Production Function Estimation

Rami Abou-Seido

August 2020

1 Introduction

Here I will present several ways to estimate production functions from the economics literature. The first two will be standard linear regression approaches. I will then discuss using a fixed effects estimator. Lastly I will present the Olley-Pakes production function estimation approach.

Production datasets will usually be a panel dataset. They will include data from a set of firms indexed by i, across some time periods indexed by t. For example, if the data is monthly, the panel would involve an entry for each firm in each month. The dataset I use includes 531 firms. It will include annual data for 6 years. t will represent the year which the data will come in.

Before I begin estimating the production functions, I must specify the shape of the function that I am estimating. I will assume that the production function has a Cobb-Douglas specification, that is:

$$Y_{it} = AL_{it}^{\beta_l} K_{it}^{\beta_k} \tag{1}$$

where

- Y_{it} is the output of firm i in period t
- L_{it} is the labor input (workers hired) of firm i in period t
- K_{it} is the capital stock (factory equipment, land, inventory etc.) of firm i in period t
- A is a constant, it can be interpreted as productivity
- β_l and β_k are the output elasticities of labor and capital respectively. These need to be estimated.

The Cobb-Douglas specification is standard in economics. The goal is to estimate β_l and β_k so the effect of hiring more workers or acquiring more capital on production can be understood. It also matters when trying to understand why wages are low or high. Higher β_l means hiring the next worker will add greater output compared to low β_l. This means the firm would be willing to pay a higher when wage β_l is higher.

Now that I have introduced the production function that is being estimated, I will show several estimation techniques in the following sections.

2 Ordinary Least Squares

By taking logarithms on all sides of equation 1 and adding an error term ϵ_{it} gives:

$$y_{it} = a + \beta_l l_{it} + \beta_k k_{it} + \epsilon_{it} \tag{2}$$

where

- $y_{it} = log(Y_{it})$
- $l_{it} = log(L_{it})$
- $k_{it} = log(K_{it})$
- $a = log(A)$

Lastly ϵ_{it} is random with mean zero.

By taking logarithms, β_l and β_k can be estimated using standard linear regression (equation 2. The intercept would be a.

Estimation of OLS

The panel data set is unbalanced. What I mean by this is that some firms have entries for all 6 years while others have output zero and are not operating in some years. There are several reasons for this, one possibility is that they go out of business before the 6 years are up. Another possibility is that they they enter into the market late. I will estimate the production function with both an unbalanced dataset and a balanced one. For the balanced one, I will only include firms that produce output all 6 years. The following is a table of the results:

	Unbalanced	Balanced
β_l	0.830	0.771
β_k	0.107	0.129
mean sq. error	0.307	0.305

Both approaches provide roughly the same mean squared error. The unbalanced data gives a larger elasticity of labor compared to the balanced. With regards to the capital stock, it is the opposite. It is hard to interpret why this is the case. There are several issues with using OLS that I will discuss in the following subsection.

Issues

Firms have information about their own production functions. For instance, workers will produce more output at a firm that has great managers. Firms with better managers might also hire more workers. There is no data on which firms have good managers, so a linear regression approach will overestimate β_l. If such data was available, good managers can be controlled for and β_l would be more precisely estimated.

In general, the fundamental issue is that firms choose how many people to hire and what kind of equipment they want to have. This decision is based on factors that can not be observed by the data analyst. This can give incorrect results when estimating the production function. I will introduce two other estimation approaches to try and deal with this issue.

3 Fixed Effect Estimator

One way to deal with the issues raised in the last section is to use a fixed effect estimator. Here we change the log specification of the production function to the following:

$$y_{it} = a + \beta_l l_{it} + \beta_k k_{it} + w_i + \epsilon_{it} \tag{3}$$

where w_i is a firm specific productivity variable that are observable to the firm but not available in the dataset. It accounts for the fact that some firms might have better managers that make it more productive compared to other firms.

Now the following terms are defined:

- $\bar{y}_i$ is the average log output across time of firm i
- $\bar{l}_i$ is the average log labor input across time of firm i
- $\bar{k}_i$ is the average log capital input across time of firm i
- ϵ_i is the average of the error term across time for firm i

By definition the following equation is true:

$$\bar{y}_i = a + \beta_l \bar{l}_i + \beta_k \bar{k}_i + w_i + \epsilon_i \tag{4}$$

Taking the difference between equation 3 and 4 gives:

$$[y_{it} - \bar{y}_i] = \beta_l [l_{it} - \bar{l}_i] + \beta_k [k_{it} - \bar{k}_i] + \epsilon_{it} - \epsilon_i \tag{5}$$

since $\epsilon_{it} - \epsilon_i$ will essentially be an error term with mean zero, β_l and β_k can be estimated using standard linear regression where:

- $y_{it} - \bar{y}_i$ is the dependant variable
- $[l_{it} - \bar{l}_i]$ and $[k_{it} - \bar{k}_i]$ are the independent variables

Estimation of Fixed Effect Approach

The following are the results from estimating the fixed effect model.

	Fixed Effect Estimates
β_l	7.477
β_k	0.142
mean sq. error	46.58

Clearly, looking at β_l the results here can not be realistic, the mean squared error is also very large. One possible explanation for why this technique did not work is because there is not enough variation within firm. For this estimator to be effective, variation in $[l_{it} - \bar{l}_i]$ and $[k_{it} - \bar{k}_i]$ is needed. Firms though usually keep capital and labor relatively constant. This means that there will not be enough variation in $[l_{it} - \bar{l}_i]$ and $[k_{it} - \bar{k}_i]$.

4 Olley-Pakes Estimator

The approach in this section will use economic theory to deal with the issues raised in section 2. By making assumptions about how the firms select labor and capital, the production function can be estimated taking into account the unobservable characteristics that affect the firms' decision. It is based off the Olley-Pakes production function estimation approach.

The log production function will be:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + w_{it} + \epsilon_{it} \tag{6}$$

Unlike the fixed effect estimator, the unobservable firm characteristics w_{it} can vary with time.

I will now consider an additional variable that is observable. It will be investment I_{it} by firm i in period t. Let $i_{it} = log(I_{it})$.

The following are theoretical assumptions that will help estimate the production function.

- $i_{it} = f_k(k_{it}, w_{it})$, the firms investment decision depends on capital stock and the firms productivity, f_k represents the decision given k_{it} and w_{it}
- f_k is invertible in w_{it}
- $k_{it} = (1 - \delta)k_{i,t-1} + i_{i,t-1}$. The capital stock will be the previous periods capital depreciated at rate δ with the addition of the investment made in the previous period.

- w_{it} is first order markov

With regards to the first two assumptions, f_k does not need to be known, just that it exists and that it is invertible with respect to w_{it}. With regards to the third assumption, δ does not need to be known, in fact the only part that matters is that k_{it} is determined from period $t-1$. It is really a time to build assumption for the capital stock. Basically if firms want to upgrade factory equipment etc., it is being assumed that it will take time to do this.

Using the assumptions stated above, the estimation will happen in two stages.

Stage 1

Firstly, given the first three assumptions, the following is true:

$$w_{it} = f_k^{-1}(i_{it}, k_{it}) \tag{7}$$

Using equation 7, the log production function from equation 6 can be re-written as:

$$\begin{aligned} y_{it} &= \beta_l l_{it} + \beta_k k_{it} + w_{it} + \epsilon_{it} \\ &= \beta_l l_{it} + \beta_k k_{it} + f_k^{-1}(i_{it}, k_{it}) + \epsilon_{it} \end{aligned} \tag{8}$$

Let $\phi_{it}(k_{it}, i_{it}) = \beta_k k_{it} + f_k^{-1}(i_{it}, k_{it})$ then:

$$y_{it} = \beta_l l_{it} + \phi_{it}(k_{it}, i_{it}) + \epsilon_{it} \tag{9}$$

This is a partially linear model. It can be estimated using standard semi-parametric methods in statistics The original Olley-Pakes paper that presents this technique uses an n'th order polynomial. This is what will be done here with $n = 3$, so:

$$\phi_{it}(k_{it}, i_{it}) = \pi_1 k_{it} + \pi_2 k_{it}^2 + \pi_3 k_{it}^3 + \pi_4 i_{it} + \pi_5 i_{it}^2 + \pi_6 i_{it}^3 + \pi_7 k_{it} i_{it} + \pi_8 k_{it}^2 i_{it} + \pi_9 k_{it} i_{it}^2 \quad (10)$$

By substituting equation 10 into equation 9, both β_l and ϕ can be estimated by running a linear regression. This completes stage 1.

Stage 2

Given the fourth assumption that w_{it} is markov, the following must be true:

$$w_{it} = E[w_{it}|w_{i,t-1}] + \sigma_{it} = h(w_{i,t-1}) + \sigma_{it} \quad (11)$$

where E is the expectation operator and σ_{it} is a random shock. From the first stage of the estimation the following is true:

$$w_{it} = \phi_{it} - \beta_k k_{it} \quad (12)$$

Combining this with equation 11 gives:

$$\phi_{it} - \beta_k k_{it} = h(\phi_{i,t-1} - \beta_k k_{i,t-1}) + \sigma_{it} \quad (13)$$

Just like in stage 1, I will assume h is a polynomial of degree $n = 3$. Equation 13 becomes:

$$\phi_{it} - \beta_k k_{it} = \pi_1(\phi_{i,t-1} - \beta_k k_{i,t-1}) + \pi_2(\phi_{i,t-1} - \beta_k k_{i,t-1})^2 + \pi_3(\phi_{i,t-1} - \beta_k k_{i,t-1})^3 + \sigma_{it} \quad (14)$$

This will allow me to estimate β_k with the following procedure:

For each possible value of β_k, complete the following steps.

1. Given β_k, find $\hat{w}_{it} = \phi_{it} - \beta_k k_{it}$. ϕ_{it} was estimated in stage 1.

2. Substitute $\hat{w}_{it}$ into equation 14:

$$\hat{w}_{it} = \pi_1 \hat{w}_{i,t-1} + \pi_2 \hat{w}_{i,t-1}^2 + \pi_3 \hat{w}_{i,t-1}^3 + \sigma_{it}$$

   and compute the linear regression

3. Compute the mean squared error

Finally choose the β_k that gives the smallest mean squared error.

This completes the Olley-Pakes estimation approach.

Estimation of Olley-Pakes Approach

The following are the results from estimating the Olley-Pakes approach.

	Olley-Pakes Estimates
β_l	0.816
β_k	0.193
mean sq. error	0.304

Compared to the OLS estimates, the elasticity of capital β_k is larger. This estimation strategy also provides the lowers mean squared error.