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 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 (**??**) to linearize it,

$$\begin{aligned} \ln Y_{it} &= \ln A_{it} + \beta_K \ln K_{it} + \beta_L \ln L_{it} \\ y_{it} &= \beta_K k_{it} + \beta_L l_{it} + v_{it}, \quad v_{it} = \ln A_{it} \end{aligned}$$

where $y_{it} = \ln Y_{it}$ and so on. Suppose that $\ln A_{it}$ can be decomposed into a time-invariant firm-specific effects, α_i and add an idiosyncratic error, u_{it}, i.e. $v_{it} = \alpha_i + u_{it}$. The model is then given by,

$$\begin{aligned} v_{it} &= c_i + u_{it}, \\ y_{it} &= c_i + \beta_K k_{it} + \beta_L l_{it} + u_{it} \end{aligned}$$

Then stack the RHS variables in the vector $\boldsymbol{x}_{it} = (k_{it}, l_{it})$ and the parameters in the vector $\boldsymbol{\beta} = (\beta_K, \beta_L)'$. The model can then be written on compact form as,

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

Pooled OLS

The simplest possible estimator for panel data is the pooled OLS (POLS) estimator, which corresponds to OLS on the entire panel, considering each (i, t) observation as i.i.d. In POLS, an attempt to identify $\boldsymbol{\beta}$ results in the following:

$$\boldsymbol{\beta} = \left(\mathbb{E}[\boldsymbol{x}_{it}'\boldsymbol{x}_{it}]\right)^{-1} \left(\mathbb{E}[\boldsymbol{x}_{it}'y_{it}] - \mathbb{E}[\boldsymbol{x}_{it}'v_{it}]\right)$$

Identification of $\boldsymbol{\beta}$ requires that the composite error u_{it} is uncorrelated with the regressors. Unpacking the composite error such that $\mathbb{E}[\boldsymbol{x}_{it}'v_{it}] = \mathbb{E}[\boldsymbol{x}_{it}'\alpha_i] + \mathbb{E}[\boldsymbol{x}_{it}'u_{it}]$ shows that both the time-varying and time-invariant part of TFP must be assumed uncorrelated with logged capital and logged employment to successfully identify $\boldsymbol{\beta}$.

An example of violation of $\mathbb{E}[\boldsymbol{x}_{it}' v_{it}] = 0$ would be if firms that have more capital (or employees) are generally more productive than firms with less capital (or less employees). This could be due to structural differences a.o. Since we find this highly likely, we proceed to consider another estimator.

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 (**??**). For each firm i, we subtract the average (over T) dependent variable on the LHS and the average (over T) regressors on the RHS.

$$y_{it} - \bar{y}_{it} = \boldsymbol{\beta}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{it}) + (\alpha_i - \alpha_i) + (u_{it} - \bar{u}_{it})$$

In this way, the time-invariant TFP α_i cancels out. Defining demeaned version of the variable v_{it} as $\ddot{v}_{it} = v_{it} - \frac{1}{T}\sum_{t=1}^{T} v_{it} = v_{it} - \bar{v}_{it}$, we get the new model:

$$\ddot{y}_{it} = \boldsymbol{\beta}\ddot{\boldsymbol{x}}_{it} + \ddot{u}_{it}$$

We then consider the identification of $\boldsymbol{\beta}$. Firstly, the data is stacked 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.

$$\begin{aligned}
\ddot{\boldsymbol{y}}_i &= \ddot{\boldsymbol{X}}_i\boldsymbol{\beta} + \ddot{\boldsymbol{u}}_i \\
\Leftrightarrow \ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{y}}_i &= \ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i\boldsymbol{\beta} + \ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i \\
\Leftrightarrow \mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{y}}_i] &= \mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i]\boldsymbol{\beta} + \mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i] \\
\Leftrightarrow \boldsymbol{\beta} &= (\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i])^{-1}\left(\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{y}}_i] - \mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i]\right)
\end{aligned}$$

, i.e. we see that $\boldsymbol{\beta}$ is exactly identified if we assume the following:

FE.1 Strict exogenity $\mathbb{E}[u_{it}|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2} \ldots, \boldsymbol{x}_{iT}, \alpha_i] = 0$, where $\boldsymbol{X}_i$ is the non-demeaned $T \times K$ matrix of regressors in all periods for firm i. This assumption implies that the time-demeaned regressors and error terms are uncorrelated $\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i] = 0$.

FE.2 the matrix $\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i]$ fulfils a (full) rank condition, i.e. $rank(\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i]) = K$ and is therefore invertible. This assumption rules out multicollinearity between demeaned regressors.

Fulfilment of these two assumptions implies that $\boldsymbol{\beta}$ is successfully identified:

$$\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i] = 0 \Rightarrow \boldsymbol{\beta} = (\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i])^{-1}\mathbb{E}[\ddot{\boldsymbol{X}}_i'\boldsymbol{y}_i]$$

Estimation of FE estimator

Using the analogy principle, estimation of the identified $\boldsymbol{\beta}$ under the assumptions of strict exogenity and full rank condition would result in the following estimate $\hat{\boldsymbol{\beta}}_{FE}$. Note that the expectation is over the firms i (and not t).

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

With $\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}}$, we obtain a $(K \times K)$ matrix, which, with a $(K \times 1)$ vector from $\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{y}}_i$ results in a $(K \times 1)$ vector $\boldsymbol{\beta}$.

Consistency of Fixed Effects estimator

Now, we turn to the consistency of the FE-estimator. Inserting $\ddot{\boldsymbol{y}}_i$ in the expression for $\hat{\boldsymbol{\beta}}_{FE}$, we obtain:

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

Under FE.1 and FE.2, we can then we apply a Law of Large Numbers (LLN) such that $p\text{-lim}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right) = \mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i]$ and $p\text{-lim}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\right) = \mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i] = 0$. Then we use Slutskys theorem: $p\text{-lim}\left(\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\right) = (\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i])^{-1}$, which is allowed since the inverse matrix is defined, due to the rank assumption, and the inverse operator is a continuous function.

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

2.0.1 Asymptotic normality of Fixed Effects estimator

For inference of the FE-estimator, the asymptotic distribution is needed. Rearranging (**??**) gives the following:

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

Earlier, we saw that FE.1 and FE.2 combined with LLN and Slutsky's theorem implied that asymptotically $\mathbb{E}[\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta})] = \boldsymbol{0}$. This means that the sum of real-valued i.i.d. r.v.s converges accordingly to the Central Limit Theorem such that $\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\right) \to_d \mathrm{N}(\boldsymbol{0}, \mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i(\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i)']) =^d \mathrm{N}(\boldsymbol{0}, \mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{X}}_i])$.

The product rule implies that the expression for $\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta})$ converges only in the distributional sense such that:

$$\Rightarrow \sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) \sim \mathrm{N}\left(\boldsymbol{0}, (\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i])^{-1}\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{X}}_i](\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i])^{-1}\right)$$

Efficiency

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

It follows, that for the FE estimator to be efficient - i.e. have the smallest possible (asymptotic) variance -, we need to make some assumptions regarding the variance of the idiosyncratic error term, ε_{it}. Specifically, we need that the variance is homoscedastic and not serially correlated, i.e. $\mathrm{Var}(\varepsilon_{it}) = \sigma^2$ and $\mathrm{Cov}(\varepsilon_{it}, \varepsilon_{is}) = 0$ for $t \neq s$, or in the stacked notation, that, $\Omega_{\boldsymbol{\varepsilon}} = \sigma^2 I_T$. This results in the following (ref Mikkels eq for asymptotic normality),

$$\mathbb{E}\left[\ddot{\boldsymbol{X}}'_i\boldsymbol{u}_i\boldsymbol{u}'_i\ddot{\boldsymbol{X}}_i\right] = \mathbb{E}\left[\ddot{\boldsymbol{X}}'_i\mathbb{E}[\boldsymbol{u}_i\boldsymbol{u}'_i|\boldsymbol{x}_i\alpha_i]\ddot{\boldsymbol{X}}_i\right]$$
$$= \mathbb{E}[\ddot{\boldsymbol{X}}'_i\sigma_u^2\boldsymbol{I}_T\ddot{\boldsymbol{X}}_i]$$

such that the asymptotic variance is,

$$\mathrm{Avar}(\hat{\beta}_{FE}) = \sigma_u^2\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i]^{-1}/N$$

Using the analogy principle, we have,

$$\hat{\text{Avar}}(\hat{\beta}_{FE}) = \hat{\sigma}_u^2 \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i \right)^{-1}, \quad \text{where} \quad \hat{\sigma}_u^2 = \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \hat{\boldsymbol{u}}_i' \hat{\boldsymbol{u}}_i,$$

$$\text{and} \quad \hat{\boldsymbol{u}}_i = \ddot{\boldsymbol{y}}_i - \ddot{\boldsymbol{X}}_i \hat{\beta}_{FE}$$

Note that we need to correct the degrees of freedom for the implicit estimation of the fixed effects, which is why we subtract N from the denominator. If these assumptions do not hold, 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 (Mikkel ref normality) using the analogy principle i.e. using the estimated residuals, such that,

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

Which is consistent under just the strict exogeneity assumption as well as the rank assumption - because consistent estimation of the error, as well as the outer matrices being invertible? (remove outer sums possibly and go full stacked like a maniac).

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 (*refer to consistency, and asymptotic normality equations*), the Wald statistic can be written as:

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

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 $H_0 : \hat{\beta} = \beta \Rightarrow$

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

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 → 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,

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

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 following Wooldridge 2010, Chapter 10.

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.

Maybe variable def.

We perform the within transformation of our data matrices as described in the methodology section (Ref) and perform pooled OLS on (ref est eq). The estimates along with standard errors and t-values are reported in Table **??**.

Table 1: Fixed Effects Regression Results

	β	Se	t-values	p-value
ℓ	0.6942	0.0147	47.2447	0.0000
k	0.1546	0.0130	11.9311	0.0000
R^2	0.477			
σ^2	0.018			

Next we proceed to test the null hypothesis of constant returns to scale. From Table **??**, 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

Wald stat	df	$\chi^2_{1(0.95)}$	p-value
135.190	1	3.842	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 **??**.

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

References

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