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-invarying (and time-varying unobserved heterogeneity) and assess their merits based on the required assumptions for proper statistical modelling.

We find evidence that the (FD does FE doesnt, lets see) 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.

NOTES:

Pooled OLS on transformed model - strict exog, implies POLS.1 fulfilled, $E(\ddot{X}'_{it}\ddot{u}_{it}) = 0$, that is, the transformed X is uncorrelated with the transformed idiosyncratic error. Such that $\hat{\beta}_{FE}$ is unbiased and consistent (along with rank / non-degeneracy condition).

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 is then:

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

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 $\boldsymbol{\beta} = (\beta_K, \beta_L)'$.

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

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

The Fixed Effects (FE) estimator doesn't suffer from same identification issues as the POLS estimator. We show this by performing a within-transformation on (2). 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 $\ddot{y}_{it}, \ddot{\boldsymbol{x}}_{it}, \ddot{u}_{it}$.

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

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 Strict exogenity $\mathbb{E}[u_{it}|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2} \ldots, \boldsymbol{x}_{iT}, \alpha_i] = 0 \Rightarrow \mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i] = 0$
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.

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

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).

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

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

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

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.

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

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.

If 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,

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

which is consistent under *just* **FE.1, FE.2**.

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 - consistency of the FE estimator, asymptotic normality and $\mathbf{R}$ has full rank, . These conditions assumes assumption **FE.1-2** are valid and we are using robust standard errors. The Wald statistic can be written as:

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

When testing for CRS we test the linear hypothesis

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

Under the H_0 the Wald statistic follows a χ^2_Q distribution.

When testing for CRS we test $H_0 : \beta_K + \beta_L = 1$, where we have $Q = 1$ degrees of freedom and $\mathbf{R}$ is a $[1 \times K]$ matrix and $\mathbf{r} = 1 \times 1$ matrix, where K is the rank $K = E[\mathbf{x}'\mathbf{x}]$ from assumption FE.2. Since we only test one linear hypothesis the Walt-statistic is simply the squared t-statistic (might be very overkill to state this mathematically, but in terms of our actual estimates:) The null-hypothesis is given by $H_0 : \mathbf{R}\hat{\boldsymbol{\beta}} = \mathbf{r} \Leftrightarrow \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_K \\ \hat{\beta}_L \end{bmatrix} = 1$. Insert the estimated values for $\hat{\beta}$ and the Walt-statistic yields:

$$W_{robust} = (\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0.154620 \\ 0.694226 \end{bmatrix} - 1)[\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0.001735 & -0.000727 \\ -0.000727 & 0.0008968 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} - 1] \approx 19.402912$$

If the assumptions for the Wald-test holds (assumptions FE.1-2 for W_{robust}), then the squared t-statistic, t^2, is equivalent to the Wald-statistic. The framework for one linear restriction is a little different, but remember the general t-statistic looks like:

$$t = \frac{\hat{\beta} - \beta}{se(\hat{\beta})}$$

The framework for testing a linear restriction when using a t-test is $H_0 : \mathbf{R}\boldsymbol{\beta} = \hat{\mathbf{r}}$, where we estimate $\hat{\mathbf{r}}$ instead of stating what it must be. From the null-hypothesis from the Wald-test we simply state $\mathbf{r} = 1$, but now we try to estimate $\hat{\mathbf{r}}$, and the simple t-test setup looks like the classical t-test framework $H_0 : \hat{\beta} = \beta \Rightarrow$

$$\begin{aligned} H_0 : \hat{\mathbf{r}} = \mathbf{r} \Leftrightarrow \hat{\mathbf{r}} = 1 \Leftrightarrow \\ \mathbf{R}\hat{\boldsymbol{\beta}} = 1 \Leftrightarrow \hat{\beta}_K + \hat{\beta}_L = 1 \end{aligned}$$

We take $\beta \Rightarrow \mathbf{r}$ as given because it is a value the statistician (we) declare $\mathbf{r} = 1$. Insert the framework into the t-statistic:

$$t = \frac{\hat{\beta}_K + \hat{\beta}_L - 1}{se(\hat{\mathbf{r}})} \Leftrightarrow \frac{0.154620 + 0.694226 - 1}{0.001178} \approx -4.404874$$

And t squared $t^2 = (-4.404873)^2 \approx 19.402906 \approx W_{robust} = 19.402912$. Differences is caused by rounding errors *run the code for exact results.*

(*dette burde stå først*:) If we use normal standard errors, we require stronger assumptions where under assumption FE.3 the FE estimator is asymptotically efficient $\rightarrow$ we can use normal standard errors. If we relax assumption FE.3, or simply reject the null-hypothesis from a serial-correlation test, like a placebo test with an AR(1) model, (*see section for serial correlation for this specific test*) we have to use robust standard errors. The only change in the Wald-statistic is the variance-covariance matrix Avar($\hat{\beta}_{FE}$), where the robust variance-covariance matrix is given by the sandwich formula.
As seen under the efficiency section. We need to differentiate between the normal and robust version of Avar(β).

Serial correlation test

In this report we utilise an autoregressive process with one lag (AR(1)) to test for serial correlation in the error term. We use 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 indicates a problem that the errors are correlated across (some) time. This warrants at least using robust standard errors, but since the time-demeaned errors are serially correlated, we have to use robust standard errors for the t-test/Wald-test to be valid. If there is arbitrary serial correlation, the t-statistic is not valid and since $\ddot{u}_{it}, \ddot{u}_{it-1}$ is correlated (*tends to 0 as T grows, so asymptotic properties are still exactly the same as always...)*. So using time-demeaned errors, we test for serial correlation in these errors by making the test robust to arbitrary serial correlation. The t-statistic will be valid, and if it still shows significance, it does indicate a problem. If this is the case we have to use robust standard errors in our model when estimating β_K, β_K, to be able to conduct valid tests.

Strict exogeneity

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 $\Delta y_{it} = \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{w}_{it}\boldsymbol{\gamma} + \Delta u_{it}$, with $\boldsymbol{w}_{it} \subseteq \boldsymbol{x}_{it}$, following **wooldridgeEconometricAnalysisCross**

Homoskedasticity

No

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.1, FE.2** and **FD.1, FD.2** hold.

Table X reports the results of all misspecification tests as described in the latter part of Section 2.

INPUT TABLE.

For the FE model we find that the strict exogeneity assumption is broken in which the $\hat{\beta}_{FE}$ is inconsistent, while the FD model, blah blah. Both models exhibit autocorrelated errors of at least degree one, why we use robust standard errors - such that the asymptotic variance estimator is still consistent, though not efficient, allowing valid inference, provided the **FE(D).1, FE(D).2**.

Next we proceed to test the null hypothesis of constant returns to scale. From Table 2, 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 2: 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 3.

Table 3: 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