchy distribution, (2) can be fully expressed as

$$1 - \left(\frac{1}{2} + \frac{1}{\pi}\text{arctan}(-\beta_0 - \beta_1 x_i - \mu)\right) = \frac{1}{2} - \frac{1}{\pi}\text{arctan}(-\beta_0 - \beta_1 x_i - \mu) \tag{3}$$

2

From equation (1) we had:

$$P(\beta_0 + \beta_1 x_i + \varepsilon_i > 0|x_i)$$

This can be rewritten by multiplying all coefficients and error terms by a positive constant c into:

$$P(\beta_0 \cdot c + \beta_1 x_i \cdot c + \varepsilon_i \cdot c > 0|x_i)$$

Here our choice probability does not change, meaning that we can identify up to a scale factor, but we cannot identify separately from the scale of the error term, therefore we could have $\hat{\beta_0} = c\beta_0$, $\hat{\beta_1} = c\beta_1$ and $\hat{\varepsilon} = c\varepsilon$. Without identification, we cannot estimate consistently, as there is not a unique maximum as the sample objective function can be maximized for different values of c.

3

When given a random sample $\{y_i, x_i\}_{i=1,..,N}$ it is possible to compute the density of y_i conditional on x_i as the following:

$$f(y_i|x_i;\beta) = G(x_i\beta)^{y_i}[1 - G(x_i\beta)]^{(1-y_i)} \tag{4}$$

Since equation (3) is the success probability, we can substitute it for $G(\cdot)$ as

$$\left(\frac{1}{2} + \frac{1}{\pi}\text{arctan}(-\beta_0 - \beta_1 x_i - \mu)\right)^{y_i}\left[1 - \left(\frac{1}{2} + \frac{1}{\pi}\text{arctan}(-\beta_0 - \beta_1 x_i - \mu)\right)\right]^{(1-y_i)} \tag{5}$$

Subsequently, the log-likelihood contribution for observation i becomes the following

$$\ell_i(\beta) = y_i \ln\left[\frac{1}{2} + \frac{1}{\pi}\text{arctan}(-\beta_0 - \beta_1 x_i - \mu)\right] + (1 - y_i)\ln\left[1 - \left(\frac{1}{2} + \frac{1}{\pi}\text{arctan}(-\beta_0 - \beta_1 x_i - \mu)\right)\right] \tag{6}$$

Since $\mu = 0$, the Cauchy distribution is symmetric around zero we have the characteristics $f(z) = f(-z)$, and therefore we can simplify equation (6) further into

$$\ell_i(\beta) = y_i \ln\left(\frac{1}{2} + \frac{1}{\pi}\text{arctan}\left(\beta_0 + \beta_1 x_i\right)\right) + (1 - y_i)\ln\left(\frac{1}{2} - \frac{1}{\pi}\text{arctan}\left(\beta_0 + \beta_1 x_i\right)\right) \tag{7}$$

4

To estimate the parameters, we are using the maximum likelihood estimator (MLE), since we are using a latent variable model. Furthermore, as we already know the distribution of y_i given x_i in (5), and thus the log-likelihood contribution from (7). In contrast to task 2, we now have $\mu = 0$, and therefore a unique set of parameter values gives the maximum likelihood. Therefore the MLE assumptions for consistency are fulfilled, and the estimator will converge asymptotically to the true parameters. Since MLE is an M-estimator, the parameter can be estimated with

$$\hat{\beta}_{MLE} = \arg\min_{\beta} \frac{1}{1000}\sum_{i=1}^{1000} -\ell_i(\beta) \tag{8}$$

Instead of just "brute forcing"to find the arg min, we utilize the derivatives of the likelihood contributions (9), to make our minimizer converge faster to a minimum

$$\nabla_\beta \ell_i(\beta) = \frac{f\left(\beta_0+\beta_1 x_i\right) x_i\left[y_i-F\left(\beta_0+\beta_1 x_i\right)\right]}{F\left(\beta_0+\beta_1 x_i\right)\left(1-F\left(\beta_0+\beta_1 x_i\right)\right)} \tag{9}$$

With F as the CDF and f as the PDF of the Cauchy distribution.
Estimation results are given in table 1, where $\hat{\beta}_0 = 0.9357$ and $\hat{\beta}_1 = 2.6798$.

5

To calculate the partial effect we will simply look at the change in the probability of success when we have discrete changes in the regressor (x_i).

$$PE_{\Delta x}\left(x^0\right) = P\left(y=1|x^0+\Delta x^0\right) - P\left(y=1|x^0\right) \tag{10}$$

We can expand this using equation (3) since that is our conditional probability expanded:

$$\left(\frac{1}{2}-\frac{1}{\pi}\arctan(-\beta_0-\beta_1(x_i+\Delta x_i)-\mu)\right)-\left(\frac{1}{2}-\frac{1}{\pi}\arctan(-\beta_0-\beta_1 x_i-\mu)\right) \tag{11}$$

Rewriting this and taking into account that $\mu = 0$

$$\frac{\arctan(\beta_0+\beta_1(x_i+\Delta x_i))}{\pi}-\frac{\arctan(\beta_0+\beta_1 x_i)}{\pi} \tag{12}$$

Equation (12) is our expression for the partial effects. We can use this equation alongside our MLE parameter estimates from table 1 to calculate the partial effects. As x is a categorical variable with three levels, there are two partial effects to estimate, as the intercept of β_0 describes the probability of success when $x = 0$. Our partial effects can be seen in the table (2).
Thereby we have that the partial effect of going from x=0 to x=1 is 0.174632552. And the partial effect of going from x=1 to x=2 is 0.03574447

6

In order to interpret our PEs, we need standard errors for these estimates, which can be estimated using the delta method. Under the delta method, we assume that our estimator $\hat{\theta}$ has an asymptotically normal distribution. By definition of MLE, we know that this assumption is valid, because of the central limit theorem, and therefore this method is justified. The delta method states then that the transformation of the MLE function of $\hat{\theta}$ also has an asymptotically normal distribution.

The Delta Method computes the standard errors for partial and marginal effects for $\mathbf{h}(\hat{\boldsymbol{\theta}})$ based on an estimated covariance matrix for $\hat{\boldsymbol{\theta}}$. Our partial effects are a function of the estimated parameters, $\mathbf{h}(\hat{\theta})$, which is a K-vector. In order to use the Delta Method we define a $K \times K$ $(= 2 \times 2)$ matrix of derivatives of $\mathbf{h}$ which we call,

$$\mathbf{g} = \nabla_{\theta}\mathbf{h}(\hat{\boldsymbol{\theta}}). \tag{13}$$

Then calculate the following asymptotic variance of $\mathbf{h}(\hat{\boldsymbol{\theta}})$

$$\text{Avar}[\mathbf{h}(\hat{\boldsymbol{\theta}})] = \mathbf{g}\,\text{Avar}(\hat{\boldsymbol{\theta}})\,\mathbf{g}' \tag{14}$$

Where g is defined as

$$\mathbf{g}_k = f(\beta_0, \beta_1)x_{11} - f(\beta_0, \beta_1)x_{10} \tag{15}$$

with $x_{11} = [1, 1]^T$, $x_{10} = [1, 0]^T$, which represents the change in $x_1 = 0 \to 1$, and $f(\cdot)$ is the Cauchy PDF. We use the same method to calculate the standard errors for the PE when $x_1 = 1 \to 2$. To calculate the covariance matrix to estimate (14), we simply just scale the inverse Hessian with $1/N$ we got from our minimization routine during MLE.
From this estimation, we get standard errors of 0.02603 for $PE_{0\to1}$ and of 0.00381 for $PE_{1\to2}$ using the delta method, which can be seen in the table 2. Standard errors can also be estimated using bootstrapping, which gave us almost identical standard errors, again see table 2.

7

To test the hypothesis of x_i has no impact on the probability of success, we could test if $\hat{\beta}_1$ is significantly different from 0. If $\beta_1 = 0$, then x_i will not have any impact on the success (y_i). Therefore we could formulate a null hypothesis of

$$H_0 : \hat{\beta}_1 = 0 \qquad H_A : \hat{\beta}_1 \neq 0$$

Since we only need to know the direction of β_1, we can test the significance of our parameter coefficients directly from our MLE estimation e.g. using a likelihood ratio test, but since we already computed the standard errors for the partial effects in task 6, we can just test the equivalent null hypothesis of them simultaneously being 0.

$$H_0 : PE_{0\to1} = PE_{1\to2} = 0 \qquad H_A : PE_{0\to1} \neq 0 \text{ or } PE_{1\to2} \neq 0$$

To test this null hypothesis we use our t-values from both of our partial effects to control whether or not they are significantly different from 0. From table 2 we can see the t-values. We know that on a significance level of $\alpha = 0.05$ our critical value is ≈ 1.9624 as we have $1000 - 2 = 998$ degrees of freedom. As both of our t-values lies above the critical value, we reject our null hypothesis. We can therefore conclude that our partial effects are significantly different from 0, leading us to the conclusion that x_i has an impact on the probability of success.

Appendix

Tabel 1: Maximum likelihood estimation results

MLE Table		
Parameter estimates	$\hat{\beta}$	t-values
x_0	0.9357 (0.1417)	6.6028
x_1	2.6798 (0.5239)	5.1154
No. observations:	1000	
mean logl	-0.358009	

Notes: Minimize used 17 iterations, 18 function evaluations and 18 evaluations of the jacobian

Tabel 2: PE results

Partial effects				
	PE-value	SE Delta method	SE Bootstrap	t-values
PE1	0.17468	0.02603	0.02659	6.7116
PE2	0.03574	0.00381	0.00389	9.3914
No. of PEs:	2			

Notes: The delta method used the inverse Hessian scaled by $1/N$ as the covariance matrix. The bootstrap method used 1000 samples with repetition. T-value derived from Delta method SE