Project 1: Linear Panel Data and Production Technology

AØKA08084U Advanced Microeconometrics

University of Copenhagen

Character Count (incl. spaces): 11864

Sebastian Hørlück, Matias Piqueras

October 10, 2021

1 Introduction

This assignment empirically investigates the Cobb-Douglas production function using panel data of $N = 441$ French firms in the period 1968–1979 ($T = 12$). The basic model specification is given by

$$y_{it} = \beta_K k_{it} + \beta_L \ell_{it} + v_{it} \tag{1}$$

where the outcome y_{it} is the log of deflated sales, k_{it} is the log of adjusted capital stock and ℓ_{it} the log of employment. We define the error term as $v_{it} := \ln A_{it}$ from the Cobb-Douglas model to capture time-varying and time-invariant factors affecting the outcome y_{it}. Our objective is to find an estimator that yields, in the following order of priority: (1) consistent and (2) efficient estimates of β_K and β_L. We base the decision of which estimator to choose on theoretical reasoning and formal testing, beginning with an inquiry into error term v_{it} and how it relates to the regressors. In this respect, we start by showing that the Pooled OLS (POLS) and the Random Effect (RE) estimators are likely to yield inconsistent estimates and then move on to discuss the Fixed Effects (FE) and First Difference (FD) estimators may alleviate this problem. Next, we focus on inference and in particular the estimation of the asymptotic variance. As a final step in outlining the econometric theory we introduce the formal tests used for model evaluation and selection before we move on to present the empirical results. We finish the assignment with a brief discussion focusing on the limitations of our results.

2 Econometric Theory

As a starting point will we assume that the error term is the composite of two components $v_{it} := c_i + u_{it}$, where c_i is the captures the unobserved time-invariant effect for i and u_{it} is the idiosyncratic error that varies across both i and t. That is, we assume that deflated sales will depend partly on firm-specific, time-invariant effects, captured in c_i. Such firm-specific effects could be the geographical location of firms, or which industry the firm operates within. Moreover, we will also assume that the expectation of c_i across firms is $\mathrm{E}(c_i) = 0$ to include an intercept to arrive at the *unobserved effects model* (UEM) (Wooldridge, 2010, p.285).

$$y_{it} = \beta_0 + \beta_K k_{it} + \beta_L \ell_{it} + c_i + u_{it} \tag{2}$$

For convenience and going forward we will often use matrix notation so $\boldsymbol{\beta} = [\beta_0, \beta_K, \beta_L]$ and $\boldsymbol{x}_{it} = [1, k_{it}, \ell_{it}]$ are row vectors. Further we use $\boldsymbol{X}$ equal to $\boldsymbol{x_{it}}$ for all $i \in N$ and $t \in T$, $\boldsymbol{y}$ equal to y_{it} for all $i \in N$ and $t \in T$, $\boldsymbol{c}$ equal to c_i for all $i \in N$, and $\boldsymbol{u}$ equal to u_{it} for all $i \in N$ and all $t \in T$. Thus the model can be written as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{c} + \boldsymbol{u}$.

2.1 Consistency

Throughout this section, we describe the consistency and efficiency of POLS, RE, FE, and FD. We estimate the partial effects $\boldsymbol{\beta}$ in (2) by solving $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N}(y_{it} - \boldsymbol{X\beta})$. This yields the POLS estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{POLS} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{y}\right)^{-1} & (3)\\ &= \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{c} + \boldsymbol{X}'\boldsymbol{u}\right) & (4) \end{aligned}$$

Taking the probability limit of this then yields

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \hat{\boldsymbol{\beta}}_{POLS} &= \boldsymbol{\beta} + \underbrace{\mathrm{E}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left[\mathrm{E}\left(\boldsymbol{X}'\boldsymbol{c}\right) + \mathrm{E}\left(\boldsymbol{X}'\boldsymbol{u}\right)\right]}_{0} & (5)\\ &= \boldsymbol{\beta} & (6) \end{aligned}$$

From (5) it is seen, that POLS rely on the assumption that the regressors are uncorrelated with both time-variant and time-invariant omitted variables, that is

$$\mathrm{E}(\boldsymbol{X}(\boldsymbol{c} + \boldsymbol{u})) = 0 \qquad \text{(POLS.1/RE.1)}$$

However, if $\mathrm{E}(\boldsymbol{X}'\boldsymbol{c}) \neq 0$ we would have an omitted variable bias, and both the POLS and RE estimator would not be consistent. This case can be illustrated through the example of industry specific affects, where some industries are more labor-intensive relative to capital than others. In this case c_i would be correlated with ℓ_{it} and thus violate the exogeneity assumption of POLS. This implies that $\boldsymbol{\beta}$ cannot be estimated consistently through POLS. We expect this, and therefore introduce the FD and FE-estimators. Moving on, we will not discuss the further assumptions needed for POLS an RE.

FD transforms the data, by first-differencing all variables, such that we have:

$$y_{it} - y_{it-1} = (\beta_0 - \beta_0) + \beta_K(k_{it} - k_{it-1}) + \beta_L(\ell_{it} - \ell_{it-1}) + \tag{7}$$
$$(c_i - c_i) + (u_{it} - u_{it-1})$$
$$\Delta y_{it} = \beta_K \Delta k_{it} + \beta_L \Delta \ell_{it} + \Delta u_{it} \tag{8}$$

From (8) shows that by transforming the data, we effectively remove the individual specific time-invariant effects, seen in (2), and therefore the endogeneity problems stemming from $\mathrm{E}(\boldsymbol{X'c}) \neq 0$ are no longer present. Similarly, for FE we also transform the data, by taking the difference from the mean of each variable within individuals. This yields:

$$y_{it} - \bar{y_i} = (\beta_0 - \beta_0) + \beta_K(k_{it} - \bar{k_i}) + \beta_L(\ell_{it} - \bar{\ell_i}) + (c_i - \bar{c_i}) + (u_{it} - \bar{u_i}) \tag{9}$$
$$\ddot{y}_{it} = \beta_K \ddot{k}_{it} + \beta_L \ddot{\ell}_{it} + \ddot{u}_{it} \tag{10}$$

where a bar above a variable indicates the mean of this variable across all $t \in T$, such that $\bar{x_i} = \frac{1}{T}\sum_{i=1}^{T} x_{it}$. Since we have defined c_i to be time-invariant, we have that $\frac{1}{T}\sum_{i=1}^{T} c_i = c_i$, which is why, c_i is cancelled out when going from (9) to (10).

The consistency of the FD and FE estimator rely on two similar assumptions. Firstly, the transformed regressors need to be exogenous to the idiosyncratic error-term. Due to the transformation of the data when estimating $\hat{\boldsymbol{\beta}}$ through FD, we have that the exogeneity assumption for FD not only depends on the explanatory variables from the current period, but also for the two surrounding periods. This can be written as

$$\mathrm{E}(\boldsymbol{\Delta x'_{it}} \Delta u_{it}) = \mathrm{E}(u_{it} | \boldsymbol{x_{it-1}}, \boldsymbol{x_{it}}, \boldsymbol{x_{it+1}}) = 0 \tag{FD.1}$$

Similarly, for the FE estimator we have that the exogeneity assumption relies on the explanatory variables from all time-periods, since the estimator is based upon a mean of the explanatory variables across all time-periods. This can be written as

$$\mathrm{E}(\boldsymbol{\ddot{x}'_{it}} \ddot{u}_{it}) = \mathrm{E}(u_{it} | \boldsymbol{x_{i1}}, \ldots, \boldsymbol{x_{iT}}) = 0 \tag{FE.1}$$

Further, for both the FD and the FE estimator we need the transformed regressors to show full rank, such that

$$\mathrm{E}(\Delta\boldsymbol{X}'\Delta\boldsymbol{X}) \text{ is invertible} \tag{FD.2}$$

and

$$\mathrm{E}(\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}}) \text{ is invertible} \tag{FE.2}$$

For now, we assume both (FD.1) and (FE.1) to hold. This assumption is not a given, and there are several cases under which the assumption might be violated. We test the assumption further in section 3 and discuss the implication of a possible violation in section 4.

2.2 Inference and Asymptotics

In this section we describe how we can derive the asymptotic variance of both the FD and FE-estimator. By combining (4) with (8) and (10) respectively, as well as a bit of rewriting, we have the following to representations of the differences in estimated effects and true effects

$$(\hat{\boldsymbol{\beta}}_{FD} - \boldsymbol{\beta}_{FD}) = (\Delta\boldsymbol{X}'\Delta\boldsymbol{X})^{-1}(\Delta\boldsymbol{X}'\Delta\boldsymbol{u}) \tag{11}$$

$$(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_{FE}) = \left(\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}}\right)^{-1}\left(\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{u}}\right) \tag{12}$$

By multiplying both sides of (11) and (12) by $\sqrt{N}$ it can be shown that these differences converge in distribution to a normal distribution with mean zero and variance

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FD} - \boldsymbol{\beta}_{FD}) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1}) \tag{13}$$

with $A = \mathrm{E}(\Delta\boldsymbol{X}'\Delta\boldsymbol{X})$ and $B = \mathrm{E}[\Delta\boldsymbol{X_i}'\mathrm{E}(\Delta\boldsymbol{u_i}\Delta\boldsymbol{u_i}'|\Delta\boldsymbol{X_i})\Delta\boldsymbol{X_i}]$, and

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_{FE}) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1}) \tag{14}$$

with $A = \mathrm{E}(\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}})$ and $B = \mathrm{E}[\ddot{\boldsymbol{X}}_{\boldsymbol{i}}'\mathrm{E}(\ddot{\boldsymbol{u}}_{\boldsymbol{i}}\ddot{\boldsymbol{u}}_{\boldsymbol{i}}'|\ddot{\boldsymbol{X}}_{\boldsymbol{i}})\ddot{\boldsymbol{X}}_{\boldsymbol{i}}]$. Both these convergences in distribution rely on the exogeneity assumption related to the idiosyncratic error-term u_{it}, since it uses the Central Limit Theorem (CLT) of $\frac{1}{\sqrt{N}}\sum_i w_i \xrightarrow{d} \mathcal{N}(0, \mathrm{V}(w_i))$ if $\mathrm{E}(w_i) = 0$, (Munk-Nielsen, 2021, p.41). As explained in section 2.1 we assumed both $\mathrm{E}(\boldsymbol{\Delta x_{it}}\Delta u_{it}) = 0$ and $\mathrm{E}(\ddot{\boldsymbol{x}}_{\boldsymbol{it}}\ddot{u}_{it}) = 0$ for consistency. Therefore we can apply the CLT to (11) and (12) to arrive at (13) and (14) respectively. Thereby we can estimate the variance of the estimators, $\widehat{\mathrm{Avar}}(\hat{\beta}_{FD})$ and $\widehat{\mathrm{Avar}}(\hat{\beta}_{FE})$ as

$$\widehat{\mathrm{Avar}}(\hat{\boldsymbol{\beta}}_{FD}) = (\Delta \boldsymbol{X}'\Delta \boldsymbol{X})^{-1}\left(\sum_{i=1}^{N}\Delta \boldsymbol{X}_{\boldsymbol{i}}'\Delta\hat{\boldsymbol{u}}_{\boldsymbol{i}}\Delta\hat{\boldsymbol{u}}_{\boldsymbol{i}}'\Delta \boldsymbol{X}_{\boldsymbol{i}}\right)(\Delta \boldsymbol{X}'\Delta \boldsymbol{X})^{-1} \tag{15}$$

$$\widehat{\mathrm{Avar}}(\hat{\boldsymbol{\beta}}_{FE}) = (\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}})^{-1}\left(\sum_{i=1}^{N}\ddot{\boldsymbol{X}}_{\boldsymbol{i}}'\hat{\ddot{\boldsymbol{u}}}_{\boldsymbol{i}}\hat{\ddot{\boldsymbol{u}}}_{\boldsymbol{i}}'\ddot{\boldsymbol{X}}_{\boldsymbol{i}}\right)(\ddot{\boldsymbol{X}}'\ddot{\boldsymbol{X}})^{-1} \tag{16}$$

where $\Delta\hat{u}_i$ and $\hat{\ddot{u}}_i$ are the residuals obtained from the FD and FE estimations respectively. These variance estimators are equivalent to (10.70) and (10.59) from (Wooldridge, 2010) respectively. (15) and (16) estimate heteroskedastic robust variances of $\boldsymbol{\beta}_{FD}$ and $\boldsymbol{\beta}_{FE}$ relying only on T being relative small compared to N (Wooldridge, 2010, p.311). Thus we allow for serial correlation in the idiosyncratic error-terms.
If we were further willing to assume that Δu_{it} is IID (FD.3), which corresponds to u_{it} being a unit root process, (15) collapses to a simpler equation, and FD will be efficient. If instead we are willing to assume that $\ddot{u}_{it}$ is IID (FE.3), (16) collapses to a simpler statement and FE will be efficient. We can test for both these cases, using a test for serial correlation, further explained in 3.

We can use (15) and (16) to test whether our estimates, (8) and (10) imply constant returns to scale in the Cobb-Douglas production function. To do this we apply a Wald test, to test the null hypotheses of

$$\mathcal{H}_0^{FD} : \beta_k^{FD} + \beta_\ell^{FD} = 1 \tag{17}$$

$$\mathcal{H}_0^{FE} : \beta_k^{FE} + \beta_\ell^{FE} = 1 \tag{18}$$

The Wald test allows for the test of multiple hypotheses relating to the estimated parameters, as well as testing hypotheses building on combinations of the estimated parameters. In our case we want to test whether the estimated Cobb-Douglas production functions shows constant returns to scale. Formally this corresponds to $\beta_k = \beta_\ell$ being equal to one.

The Wald test used for testing, whether the estimated production function shows constant returns to scale, tests the null hypotheses of $R\boldsymbol{\beta} = r$, where $R = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $r = \begin{pmatrix} 1 \end{pmatrix}$. The Wald test statistic is calculated as

$$W = (R\hat{\boldsymbol{\beta}} - r)'(\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}))^{-1}(R\hat{\boldsymbol{\beta}} - r) \tag{19}$$

Following (Wooldridge, 2010, p.42) it can be shown, that under the null, the Wald test statistic will converge in distribution to a χ^2_Q distribution with Q amounting to the number of hypotheses being tested and the degrees of freedom in the distribution. This convergence rely on the asymptotic distribution of $\hat{\boldsymbol{\beta}} \xrightarrow{d} \mathcal{N}(\boldsymbol{\beta}, \text{Avar}(\hat{\boldsymbol{\beta}}))$ which again rely on the consistency of $\hat{\boldsymbol{\beta}}$, and the assumptions required for this.
The results of the Wald test are shown in section 3.

3 Empirical Results

As a first step in our analysis we evaluate (FD.3) and (FE.3) by testing for serial correlation in the idiosyncratic error term. We do this following the steps laid out by Wooldridge (2010, p.320), using the error term obtained from FD $e_{it} \equiv \Delta u_{it}$ to run the pooled OLS regressions

$$\hat{e}_{it} = \hat{\rho}_1 \hat{e}_{i,t-1} + \text{error}_{it}, \qquad t = 3, 4, \ldots, T; i = 1, 2, \ldots, N. \tag{20}$$

The quantity of interest here is $\hat{\rho}_1$, since it can be shown that under FE3 the correlation will be $\text{Corr}(e_{it}, e_{i,t-1}) = -.5$. However, if we instead think that e_{it} follows a random-walk as under FD3, then the correlation should be $\text{Corr}(e_{it}, e_{i,t-1}) = 0$. The results from estimating (20) are reported in table 1. The coefficient $\hat{\rho}_1$ is -0.199 with a standard error of .015 and significant at $p < 0.001$. This provides evidence of negative serial correlation in Δu_{it} warranting the use of heteroscadasticity-robust standard errors derived in (15) and (16) for the estimation of FE and FD. Moreover, the results are coherent with Wooldridge (2010, p.321) observation that the behavior of the error term is likely to lie somewhere in-between (FE.3) and (FD.3) in many empirical settings, in our case marginally suggesting FD being more efficient.

Table 1: Test of Serial Correlation

	Dependent variable:
	$\hat{e}_{it}$
$\hat{e}_{i,t-1}$	-0.199***
	(0.015)
Observations	4,410
R^2	0.039
Adjusted R^2	0.039
Note:	*p<0.1; **p<0.05; ***p<0.01

In table 2 we report the FE and FD estimates from estimating (8) and (10) respectively. Standard errors are heteroscedasticity-robust. We can see from the FE estimates than a one percent increase in ℓ_{it} suggests an increase of about 0.155% increase in y_{it}, with a standard error of 0.03%. The corresponding estimate for FD suggests an 0.063% increase with a standard error of 0.023%. Also the estimates for k_{it} differ significantly between the models, with an estimate of 0.694% and a standard error of 0.042% for FE compared to 0.549% and 0.029% for FD. All estimates are significant at $p < 0.001$. However, the large difference in the estimated effect size between the models likely renders any interpretation thereof problematic. In fact, the difference between estimates suggests that we should be worried about assumption FE.1 and FD.1 being violated, causing estimates to be inconsistent (Wooldridge, 2010, p.321). We consider attributing such differences to sampling error ill-advised without any insights into the data-generating process. We discuss this further in sections 3.1 and 4.

Table 2: Main results

	Dependent variable:	
	Log deflated sales	
	FE	FD
	(1)	(2)
Log adjusted capital	0.155***	0.063***
	(0.030)	(0.023)
Log employment	0.694***	0.549***
	(0.042)	(0.029)
Observations	5,292	4,851
R^2	0.477	0.165
Adjusted R^2	0.429	0.165
Note:	*p<0.1; **p<0.05; ***p<0.01	

As a final step in this part of the analysis we explicitly evaluate if the production function shows constant returns to scale by testing the *null* hypotheses (17) and (18) using the Wald test statistic derived in (19). For FE the test statistic is $\chi^2 = 19.403$ and for FD $\chi^2 = 251.73$, in both cases significant at $p < 0.001$ leading us to reject the *null*-hypothesis. More specifically, as $\beta_k + \beta_\ell < 1$ for both estimators, we find diminishing returns to scale to be a more accurate characterisation of the relation between employment, capital and deflated sales. We want to stress again, that these results should be interpreted with caution since (FD.1) and (FE.1) are likely violated.

3.1 Test of strict exogeneity

Since the assumption of strict exogeneity as per (FE.1) and (FD.1) are key for our confidence in interpreting the results in 2, we test these assumptions formally. As described in (Wooldridge, 2010, p.325) we can do this by estimating $\ddot{y}_{it} = \ddot{\boldsymbol{x}}_{it}\boldsymbol{\beta} + \ddot{\boldsymbol{x}}_{it+1}\boldsymbol{\delta} + \ddot{u}_{it}$, since strict exogeneity would imply $\delta = 0$. We present these results in table 3. From here we see that $\delta \neq 0$ since both β_K and β_L are significant and positive. This implies that strict exogeneity is violated and neither FD or FE produce consistent estimates of $\boldsymbol{\beta}$. Further, since δ is positive, we assume the bias to arise from (FD.1) and (FE.1) being violated is positive, indicating that (8) and (10) both overestimate $\boldsymbol{\beta}$, thus not challenging the conclusion of the Cobb-Douglas production function showing diminishing returns to scale.

Table 3: Fixed effects with leads and lags

	Dependent variable:	
	Log deflated sales	
	(1)	(2)
Log adjusted capital	0.028	0.143***
	(0.038)	(0.045)
Log employment	0.541***	0.643***
	(0.043)	(0.036)
Log adjusted capital$_{t+1}$	0.167***	
	(0.046)	
Log employment$_{t+1}$	0.142***	
	(0.028)	
Log adjusted capital$_{t-1}$		0.007
		(0.033)
Log employment$_{t-1}$		0.057
		(0.037)
Observations	4,851	4,851
R^2	0.478	0.448
Adjusted R^2	0.426	0.392
Note:	*p<0.1; **p<0.05; ***p<0.01	

4 Discussion

As shown in section 3, there are relatively large differences in the estimates of $\boldsymbol{\beta}$ based on whether the FD or the FE estimator is used. The significance of the difference in the estimates can be tested through a Hausmann test, which we will not conduct in this paper. The differences in estimates could stem from the exogeneity assumptions used to derived the estimators being violated. This would cause both estimators to be inconsistent, and have different probability limits, (Wooldridge, 2010, p.321-322). Generally speaking, correlation between x_{it} and u_{is} can appear in either the same period, $t = s$, in the lagged regressor, $t < s$, and/or in the future regressor, $t > s$. Any of these cases causes inconsistency of both FD and FE, and challenges any conclusions based on the results from section 3.

As seen in section 3.1, the leaded regressors are significant. This implies two main problems. Firstly, we have only a limited number of years for each individual in our dataset. For every leaded variable we include, we limit the number of years we can include in our dataset. Since we do not know exactly for how many years the correlation between the idiosyncratic error-term and the regressors will persist, we do no know how many years we would have to drop to arrive at a correct model specification, and thus find consistent estimates of $\boldsymbol{\beta}$. Secondly, it is generally not meaningful to include too many leaded variables in a economic regression, since the need for leaded variables reduces the timeliness of the regression output. Even if it the goal is not to estimate the output for the most recent year possible, the need for leaded variables reduces the data available to esimate $\boldsymbol{\beta}$.

The endogeneity problems imply that the FD and FE estimators might not be sufficient to estimate $\boldsymbol{\beta}$ consistently. However, the results from section 3.1 imply that we can still conclude the Cobb-Douglas production function to show diminishing returns to scale.

References

Munk-Nielsen, A. (2021). *Linear models for panel data: 2nd handout.* University Lecture.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.