Assignment 1

Mathias Porsgaard, Mikkel Reich, Noah Carelse

Faculty of Social Sciences
University of Copenhagen

1 Cobb-Douglas production

In this assignment we want to test whether the Cobb-Douglas production function (CD) exhibits constant returns to scale (CRS) at a firm level using panel data on French manufacturing firms with panel data methods. The Cobb-Douglas production function is given by,

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

where K is capital, L is labor, A is total factor productivity (TFP) and β_K, β_L are the output elasticities of capital and labor, respectively. It follows that for CD to have CRS, it must be homogeneous of degree one, i.e. $\beta_K + \beta_L = 1$, as seen from,

$$F(\lambda K, \lambda L) = A(\lambda K)^{\beta_K}(\lambda L)^{\beta_L} = \lambda^{\beta_K+\beta_L} AK^{\beta_K} L^{\beta_L} = \lambda F(K, L) \iff \beta_K + \beta_L = 1.$$

which forms the linear hypothesis we want to test i.e. $H_0 : \beta_K + \beta_L = 1$ vs. $H_A : \beta_K + \beta_L \neq 1$. While K and L are certainly important production inputs, we cannot hope to observe **all** input factors that make up production such as managerial quality or structure i.e. A is unobservable. To investigate the functional form of the production function, we estimate simple linear panel models, to remove the time-invariying (and time-varying unobserved heterogeneity) and assess their merits based on the required assumptions for proper statistical modelling.

We find evidence that only the FD estimator consistently estimates the partial effects of capital and labor, while controlling for time-constant fixed effects, c_i. Furthermore, the French manufacturing firms does not exhibit CRS, but rather decreasing returns to scale.

2 Methodology

We start by taking logs of (1) to linearize it,

$$\ln Y_{it} = \ln A_{it} + \beta_K \ln K_{it} + \beta_L \ln L_{it} \quad \Longrightarrow y_{it} = \beta_K k_{it} + \beta_L l_{it} + \varepsilon_{it}, \quad \varepsilon_{it} = \ln A_{it}$$

where $y_{it} = \ln Y_{it}$ and so on. Suppose that $\ln A_{it}$ can be decomposed into a time-invariant firm-specific effect, c_i and add an idiosyncratic error, u_{it}, such that the composite error is $v_{it} = c_i + u_{it}$. Our model can then be written as:

$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + c_i + u_{it}$$

We write the model on compact form by stacking the regressors in the vector $\boldsymbol{x}_{it} = (k_{it}, l_{it})$ and the parameters in the vector $\boldsymbol{\beta} = (\beta_K, \beta_L)'$.

$$y_{it} = c_i + \boldsymbol{x}_{it}\boldsymbol{\beta} + u_{it}, \quad t = 1, 2, \ldots, 8, \quad i = 1, 2, \ldots, 441 \tag{2}$$

All distributional results are based on fixed T large N asymptotic approximations, synonymous with Micro panel econometric theory in which the cross-sectional dimension is large relative to the time dimension. We consider three estimators: Pooled OLS, Fixed Effects and First-Differences.

Pooled OLS

The pooled OLS (POLS) estimator performs OLS on the entire panel, treating each (i, t) observation as i.i.d. In POLS, identification of $\boldsymbol{\beta}$ requires $\mathbb{E}[\boldsymbol{x}_{it}'v_{it}] = \mathbb{E}[\boldsymbol{x}_{it}'\alpha_i] + \mathbb{E}[\boldsymbol{x}_{it}'u_{it}] = 0$. An example of $\mathbb{E}[\boldsymbol{x}_{it}'v_{it}] \neq 0$ is that if there is some firm-specific effect on the log of sales that correlates with the log of capital or employment, i.e. $\mathbb{E}[\boldsymbol{x}_{it}'c_i] \neq 0$. Since we find this likely, we consider the class of Fixed Effects Methods to alleviate the omitted variable problem in POLS.

Fixed Effects and First Differences

The Fixed Effects (FE) and First Differences (FD) estimators doesn't suffer from same identification issues as the POLS estimator. We show this for FE by performing a within-transformation on (2) and generalize for FD. For each firm i, we subtract the average (over T) dependent variable on the LHS and the average (over T) regressors on the RHS. In this way, the time-invariant TFP α_i cancels out. Define the demeaned variable as $\ddot{y}_{it}, \ddot{\boldsymbol{x}}_{it}, \ddot{u}_{it}$.

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

Identification of $\boldsymbol{\beta}$ is considered. For simplicity, we stack the data over the time-dimension such that $\ddot{\boldsymbol{y}}_i$ and $\ddot{\boldsymbol{u}}_i$ are $T \times 1$ vectors and $\ddot{\boldsymbol{X}}$ is a $T \times K$ matrix. Using the model equation, premultiplying $\ddot{\boldsymbol{X}}_i$, taking expectations and rearranging, we get the following.

$$\Rightarrow \boldsymbol{\beta} = (\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i])^{-1}\left(\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{y}}_i] - \mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i]\right)$$

, i.e. we see that $\boldsymbol{\beta}$ is identified as $\boldsymbol{\beta} = (\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i])^{-1}\mathbb{E}[\ddot{\boldsymbol{X}}_i'\boldsymbol{y}_i]$ if we assume the following:

FE.1, FD.1 Strict exogenity $\mathbb{E}[u_{it}|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2} \ldots, \boldsymbol{x}_{iT}, c_i] = 0 \Rightarrow \mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i] = 0$
The time-varying error term must be exogenous to both log of capital and employment, reflecting no dependence of production shocks on log of capital or employment, neither contemporaneous, leaded or lagged. For FE, this implies $\mathbb{E}[\ddot{u}_{it}|\ddot{\boldsymbol{x}}_{it}] = 0$, while it implies $\mathbb{E}[\Delta\ddot{\boldsymbol{x}}_{it}\Delta u_{it}] = 0$

FE.2, FD.2 (full) rank condition: $\text{rank}(\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i]) = K$ for FE and $\text{rank}(\mathbb{E}[\Delta\boldsymbol{X}_i'\Delta\boldsymbol{X}_i]) = K$
Log of capital and employment can't be linearly dependent.

With FE, estimation of $\boldsymbol{\beta}$ under FE.1 and FE.2 using the analogy principle gives the following.

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{y}}_i \tag{3}$$

, where we emphasize that the former expectation for the population is over the firms i (and not t).

Consistency of Fixed Effects estimator

We evaluate the consistency of the FE-estimator $\hat{\boldsymbol{\beta}}_{FE}$ by inserting $\ddot{\boldsymbol{y}}_i$ in (3), and generalize for the FD-estimator.

$$\Rightarrow \hat{\boldsymbol{\beta}}_{FE} = \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\beta + \ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\right) = \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i\right)$$

Under FE.1 and FE.2, we can apply a Law of Large Numbers (LLN) and Slutsky's theorem (using that the inverse of a matrix is continuous mapping) which shows us that $\hat{\boldsymbol{\beta}}_{FE}$ is consistent for β,

$$p\text{-lim}(\hat{\boldsymbol{\beta}}_{FE}) = \boldsymbol{\beta} + \left(\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i]\right)^{-1}\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i] = \boldsymbol{\beta} + \left(\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i]\right)^{-1} * \mathbf{0} = \boldsymbol{\beta}$$

Asymptotic normality of Fixed Effects estimator

Rearranging (3) and using $\mathbb{E}[\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta})] = \mathbf{0}$ under FE.1 and FE.2, we see that the sum of products of regressors and time-varying errors converge according to the Central Limit Theorem (CLT). The product rule then implies that the product of the two matrices converges (only) in distribution. Again, we generalize for the FD-estimator.

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) = \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i\right) \xrightarrow{d} \text{N}\left(\mathbf{0}, (\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i])^{-1}\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i\ddot{\boldsymbol{u}}_i'\ddot{\boldsymbol{X}}_i](\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i])^{-1}\right)$$

Efficiency

FE.3: Homoskedasticity. For $\hat{\beta}$ to be efficient - have the smallest possible asymptotic variance - it must hold that $\mathbb{E}(\boldsymbol{u}_i\boldsymbol{u}_i'|\boldsymbol{x}_i, c_i) = \mathbb{E}(\boldsymbol{u}_i\boldsymbol{u}_i') = \Omega_u = \sigma_u^2\boldsymbol{I}_T$, i.e. the error term u_{it} is homoscedastic and serially uncorrelated.

FD.3: Similarly for $\hat{\beta}_{FD}$ to be efficient we need $\mathbb{E}(\boldsymbol{e}_i\boldsymbol{e}_i'|\boldsymbol{x}_i, c_i) = \sigma_e^2\boldsymbol{I}_{T-1}$

If any of these assumptions do not hold, as we will see, we can still obtain consistent estimates of the standard errors using robust or clustered standard errors, though often at the cost of efficiency. In which case we estimate the middle part of the sandwich from (**??**) using the analogy principle i.e. using the estimated residuals, such that,

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

which is consistent under *just* **FE.1, FE.2**, and similarly for the FD estimator.

Wald-test

We utilise the Wald-test to test for CRS, where the null-hypothesis is $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$. Under weak regularity conditions, that **FE.1-2** holds and we are using cluster robust standard errors, the Wald statistic can be written as:

$$W := (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}\widehat{\text{Avar}(\hat{\boldsymbol{\beta}}_{FE})}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

The null- and alternative-hypothesis for testing CRS is:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \Leftrightarrow \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}\beta_K \\ \beta_L\end{bmatrix} = 1 \Leftrightarrow \beta_K + \beta_L = 1$$
$$H_A : \beta_K + \beta_L \neq 1$$

If the assumptions for the Wald-test holds we also see the squared t-statistic, t^2, is equivalent to the Wald-statistic: $t^2 = (\frac{\mathbf{R}\hat{\boldsymbol{\beta}}-1}{se(\hat{\boldsymbol{\beta}})})^2$ where $\text{se}(\hat{\boldsymbol{\beta}}) = \sqrt{\text{Avar}(\hat{\beta}_K) + \text{Avar}(\hat{\beta}_L) + 2\text{ACov}(\hat{\beta}_K, \hat{\beta}_L)}$. For a joint test this pattern follows a F-distribution, not a squared t-distribution. This is important because in the strict exogeneity test for the FE and FD models it also tests for joint significance in their leaded terms, where we test for two linear restrictions.

Serial correlation test

We utilise an autoregressive process with one lag (AR(1)) to test for serial correlation in the error term by using an auxilliary regression:

$$\hat{\ddot{u}}_{it} = \rho\hat{\ddot{u}}_{it-1} + \varepsilon_{it}$$

If $\hat{\ddot{u}}_{it-1}$has a significant impact on $\hat{\ddot{u}}_{it}$ it warrants at minimal using cluster robust standard errors. Since time-demeaned errors are by structure serially correlated, they die out under asymptotic properties when $T \to \infty$ for the FE model. This is *not* for the case for the FD model, since $\text{Corr}(\Delta u_{it}, \Delta u_{it-1}) = -0.5$ if FE.3 holds (Wooldridge 2010, Chapter 10.6.3).

Strict exogeneity test

To test for strict exogeneity we test whether any (is it any, or...?) $\boldsymbol{x_{is}}$ is correlated with $\boldsymbol{u}_{it}$ for $s \neq t$. This is most easily done by testing the significance of parameter estimates on leaded explanatory variables, $\boldsymbol{x}_{it+1}$, that is estimate,

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{w}_{it+1}\boldsymbol{\delta} + c_i + u_{it}$$

using FE, where $\boldsymbol{w}_{it+1} \subseteq \boldsymbol{x}_{it+1}$. Under strict exogeneity $\boldsymbol{\delta} = 0$, which is done by a regular t-test. Similarly for the FD model we use $\Delta y_{it} = \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{w}_{it}\boldsymbol{\delta} + \Delta u_{it}$, with $\boldsymbol{w}_{it} \subseteq \boldsymbol{x}_{it}$, following Wooldridge 2010, Chapter 10 - note that the respective $\boldsymbol{\delta}$'s are different, and this slight abuse of notation is done for tabulating convenience.

3 Empirical results

The data consists of $N = 441$ French manufacturing firms over 12 years ($T = 12$) from 1968 to 1979, and is thus a panel data set with $NT = 5292$ observations. The data is balanced.

We estimate the models using the methods outlined in Section 2. The estimates along with robust standard errors are reported in Table 1. Coefficient significance and sign are consistent between methods that is we find a positive partial effect on log deflated sales from increasing any of the inputs conditional on the unobserved heterogeneity, as one would except. With labor increases resulting in a factor ≈ 4.5 and ≈ 8.7 relative sales increase, in percent, respectively.

Table 1: Estimation results

Model	β_L	β_K	(N,T)	R^2	σ^2
FE	0.6942^{***} (0.0417)	0.1546^{***} (0.0299)	(441, 12)	0.477	0.018
FD	0.5487^{***} (0.0292)	0.0630^{***} (0.0232)	(441, 11)	0.165	0.014

Notes: Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

However, for the estimates to carry any meaning, at all, we must test whether assumptions **FE(D).1** hold.

Table 2 reports the results of the autocorrelation tests as described in Section 2. Both models exhibit significant first-order autocorrelation additional to the structurally induced autocorrelation implied by the transformations, and as such we use robust standard errors throughout.

Table 3 reports the results of all strict exogeneity tests as described in the latter part of Section 2.

Table 2: Serial correlation tests on residuals

	FD	FE
Lag residual ($\hat{e}_{it-1}$ / $\hat{\ddot{u}}_{it-1}$)	-0.1987^{***} (0.0148)	0.5316^{***} (0.0123)
t-stat	-13.4493	43.2811

Notes: Each column reports a regression of residuals on their first lag: $\hat{e}_{it} = \rho\hat{e}_{it-1} + v_{it}$ for first-differenced (FD) residuals and $\hat{\ddot{u}}_{it} = \rho\hat{\ddot{u}}_{it-1} + v_{it}$ for fixed-effects (FE) residuals. Robust standard errors in parentheses. $^{***}p < 0.01$.

We test strict exogeneity by adding leads of the regressors. In the FE specifications all subsets of $\ddot{\boldsymbol{x}}_{\boldsymbol{it+1}}$ are significant, whereas none of the $\boldsymbol{x_{it}}$ subsets are significant in the FD specifications (note the distinction between transformed leaded variables and levels in the FE vs FD case). Consequently, we reject strict exogeneity for the FE model and fail to reject it for the FD model. The joint tests test the linear hypothesis

$$\boldsymbol{R\beta} = \boldsymbol{r} \implies= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_L \\ \beta_K \\ \gamma_{L+1} \\ \gamma_{K+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This implies that $\hat{\beta}_{\text{FE}}$ is inconsistent, while $\hat{\beta}_{\text{FD}}$ is consistent under the usual regularity conditions, and for further inference we should rely on the FD model. The full set of results is reported in Table 3.

Table 3: Strict exogeneity tests

	FE_1	FE_2	FE_3	FD_1	FD_2	FD_3
β_L	0.5681^{***}	0.6479^{***}	0.5408^{***}	0.5484^{***}	0.5473^{***}	0.5483^{***}
	(0.0231)	(0.0162)	(0.0431)	(0.0294)	(0.0294)	(0.0293)
β_K	0.1495^{***}	0.0210	0.0280	0.0629^{**}	0.0612^{**}	0.0565^{**}
	(0.0134)	(0.0231)	(0.0375)	(0.0232)	(0.0234)	(0.0241)
δ_L	0.1532^{***}	–	0.1419^{***}	-0.0002	–	0.0045
	(0.0225)		(0.0283)	(0.0011)		(0.0030)
δ_K	–	0.1793^{***}	0.1667^{***}	–	-0.0009	-0.0046^{*}
		(0.0258)	(0.0457)		(0.0009)	(0.0026)
$H_0 : \delta_L = \delta_K = 0$			44.111			3.406
			p=0.000			p=0.182

Notes: Robust standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. "Joint test" reports the Wald statistic for the joint significance of δ_L and δ_K in columns FE_3 and FD_3; the p-value is shown below each statistic using ·. The Wald test is asymptotically $\chi^2(2)$ under the null of two linear restrictions $\delta_L = \delta_K = 0$.

Turning to residual dynamics, both FE and FD residuals exhibit at least first-order serial correlation (see Table 2). We therefore report robust standard errors so that the asymptotic variance estimator remains consistent (though not efficient), allowing valid inference provided **FE(D).1** and **FE(D).2** hold.

Lastly we proceed to test the null hypothesis of constant returns to scale. From Table 4, we reject the null at the 1 % significance level with a p-value of ≈ 0. Thus, we conclude that the (Cobb-Douglas) production function does not exhibit constant returns to scale, for French manufacturing firms in the period 1968-1979.

Table 4: Wald Test Results - Robust

Wald stat	df	$\chi^2_{1(0.95)}$	p-value
19.403	1	3.841	0.000

In addition one might want to test whether the production function exhibits increasing or decreasing returns to scale. This can be done by testing the null hypothesis of $\beta_K + \beta_L \geq 1$ against the alternative of $\beta_K + \beta_L < 1$ (decreasing returns to scale) or vice versa for increasing returns to scale. Importantly we need to consider that the test includes the boundary i.e. $\beta_K + \beta_L = 1$ such that we cant draw a conclusion if the null is accepted, which is why the tests are read somewhat opposite of what one might expect. One might suspect that this particular sector exhibits increasing returns to scale, as firms in manufacturing often benefit from economies of scale, which indeed is the case as seen from Table 5.

Table 5: One-Sided Tests for Returns to Scale

Hypothesis	Estimate	t-stat	p-value	Reject H_0
Increasing RTS ($\beta_K + \beta_L > 1$)	0.849	-11.627	1.000	No
Decreasing RTS ($\beta_K + \beta_L < 1$)	0.849	-11.627	0.000	Yes

References

Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.