Project 1: Linear Panel Data and Production Technology

Advanced Microeconometrics, Fall 2025

University Copenhagen

Group 30: Josefin Müller, Arthur Scholz, Inga Stehr

Due Date: Sunday October, 6th 2025 at 22:00

Contents

1 Introduction

Understanding the type of returns to scale is essential for assessing how efficiently firms can expand production. In economics, the Cobb–Douglas production function is commonly used to describe how output depends on the two key inputs of production: capital (K) and labor (L). Assuming a production technology of the form $F(K, L) = AK^{\beta_K} L^{\beta_L}$, the parameters β_K and β_L capture the output elasticities with respect to each input. In this project, we empirically estimate these parameters using panel data for French manufacturing firms and test whether production exhibits constant returns to scale, that is, whether doubling both inputs exactly doubles output ($\beta_K + \beta_L = 1$). After model selection we decide to use a Fixed Effects estimator. We do not find constant returns to scale. However, due to endogenity problems, our estimator is not consistent and the results therefore not reliable.

The dataset consists of a balanced panel of $N = 441$ French manufacturing firms observed over the period 1968–1979. For each firm and year, the data include the log of deflated sales (y_{it}), the log of employment (ℓ_{it}), and the log of adjusted capital stock (k_{it}). Since yearly cross-sectional means are subtracted, variables are expressed relative to the annual average, effectively removing time effects.

2 Econometric Theory

We start from a Cobb–Douglas production function, where $A > 0$ denotes total factor productivity (TFP). Taking logs and defining $y_{it} = lnF(k_{it}, \ell_{it})$ gives:

$$F(k_{it}, \ell_{it}) = A_{it} k_{it}^{\beta_K} \ell_{it}^{\beta_L} \quad \Rightarrow \quad y_{it} = \beta_K k_{it} + \beta_L \ell_{it} + v_{it}, \quad \text{where } v_{it} := \ln(A_{it}). \tag{1}$$

We proceed by decomposing the unobservable TFP component v_{it} into a time-invariant firm-specific effect c_i and an idiosyncratic time-varying component u_{it}: $v_{it} = c_i + u_{it}$.

A key concern is that TFP (v_{it}) is unobserved. If either component is correlated with the regressors, omitted variable bias arises and the parameters cannot be estimated consistently. To address this, we apply different panel data estimators to reduce the influence of unobserved heterogeneity and obtain consistent and efficient estimates of the production parameters. If we suspect $E[\mathbf{x}_{it}' u_{it}] = 0$ we can choose a model with c_i, either a pooled OLS or a Random Effect model. If we suspect $E[\mathbf{x}_{it}' c_i] \neq 0$ we have to use either a Fixed Effect or a First Differences model. Those use transformed data to eliminate c_i.

2.1 Pooled OLS (POLS)

The pooled OLS estimator serves as a natural benchmark for the Cobb–Douglas model before accounting for unobserved heterogeneity. All firm-year observations are stacked into a single regression,

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}, \quad \text{where } v_{it} = c_i + u_{it}. \tag{2}$$

It treats all observations as i.i.d., implicitly assuming that both components of TFP c_i and u_{it} are uncorrelated with the regressors. Under unconditional contemporaneous exogeneity ($E[\mathbf{x}_{it}'v_{it}] = 0, \quad \forall\, t$) and a mild rank condition ($\text{rank}(E[X_i'X_i]) = K$), POLS provides consistent and asymptotically normal estimates of $\boldsymbol{\beta}$.

In production data, the exogeneity assumption is unlikely to hold. The time-invariant component c_i captures persistent firm heterogeneity (e.g. managerial ability), which is typically correlated with long-run input levels. The time-varying component u_{it} reflects shocks (e.g. demand or technology), to which firms adjust capital and labor contemporaneously. Thus, both parts of TFP are systematically related to input choices, violating the conditions for consistency of Pooled OLS.

2.2 Fixed Effects (FE)

The Fixed Effects estimator addresses one key weakness of POLS: the likely correlation between the time-invariant unobserved effect c_i and the regressors. By transforming the data within each firm over time, FE removes all time-invariant components c_i and isolates the within-firm variation,

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}'\boldsymbol{\beta} + \ddot{u}_{it}, \quad \text{where } \ddot{y}_{it} = y_{it} - \bar{y}_i,\ \ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i, \bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}. \tag{3}$$

This "within" transformation eliminates c_i and yields consistent estimates of $\boldsymbol{\beta}$ under the following conditions:

FE.1 (Strict exogeneity): the idiosyncratic error u_{it} is uncorrelated with all past, present, and future regressors conditional on c_i, i.e. $E[u_{it} \mid \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, c_i] = 0$.

FE.2 (Rank condition): the within-transformed regressors have full column rank, $\text{rank}(E[\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i]) = K$.

FE.3 (Efficiency condition): the idiosyncratic errors are conditionally homoskedastic and serially uncorrelated, $E[\mathbf{u}_i\mathbf{u}_i' \mid \mathbf{x}_i, c_i] = \sigma_u^2 I_T$.

Assumptions FE.1–FE.2 ensure identification and consistency, while FE.3 provides asymptotic efficiency. In empirical applications, FE.3 is often unrealistic since production shocks may persist over time. In such cases, robust standard errors are used to obtain valid inference under FE.1–FE.2 alone.

In the context of production, FE estimation is typically preferred over POLS because it controls for unobserved, time-invariant heterogeneity such as managerial ability or firm technology. Compared to FD, FE preserves more information and is less sensitive to measurement error, while unlike RE and POLS, it does not rely on the strong and often unrealistic assumption that firm productivity is uncorrelated with input choices.

2.3 First Differences (FD)

The First Differences estimator provides an alternative transformation that also removes the time-invariant component c_i, but instead of demeaning, it differences each observation over time:

$$\begin{aligned} y_{it} - y_{i,t-1} &= (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})\boldsymbol{\beta} + (c_i - c_i) + (u_{it} - u_{i,t-1}) \\ \Delta y_{it} &= \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, \quad t = 2, \ldots, T \end{aligned} \tag{4}$$

The identification of $\boldsymbol{\beta}$ is based solely on within-firm changes in inputs and output. The FD estimator is thus consistent under strict exogeneity similar to FE.1 and full rank, while its efficiency condition FD.3 is weaker, as it requires uncorrelated errors only in differences rather than in levels.

In the production context, the FD estimator captures short-run productivity dynamics by relating changes in inputs to changes in output. Compared to FE, FD may be less efficient when serial correlation is weak but more robust when shocks have persistent components. However, this is rarely realistic in production data, since firm productivity usually changes gradually over time rather than following a random walk.

2.4 Random Effects (RE)

The Random Effects estimator combines features of the POLS and FE approaches. Instead of eliminating c_i, it models the unobserved firm-specific effect as a random variable drawn from a population distribution that is uncorrelated with the regressors. This allows for a more efficient estimator than FE when the random-effects structure is valid, i.e. that not only the strict exogeneity condition similar to FE.1 holds but additional $E[c_i|\mathbf{x}_i] = 0$.

By exploiting the variance–covariance structure of the composite error $v_{it} = c_i + u_{it}$, the RE estimator quasi–time-demeans the data:

$$y_{it} - \hat{\lambda}\bar{y}_i = (\mathbf{x}_{it} - \hat{\lambda}\bar{\mathbf{x}}_i)\boldsymbol{\beta} + (v_{it} - \hat{\lambda}\bar{v}_i), \quad \text{where } \hat{\lambda} = 1 - \sqrt{\frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + T\hat{\sigma}_c^2}}. \tag{5}$$

Furthermore, identification requires that the quasi–demeaned regressors have full column rank (RE.2). Under the additional assumptions of conditional homoskedasticity (RE.3(a)) and no serial correlation (RE.3(b)) the estimator becomes asymptotically efficient.

In the context of firm-level production, c_i is almost certainly correlated with input levels, making the RE assumptions unrealistic. Hence, FE typically serves as the preferred estimator, while RE may still be useful for efficiency when the exogeneity assumption is plausible.

2.5 Specification Tests - Hausman Test

The Hausman test evaluates whether the key restriction for the RE estimator, $E[c_i|\mathbf{x}_i] = 0$, is satisfied, effectively serving as a model selection test between RE and FE. The test compares the two estimators:

$$H = (\hat{\boldsymbol{\beta}}_{\boldsymbol{FE}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{RE}})' \left[\widehat{\text{Avar}(\hat{\boldsymbol{\beta}}_{\boldsymbol{FE}})} - \widehat{\text{Avar}(\hat{\boldsymbol{\beta}}_{\boldsymbol{RE}})} \right]^{-1} (\hat{\boldsymbol{\beta}}_{\boldsymbol{FE}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{RE}}) \tag{6}$$

which under the null hypothesis is asymptotically distributed as χ^2_K, where K is the number of regressors. We test:

$$\begin{aligned} H_0 &: E[c_i \mathbf{x}_{it}] = 0 \quad \Rightarrow \text{ RE consistent and efficient,} \\ H_1 &: E[c_i \mathbf{x}_{it}] \neq 0 \quad \Rightarrow \text{ RE inconsistent, FE preferred.} \end{aligned} \tag{7}$$

2.6 Constant returns to scale

A production function $F(K, L)$ is said to be homogeneous of degree x if, for any scalar $\lambda > 0$,

$$F(\lambda K, \lambda L) = \lambda^x F(K, L). \tag{8}$$

The parameter x determines the type of returns to scale: decreasing ($x < 1$), constant ($x = 1$), or increasing ($x > 1$).

In the Cobb-Douglas equation the degree of homogeneity is given by $\beta_K + \beta_L$. Under constant returns to scale, this sum equals one. Hence, we test:

$$H_0 : \beta_K + \beta_L = 1 \quad \text{vs.} \quad H_1 : \beta_K + \beta_L \neq 1, \tag{9}$$

where a rejection of H_0 indicates either increasing ($\beta_K + \beta_L > 1$) or decreasing ($\beta_K + \beta_L < 1$) returns to scale. The restriction is evaluated using a Wald test.

3 Empirical Analysis

We begin with a Pooled OLS regression as a baseline and use robust standard errors to account for within-firm autocorrelation over time. While the results display high explanatory power and significant estimators, unobserved firm heterogeneity is likely correlated with input choices. Therefore the POLS estimator is not reliable for causal inference as it is violating the exogeneity condition POLS.1.

Next, we estimate the Fixed Effects (FE) and First-Difference (FD) models, which both remove time-invariant unobserved heterogeneity. The FE estimator, exploiting all within-firm variation, produces stable and significant coefficients for labor ($\beta_L = 0.6942$) and capital ($\beta_K = 0.1546$). The FD estimator, in contrast, relies only on year-to-year changes and generates significant but noisier results.

We examine the residual structure with a serial correlation test, which rejects the null of no autocorrelation. This is consistent with the idea that productivity shocks exist persistently across years. While this result is in favor of FD we decide to rather stick to FE and use robust variance estimators, as then FE remains consistent even with serial correlation and is more efficient by preserving within-firm variation.

We then turn to the Random Effects (RE) model. RE are more efficient than FE if the assumption $E[c_i|\mathbf{x}_i] = 0$ is valid. However, the Hausman test strongly rejects this restriction (p-value: < 0.001). This result is consistent with the economically based idea that persistent firm characteristics, such as time-invariant long-run productivity differences, are systematically related to input choices in capital and labor. This makes the RE estimator inconsistent and leaves FE as the preferred specification for our analysis.

Turning to the main research question of constant returns to scale (CRS), we perform a Wald test on the restriction $\beta_K + \beta_L = 1$ as described in (9). The test strongly rejects the CRS hypothesis, with a test statistic of 19.40 and a p-value < 0.001. This implies that output does not scale proportionally with capital and labor inputs.

Finally, we test the strict exogeneity assumption of the FE model by including future values of inputs in a Lead-Test. The lead terms for capital and labor are all strongly significant (p-value: < 0.001), indicating that firms adjust input decisions (capital and labor) in anticipation of future TFP shocks. This violation implies that, even after adding fixed effects, our input coefficients remain inconsistent and should therefore be interpreted with caution.

4 Discussion and Conclusion

In summary, our empirical analysis of the Cobb-Douglas production function finds that POLS is inconsistent but provides a baseline for comparison. FE dominates FD when robust standard errors are used to correct for serial correlation. The Hausman test rejects RE, confirming that FE is the preferred estimator. Using FE, we do not find evidence of constant returns to scale, showing that the sum of the output elasticities with respect to labor and capital does not equal one. However, since strict exogeneity is violated, the estimates are inconsistent. This underscores the limitations of conventional panel estimators and points to the need for more advanced approaches such as IV or structural methods for credible estimation and inference.

5 Tables and Figures

Model	Variable	Beta	SE
Pooled OLS	Intercept	0.0000	0.0161
	lemp	0.6748***	0.0366
	lcap	0.3100***	0.0324
	R^2	0.914	
	σ^2	0.131	
Fixed Effects	lemp	0.6942***	0.0417
	lcap	0.1546***	0.0299
	R^2	0.477	
	σ^2	0.018	
First Difference	lemp	0.5487***	0.0292
	lcap	0.0630***	0.0232
	R^2	0.165	
	σ^2	0.014	
Random Effects	Intercept	0.0000	0.0168
	lemp	0.7201***	0.0333
	lcap	0.2002***	0.0260
	R^2	0.647	
	σ^2	0.018	

Table 1: Estimation results for Cobb–Douglas production function using different panel estimators. Significance levels: *p<0.1, **p<0.05, ***p<0.01.

Test	Test Statistic	df	p-value	5% Critical Value
Hausman (FE vs. RE)	10.57	2	0.0051	5.99
CRS Wald ($\beta_K + \beta_L = 1$)	19.40	1	< 0.001	3.84
Strict exogeneity (lead test)	88.59	2	< 0.001	5.99

Table 2: Model Specification and Hypothesis Tests based on χ^2

Variable	Coefficient	SE
$e_{i,t-1}$	0.4727	0.0126
R^2	0.227	
σ^2	0.012	

Table 3: Serial Correlation in OLS Residuals

Model	Condition 1	Condition 2	Condition 3
POLS	**POLS.1:** $E[\mathbf{x}_{\mathbf{it}}' v_{it}] = 0\ \forall t$. $\Rightarrow$ contemporaneous exogeneity: $E[u_{it} \vert \mathbf{x}_{\mathbf{it}}] = 0\ \forall t$ and $E[c_i \vert \mathbf{x}_{\mathbf{i}}] = 0$.	**POLS.2:** $\operatorname{rank}\big(E[\mathbf{X}_{\mathbf{i}}'\mathbf{X}_{\mathbf{i}}]\big) = K$	–
FE	**FE.1:** Strict exogeneity: $E[u_{it} \mid \mathbf{x}_{\mathbf{i1}}, \dots, \mathbf{x}_{\mathbf{iT}}, c_i] = 0\ \forall t$. $\Rightarrow$ conditional homoskedasticity and no serial correlation.	**FE.2:** $\operatorname{rank}\big(E[\ddot{\mathbf{X}}_{\mathbf{i}}'\ddot{\mathbf{X}}_{\mathbf{i}}]\big) = K$	**FE.3:** $E[\mathbf{u}_{\mathbf{i}}\mathbf{u}_{\mathbf{i}}' \mid \mathbf{x}_{\mathbf{i}}, c_i] = \sigma_u^2 \mathbf{I}_{\mathbf{T}}$
FD	**FD.1:** $E[\mathbf{u}_{\mathbf{it}} \mid \mathbf{x}_{\mathbf{i}}, c_i] = 0\ \forall t$. with $e_{it} = \Delta u_{it},\ t = 2, 3, \dots, T$.	**FD.2:** $\operatorname{rank}\big(E[\Delta\mathbf{X}_{\mathbf{i}}'\Delta\mathbf{X}_{\mathbf{i}}]\big) = K$	**FD.3:** $E[\mathbf{e}_{\mathbf{i}}\mathbf{e}_{\mathbf{i}}' \mid \mathbf{x}_{\mathbf{i}}, c_i] = \sigma_e^2 \mathbf{I}_{\mathbf{T-1}}$
RE	**RE.1:** (a) $E[u_{it} \mid \mathbf{x}_{\mathbf{i}}, c_i] = 0\ \forall t$; (b) $E[c_i \mid \mathbf{x}_{\mathbf{i}}] = E[c_i] = 0$. $\Rightarrow$ homoskedasticity and no serial correlation.	**RE.2:** $\operatorname{rank}\big(E[\mathbf{X}_{\mathbf{i}}'\boldsymbol{\Omega}^{-1}\mathbf{X}_{\mathbf{i}}]\big) = K$	**RE.3:** (a) $E[\mathbf{u}_{\mathbf{i}}\mathbf{u}_{\mathbf{i}}' \mid \mathbf{x}_{\mathbf{i}}, c_i] = \sigma_u^2 \mathbf{I}_{\mathbf{T}}$; (b) $E[c_i^2 \mid \mathbf{x}_{\mathbf{i}}] = \sigma_c^2$.

Table 4: Model Conditions for POLS, FE, FD, and RE