Advanced Microeconometrics: Project 1

Konstantinos Katharakis
Victor Yiming Li
Jeronya Mbiatat Petnkeu

October 5th, 2025

1 Introduction

This paper investigates whether French manufacturing firms operate under constant returns to scale (CRS), that is, whether the sum of the output elasticities of capital and labour equals one. Using panel data on 441 firms from 1968 to 1979, we estimate a Cobb–Douglas production function of the form

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

employing a Fixed Effects (FE) model. The dataset includes the log of deflated sales (y_{it}), the log of employment (l_{it}), and the log of adjusted capital stock (k_{it}). Our FE estimates indicate that a 1% increase in employment raises output by 0.69%, while a 1% increase in capital raises output by 0.15%. We reject the hypothesis of constant returns to scale (as $\beta_L + \beta_K < 1$). This suggests the presence of decreasing returns to scale.

2 Econometric Theory

2.1 General Setup

We start off by specifying the log-linearized Cobb–Douglas production function in equation (2), with $v_{it} = \ln A_{it}$ capturing the logarithm of total factor productivity (TFP).

$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + v_{it}. \tag{2}$$

Note that TFP is inherently unobservable in the data. We impose that TFP consists of a time-invariant productivity factor c_i, such as managerial talent, and a time-varying element u_{it}, for example, power outages. Defining $\mathbf{x}_{it}$ as the row vector of the log levels of capital and employment, and $\boldsymbol{\beta}$ as the column vector of coefficients associated with capital and employment, β_K and β_L, we can write equation (3) as the following unobserved model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}. \tag{3}$$

We assume throughout that the underlying population model is linear in parameters and that observations are independent and identically distributed across firms.

2.2 The role of c_i

The specification of our econometric model critically depends on whether we believe that the time-invariant productivity factor c_i is uncorrelated with the log levels of capital and labor in every time period t, formally:

$$\mathbb{E}(\mathbf{x}_{it}' c_i) = \mathbf{0}, \qquad t = 1, 2, \ldots, T \tag{4}$$

where $\mathbf{0}$ denotes a 2×1 zero vector. If we believe equation (4) holds, then we can consider using Random Effects Methods, such as Pooled OLS or Random Effects. [1] We believe that it is unlikely that the amount of capital and labor used in production is uncorrelated with the time-invariant productivity factor c_i. For instance, more talented managers can make more optimal labor and capital input choices, which would invalidate $\mathbb{E}(\mathbf{x}_{it}' c_i) = \mathbf{0}$. Since managerial quality is unobserved, failing to control for it introduces omitted variable bias in both Pooled OLS and Random Effects and thus, will lead to inefficient estimators.

2.3 The Fixed Effects Model

The FE model addresses this correlation by transforming the data through within-group demeaning. Subtracting the unit-specific time-average from the dependent variable, regressors, and error yields the model in (5), where c_i drops out across all observations, as it is constant over time. By removing c_i from the model, arbitrary correlation between c_i and the regressors is now permitted. Defining $\ddot{y}_{it} = y_{it} - \bar{y}_i$, $\ddot{k}_{it} = k_{it} - \bar{k}_i$, $\ddot{l}_{it} = l_{it} - \bar{l}_i$, $\ddot{u}_{it} = u_{it} - \bar{u}_i$, and $\ddot{\mathbf{x}}_{it}$ as the row vector of time-demeaned regressors $\begin{bmatrix} \ddot{k}_{it} & \ddot{l}_{it} \end{bmatrix}$, we write the model we want to estimate in (6).

$$y_{it} - \bar{y}_i = \beta_K (k_{it} - \bar{k}_i) + \beta_L (l_{it} - \bar{l}_i) + (c_i - \bar{c}_i) + (u_{it} - \bar{u}_i) \tag{5}$$

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it} \boldsymbol{\beta} + \ddot{u}_{it} \tag{6}$$

For the Fixed Effects model to yield an estimator that is identified and consistent, we have to impose some assumptions:

$$\textbf{FE.1:} \quad \mathbb{E}(u_{it} \,|\, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i) = 0, \quad t = 1, \ldots, T. \tag{7}$$

$$\textbf{FE.2:} \quad \operatorname{rank}\left(\sum_{t=1}^{T} \mathbb{E}(\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}) \right) = K \tag{8}$$

FE.1 assumes that the idiosyncratic component of TFP has a zero conditional mean given the firm's capital and labor inputs across all time periods, as well as the time-invariant firm-

[1] Random Effects actually requires a stronger form of exogeneity: $\mathbb{E}(c_i \mid \mathbf{x}_i) = \mathbb{E}(c_i) = 0$. However, using the Law of Iterated Expectations, one can show that this stronger form of convergence implies that equation (4) also holds: $\mathbb{E}(\mathbf{x}_{it}' c_i) = \mathbb{E}[\mathbf{x}_{it}' \, \mathbb{E}(c_i \mid \mathbf{x}_i)] = \mathbb{E}[\mathbf{x}_{it}' \cdot 0] = \mathbf{0}$

specific effect c_i. We further elaborate on this assumption in the following section. FE.2 imposes that the rank of the matrix of the time-demeaned regressors equals the number of regressors, which in our case are two. Capital and labor choices are likely to vary from period to period, so this is expected to hold, but we confirm it with a rank test. If we further impose an assumption on the variance structure of the time-variant part of TFP:

$$\textbf{FE.3:} \quad \mathbb{E}(\mathbf{u}_i \mathbf{u}_i' \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i) = \sigma_u^2 \mathbf{I}_T, \tag{9}$$

where $\mathbf{u}_i = [\, u_{i1} \;\; u_{i2} \;\; \ldots \;\; u_{iT} \,]'$ is the $T \times 1$ vector of idiosyncratic errors, then the Fixed Effects model yields an asymptotically efficient estimator. FE.3 implies that the error of the time-variant part of TFP is homoskedastic and that there is no serial correlation. In this setting, the homoskedasticity assumption is questionable, as the variance of TFP-related shocks, such as supply-chain disruptions, is unlikely to remain constant over time.

2.4 The Fixed Effects Estimator

Under FE.1 and FE.2, we can estimate β_K and β_L using

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{y}_{it} \right). \tag{10}$$

Under the same assumptions, the estimator satisfies a $\sqrt{N}$-asymptotic normal distribution:

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\big(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\big), \tag{11}$$

where $\mathbf{A} = \mathbb{E}[\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i], \mathbf{B} = \mathbb{E}[\ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i \ddot{\mathbf{u}}_i' \ddot{\mathbf{X}}_i]$ and $\ddot{\mathbf{X}}_i$ is the $T \times K$ matrix stacking $\mathbf{x}_{it}$ over t. Since we believe that FE.3 is unlikely to hold, we rely on the robust variance estimator, which enables valid inference even when FE.3 is violated:

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{FE}) = (\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1} \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \hat{\ddot{\mathbf{u}}}_i \hat{\ddot{\mathbf{u}}}_i' \ddot{\mathbf{X}}_i \right) (\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1}, \tag{12}$$

,
where $\hat{\ddot{\mathbf{u}}}_i = [\, \hat{\ddot{u}}_{i1} \;\; \hat{\ddot{u}}_{i2} \;\; \ldots \;\; \hat{\ddot{u}}_{iT} \,]'$ and $\hat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}' \hat{\boldsymbol{\beta}}_{FE}$.

2.5 Wald Test

Based on the estimated coefficients $\hat{\boldsymbol{\beta}}_{FE}$, we conduct a Wald test of the following hypotheses to assess whether production exhibits constant returns to scale (CRS):

$$H_0 : \beta_K + \beta_L = 1 \quad \text{vs.} \quad H_1 : \beta_K + \beta_L \neq 1 \tag{13}$$

The estimated asymptotic variance in (12) provides the basis for conducting the Wald test. The test statistic takes the form:

$$W = (\mathbf{R}\hat{\boldsymbol{\beta}}_{FE} - r)' \left[\mathbf{R}\,\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{FE})\,\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}}_{FE} - r), \tag{14}$$

where based on our hypothesis test specified in (13), $\mathbf{R} = [1 \quad 1]$ and $r = 1$. The Wald statistic follows a χ^2_Q distribution under H_0. Thus, we reject H_0 at significance level α if W exceeds the $(1-\alpha)$-quantile of χ^2_Q, where Q denotes the number of linear restrictions. In our case, $Q = 1$, reflecting the single linear restriction.

3 Empirical Results

3.1 Model Selection

To assess potential model misspecification, we analysed the data using four different estimators: Pooled OLS, Fixed Effects (FE), Random Effects (RE), and First Differences (FD). As discussed in section 2.2, Pooled OLS ignores unobserved firm-specific effects and thus is likely biased and not suitable for causal inference. To determine whether the Random Effects estimator is preferable to the Fixed Effects estimator, we perform a Hausman test. The test's null hypothesis states that both FE and RE are consistent, but that RE is more efficient. Our test yields a statistic of 24.38 (see Table 2) with a p-value far smaller than 0.001, leading us to reject the null and conclude that the RE estimator is inconsistent.

The FD estimator relies on the same exogeneity assumption as FE (and a mild rank condition) and yields a consistent estimator if FE.1 holds. The choice between FE and FD depends on whether the idiosyncratic component of TFP is serially correlated. If the errors follow a random walk, $u_{it} = u_{i,t-1} + \varepsilon_{it}$, the FD estimator becomes more efficient by removing most of this serial correlation. We consider it unlikely that such serial correlation exists, as some idiosyncratic TFP shocks, such as strikes, supply-chain disruptions, or power outages, are typically short-lived rather than permanently persistent, making FE the preferred choice. In a more detailed analysis, this assumption should be explicitly examined.

3.2 Empirical Results

The regression results (Table 1) show that output was more responsive to labour than to capital, with estimated elasticities of 0.69 and 0.15 respectively in the preferred FE model. The combined elasticity of 0.84 indicates decreasing returns to scale, suggesting that expanding production was less efficient. We used the Wald test to test whether the coefficients are statistically different from zero. Table 2 displays strong statistical evidence that the sum of capital and employment elasticities is significantly different from 1. These findings depart from the expected constant returns to scale under standard theory and may reflect the labour-intensive structure and inefficiencies of the French manufacturing sector during the period.

3.3 Assumptions and Limitations

Our findings crucially depend on the strict exogeneity assumption (FE.1). Yet, this assumption may not hold, since firms might adjust input decisions in response to expected future productivity shocks, for example, in anticipation of rising demand. In a more comprehensive setting, this assumption should therefore be explicitly tested. We nevertheless rely on the Fixed Effects model, as we think it most effectively controls for time-invariant firm heterogeneity. Since other linear models such as Random Effects and First Differences share the same exogeneity requirement, a violation of FE.1 would motivate the use of more advanced estimation approaches.

4 Discussion and Conclusion

The analysis confirms that French manufacturing firms during 1968–1979 operated under decreasing returns to scale, with labour exerting a stronger influence on output than capital. This implies that firms faced rising costs as they expanded and often operated below an efficient scale. Inefficiencies were likely amplified by factors such as labour quality, availability, and disruptions like strikes, which could halt production while fixed costs continued to accrue. External shocks, including the 1973 oil crisis (Schumacher, 1985) and the collapse of the Bretton Woods system (Melvin and Norrbin, 2017), may have further distorted production and constrained efficiency, underscoring the combined impact of internal and macroeconomic factors on firm productivity.

Output

Table 1: Regression Results: Alternative Estimation Methods

	Pooled OLS			Fixed Effects			First Differences			Random Effects		
	β	SE	t	β	SE	t	β	SE	t	β	SE	t
Intercept	0.000	(0.016)	0.00	—	—	—	—	—	—	0.000	(0.005)	0.00
Log capital	0.310***	(0.032)	9.58	0.155***	(0.030)	5.16	0.063**	(0.023)	2.71	0.319***	(0.010)	31.64
Log employment	0.675***	(0.037)	18.45	0.694***	(0.042)	16.67	0.549***	(0.029)	18.82	0.667***	(0.011)	58.67
R^2		0.914			0.477			0.165			0.923	
σ^2		0.131			0.018			0.014			0.114	

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Dependent variable: Log output. Panel data regression results across four estimation methods.

Table 2: Tests

Statistic	**Hausman Test**	**Wald Test**
Test Statistic	24.38	19.4029
p-value	0.000005	0.00001
	Critical Values	
5% level	5.99*	3.8415*
1% level	9.21**	6.6349**
0.001% level	23.03***	10.8276***

References

[1] Melvin, M. and Norrbin, S. (2017) 'Chapter 2 - International Monetary Arrangements', in Melvin, M. and Norrbin, S. (eds.) International Money and Finance (9th ed.). Academic Press, pp. 25–57. Available at: https://doi.org/10.1016/B978-0-12-804106-2.00002-2.

[2] Schumacher, D. (1985) The 1973 Oil Crisis and its Aftermath. In: Energy: Crisis or Opportunity?. Palgrave, London. https://doi.org/10.1007/978-1-349-17797-4_2