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 using classical unobserved effect 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. The aim is thus to investigate the functional form of the production function using methods that can account for (some) unobserved heterogeneity. We do this by estimating simple linear panel models, that remove the time-invaryingunobserved heterogeneity and assess their individual merits based on the required assumptions for proper statistical modelling.

We find evidence that the strict exogeneity assumption needed for consistent estimation of the partial effects of capital and labor, while controlling for time-constant fixed effects, c_i, is violated. Due to this endogeneity issue, inference should be interpreted with (a lot of) caution, where we nonetheless find that the French manufacturing firms does not exhibit CRS, but rather decreasing returns to scale.

2 Methodology

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

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

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 $\ln A_{it} = v_{it} = c_i + u_{it}$. 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 $\boldsymbol{\beta} = (\beta_K, \beta_L)'$.

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

We will consider three estimators: Pooled OLS, Fixed Effects and First-Differences.

Note on asymptotics: 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.

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, say the TFP level from our Cobb-Douglas production function on the log of sales, it will correlate with the log of capital and/or employment, i.e. $\mathbb{E}[\boldsymbol{x}_{it}'c_i] \neq 0$. Therefore we consider the class of Unobserved 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, and allow for arbitrary correlation between the regressors and c_i. We show this for FE by performing a within-transformation on (**??**) and similar results hold using the FD operator. 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 c_i cancels out. Define the demeaned variable as $\ddot{y}_{it}, \ddot{\boldsymbol{x}}_{it}, \ddot{u}_{it}$.

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

and analogous for FD with $\Delta y_{it} = \Delta\boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$ where $\Delta\boldsymbol{x}_{it} = \boldsymbol{x}_{it} - \boldsymbol{x}_{it-1}$ and so on. Correct identification is assumed, for unbiased estimation of $\boldsymbol{\beta}$. 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 idiosyncratic 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 $\mathbb{E}[\Delta\ddot{\boldsymbol{x}}_{it}\Delta u_{it}] = 0$ for FD.

FE.2, FD.2 (full) rank condition: $\text{rank}(\mathbb{E}[\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{X}}_i]) = K$ for FE and $\text{rank}(\mathbb{E}[\Delta\boldsymbol{X}_i'\Delta\boldsymbol{X}_i]) = K$ for FD.
The expectation must be invertible, which also implies that 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), and similarly for FD with $\Delta\boldsymbol{X}_i$.

Consistency of Fixed Effects estimator

We evaluate the consistency of the FE-estimator $\hat{\boldsymbol{\beta}}_{FE}$ by inserting $\ddot{\boldsymbol{y}}_i$ in (**??**),

$$
\Rightarrow \hat{\boldsymbol{\beta}}_{FE} = \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\beta + \ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\right) = \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}_i'\ddot{\boldsymbol{u}}_i\right) \tag{4}
$$

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 a continuous mapping) which shows that $\hat{\boldsymbol{\beta}}_{FE}$ is consistent for $\boldsymbol{\beta}$,

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

This is easily shown with an analogous approach for the FD-estimator.

Asymptotic normality of Fixed Effects estimator

Rearranging (**??**) 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.

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) = \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, (\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i])^{-1}\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{u}}_i\ddot{\boldsymbol{u}}'_i\ddot{\boldsymbol{X}}_i](\mathbb{E}[\ddot{\boldsymbol{X}}'_i\ddot{\boldsymbol{X}}_i])^{-1}\right) \tag{5}$$

Efficiency

FE.3: For $\hat{\boldsymbol{\beta}}_{FE}$ 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 or in other words; white noise in levels.

FD.3: Similarly for $\hat{\boldsymbol{\beta}}_{FD}$ to be efficient we need $\mathbb{E}(\boldsymbol{e}_i\boldsymbol{e}'_i|\boldsymbol{x}_i, c_i) = \sigma_e^2\boldsymbol{I}_{T-1}$ which implies that u_{it} is white noise in differences.

Both assumptions cannot hold simultaneously, due to e_{it} having $\mathrm{corr}(e_{it}, e_{it-1}) = -1/2$ under FE.3 in which case FE is efficient. Contrarily, if FE.3 does not hold, FD.3 might, and it might not. In any case, rejecting the hypothesis outlined in the serial correlation section, warrant the use of a cluster robust variance estimator, provided the usual regularity conditions hold.

If any of these assumptions do not hold, we can still obtain consistent estimates of the standard errors using 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,

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

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

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, the Wald statistic can be written as:

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

where $\widehat{\mathrm{Avar}(\hat{\boldsymbol{\beta}}_{FE})}$ is the robust asymptotic variance estimator. If **FE.3** hold the Wald-statistic is the same but we assume u_{it} is homoscedastic and non serially correlated when inserting $\widehat{\mathrm{Avar}(\hat{\boldsymbol{\beta}}_{FE})}$. 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 that 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 $\mathrm{se}(\hat{\boldsymbol{\beta}}) = \sqrt{\mathrm{Avar}(\hat{\beta}_K) + \mathrm{Avar}(\hat{\beta}_L) + 2\mathrm{ACov}(\hat{\beta}_K, \hat{\beta}_L)}$. The Wald test is asymptotically χ^2_Q distributed under the null. The squared t-distribution does not follow the χ^2_Q distribution, unless the degrees of freedom are $Q = 1$. Therefore the t-distribution can not be used for joint tests since with two linear restrictions (or above) the linear hypothesis follows an F-distribution: $W \xrightarrow{d} F(Q, N-K) \xrightarrow{d} \chi^2_Q/Q$, where K is the amount of estimated parameters in $\boldsymbol{\beta}$. This is important because in the strict exogeneity test for the FE and FD models we tests for joint significance in their leaded terms, where we test for two linear restrictions by using the Wald-test.

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:

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

While time-demeaned errors are by structure serially correlated, $\text{corr}(\hat{\ddot{u}}_{it}, \hat{\ddot{u}}_{is}) = -1/(T-1), \forall s \neq t$ it dies out under asymptotic properties when $T \to \infty$. Under **FE.3** we only have to use period $\{T, T-1\}$ because we assume constant σ^2_u across all time periods, but $\sigma^2_{\ddot{u}}$ is structurally correlated so its calculated as $\sigma^2_{\ddot{u}} = \sigma^2_u(1-1/T)$. Therefore any two selected periods we take, will yield the same result if u_{it} is homoscedastic. If we only assume **FE.1-2** and use cluster robust errors in the auxiliary regression, we have to use all periods and use the estimate for the null-hypothesis $H_0 : \rho = -1/(T-1)$, because the errors σ^2_u are not the same on the diagonal covariance matrix. 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). In the FD model, the null is that the differenced errors are white noise, and as such we merely test for significance of ρ.

Strict exogeneity test

To test for strict exogeneity we test whether $\boldsymbol{x_{is}}$ 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 dataset is balanced.

We estimate the models using the methods outlined in Section **??**. Table **??** reports the results of the autocorrelation tests with both models exhibiting significant first-order autocorrelation additional to the mechanically induced autocorrelation implied by the transformation(s), and as such we use cluster robust standard errors - to ensure consistent variance estimation. The estimates along with robust standard errors are reported in Table **??**. 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. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

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^{***} (0.0248)
t-stat	-13.4493	25.1137

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. Standard errors in parentheses, robust in the FE model. $^{***}p < 0.01$. Note that we are testing different hypothesis as per the methodology section, but both that both test statistics are asymptotically normally distributed.

However, for the estimates to carry any meaning, at all, we must test whether assumptions **FE(D).1** hold. We test strict exogeneity by adding leads of the regressors as described in the latter part of Section **??**. 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 as per Wooldridge 2010). Consequently, we would reject strict exogeneity for the FE model and fail to reject it for the FD model. The hypothesis for the joint tests are,

$$\boldsymbol{R\beta} = \boldsymbol{r} \Longrightarrow = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_L \\ \beta_K \\ \delta_L \\ \delta_K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$H_0 : \delta_L = \delta_K = 0 \quad \text{vs} \quad H_A : \text{any } \delta_i \neq 0, i = L, K$$

This suggests that the strict exogeneity assumption on the *population* might not hold, such that neither $\hat{\beta}_{\text{FE}}$ and $\hat{\beta}_{\text{FD}}$ are consistent, even if one finite sample model suggests otherwise which simply could be due to the violation having a certain structure which might get washed out in the transformation(s). The full set of results is reported in Table **??**.

Table 3: Strict exogeneity tests

	FE_1	FE_2	FE_3	FD_1	FD_2	FD_3
β_L	0.5681^{***} (0.0231)	0.6479^{***} (0.0162)	0.5408^{***} (0.0431)	0.5484^{***} (0.0294)	0.5473^{***} (0.0294)	0.5483^{***} (0.0293)
β_K	0.1495^{***} (0.0134)	0.0210 (0.0231)	0.0280 (0.0375)	0.0629^{**} (0.0232)	0.0612^{**} (0.0234)	0.0565^{**} (0.0241)
δ_L	0.1532^{***} (0.0225)	–	0.1419^{***} (0.0283)	-0.0002 (0.0011)	–	0.0045 (0.0030)
δ_K	–	0.1793^{***} (0.0258)	0.1667^{***} (0.0457)	–	-0.0009 (0.0009)	-0.0046^{*} (0.0026)
$H_0 : \delta_L = \delta_K = 0$			44.111 p=0.000			3.406 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 $\underset{\cdot}{\cdot}$. The Wald test is asymptotically $\chi^2(2)$ under the null of two linear restrictions $\delta_L = \delta_K = 0$.

While the above suggest that both estimators are inconsistent and not asymptotically normal, we continue with the main hypothesis test of the assignment as if they were. A proper paper would have to employ different models that can account for the endogeneity problem.

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.

Table 4: Robust Wald Test for CRS in CD

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

Naturally one might then be interested in testing whether the production function exhibits increasing or decreasing returns to scale. This can be done by constructing one-sided tests where we are looking for a rejection of the null, due to the boundary case of CRS being included in the null in which results are inconclusive. To test for decreasing return to scale we test: $H_0 : \beta_K + \beta_L \geq 1$ (increasing or constant RTS) against the alternative of $H_A : \beta_K + \beta_L < 1$ (decreasing RTS) or vice versa for increasing RTS. Since the squared t-statistic yields the same numerical result as the Wald stat in table **??**, the only difference for increasing RTS and decreasing RTS is the *critical value* region, for the t-statistic, have a different sign (+) for increasing and (-) for decreasing. The results are $t = \pm\sqrt{(19.403)} = \pm 4.404$. Inserting the estimates from our FE model yields $t = -4.404$, which illustrates the degree of missing information in not knowing the sign going from Wald to t-test.

For a one-sided test with decreasing RTS, the critical value region is $(-\infty, -1.648]$. Since the t-statistic is within the critical region, we reject the H_0 of constant or increasing returns to scale. As such we cannot reject the H_A of decreasing returns to scale. The H_0 for constant or decreasing return to scale have the opposite critical region $[1.648, \infty)$ where the t-statistic does not lie within. We reach the same conclusion for the FD model, with $t \approx -12.249$. As such the French manufacturing firms exhibit decreasing returns to scale, under the assumption that **FE(D).1** and **FE.2/FD.2** hold.

4 Discussion and conclusion

The most glaring problem of the methods proposed in this paper, is the potential violation of **FE(D).1**, possibly rendering the empirical analysis unreliable at best. A proper paper would dive deeper into this issue, using more advanced models that can account for this endogeneity such as panel IV or GMM.

Another potential issue is the assumption of Cobb-Douglas production technology. It could very well be an appropriate approximation. However, a broader class of production functions such as the non-CRS CES class could account for at least an elasticity of substitution between inputs different from one. Similarly, there are other measurable production inputs that could be accounted for directly - e.g. human capital, though with some uncertainty.

Lastly, while we look at firm-specific fixed effects that are constant over time, it could be very reasonable to assume that there are common temporal macro shocks among other factors that affect French manufacturing firms TFP equally. A model extension including time-dummies (or perhaps Two-Way FE) and/or interaction terms could provide better estimates, and thus conclusions - though it wouldn't alleviate the strict exogeneity problem. Since we find that **FE(D).1**, might be violated, we cannot conclude which of the two models, is best. If the assumption is violated in the original equation before transformation, then it is violated for both FE and FD models. Since the FD model might remove the structural problems when FE.1 is violated, the strict exogeneity test shows a "better" result for the FD model. This does not mean the actual assumption is violated, but we simply can not reject that it isn't. Other/modified models should be considered to alleviate these considerations.

References

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