Introduction
Model and data
Model
Data
Econometric issues
Simultaneity problem
Endogenous exit
Estimation methods
Input prices as instruments
Panel data: Fixed-effects
Dynamic panel data: GMM
Control function methods
Endogenous exit
Determinants of productivity
What determines productivity?
TFP dispersion in equilibrium
How can firms improve their TFP?
R&D and productivity
Knowledge capital model
An application
Exercises
Exercise 1
Exercise 2
Exercise 3

3. Production Functions

3.1 Introduction

Production functions (PF) are important primitive components of many economic models. The estimation of PFs plays a key role in the empirical analysis of issues such as productivity dispersion and misallocation, the contribution of different factors to economic growth, skill-biased technological change, estimation of economies of scale and economies of scope, evaluation of the effects of new technologies, learning-by-doing, or the quantification of production externalities, among many others.

In empirical IO, the estimation of production functions can be used to obtain firms' costs. Cost functions play an important role in any empirical study of industry competition. As explained in chapter 1, data on production costs at the firm-market-product level is rare. For this reason cost functions are often estimated in an indirect way, using first order conditions of optimality for profit maximization (see chapter 4). However, cross-sectional or panel datasets with firm-level information on output and inputs of the production process are more commonly available. Given this information, it is possible to estimate the industry production function and use it to obtain firms' cost functions.

There are multiple issues that should be taken into account in the estimation of production functions.

(a) Measurement issues. There are important issues in the measurement of inputs, such as differences in the quality of labor, or the measurement error that results from the construction of the capital stock using a perpetual inventory method. There are also issues in the measurement of output. For instance, the problem of observing revenue instead of output in physical units.

(b) Specification assumptions. The choice of functional form for the production function is an important modelling decision, especially when the model includes different types of labor and capital inputs that may be complements or substitutes.

(c) Simultaneity / endogeneity. This is a key econometric issue in the estimation of production functions. Observed inputs (for instance, labor and capital) can be correlated with unobserved inputs or productivity shocks (for instance, managerial ability, quality of land, materials, capacity utilization). This correlation introduces biases in some

estimators of PF parameters.

(d) Multicollinearity between observed inputs is also a relevant issue in some empirical applications. The high correlation between observed labor and capital can seriously reduce the precision in the estimation of PF parameters.

(e) Endogenous exit. In panel datasets, firm exit from the sample is not exogenous and it is correlated with firm size. Smaller firms are more likely to exit compared to larger firms. Endogenous exit can introduce selection-bias in some estimators of PF parameters.

In this chapter, we concentrate on the problems of simultaneity, multicollinearity, and endogenous exit, and on different solutions that have been proposed to deal with these issues. For the sake of simplicity, we discuss these issues in the context of a Cobb-Douglas PF. However, the arguments and results can be extended to more general specifications of PFs. In principle, some of the estimation approaches can be generalized to estimate nonparametric specifications of PF. Griliches and Mairesse (1998), Bond and Van Reenen (2007), and Ackerberg et al. (2007) include surveys of this literature.

3.2 Model and data

3.2.1 Model

Basic framework

A Production Function (PF) is a description of a production technology that relates the physical output of a production process to the physical inputs or factors of production. A general representation is:

$$Y = F(X_1, X_2, ..., X_J, A) \tag{3.1}$$

where Y is a measure of firm output, X_1, X_2, .., and X_J are measures of J firm inputs, and A represents the firm's technological efficiency. The marginal productivity of input j is $MP_j = \partial F / \partial X_j$.

Given the production function $Y = F(X_1, X_2, ..., X_J, A)$ and input prices $(W_1, W_2, ..., W_J)$, the cost function $C(Y)$ is defined as the minimum cost of producing the amount of output Y:

$$\begin{aligned} C(Y) \quad &= \min_{\{X_1, X_2, ..., X_J\}} W_1 X_1 + W_2 X_2 + ... + W_J X_J \\ &\text{subject to: } Y \geq F(X_1, X_2, ..., X_J, A) \end{aligned} \tag{3.2}$$

The marginal conditions of optimality imply that for every input j:

$$W_j - \lambda \ F_j(X_1, X_2, ..., X_J, A) = 0, \tag{3.3}$$

where $F_j(X_1, X_2, ..., X_J, A)$ is the marginal productivity of input j, and λ is the Lagrange multiplier of the restriction.

Cobb-Douglas production and cost functions

A very common specification is the Cobb-Douglas PF (Cobb and Douglas, 1928):

$$Y = L^{\alpha_L} \ K^{\alpha_K} \ A \tag{3.4}$$

where L and K represent labor and capital inputs, respectively, and α_L and α_K are technological parameters that are assumed the same for all the firms in the market and industry under study. This Cobb-Douglas PF can be generalized to include more inputs, for instance, $Y = L^{\alpha_L}\, K^{\alpha_K}\, R^{\alpha_R}\, E^{\alpha_E}\, A$, where R represents R&D and E is energy inputs. We can also distinguish different types of labor (for instance, blue collar and white collar labor) and capital (for instance, equipment, buildings, and information technology). For the Cobb-Douglas PF, the productivity term A is denoted the *Total Factor Productivity* (TFP). The marginal productivity of input j is $MP_j = \alpha_j \frac{Y}{X_j}$. All the inputs are complements in production, that is, the marginal productivity of any input j increases with the amount of any other input k:

$$\frac{\partial MP_j}{\partial X_k} = \frac{\alpha_j}{X_j}\frac{\alpha_k}{X_k} Y > 0 \tag{3.5}$$

Note that this is not necessarily the case for other production functions such as the Constant Elasticity of Substitution (CES) or the Translog

More generally, we can consider a Cobb-Douglas PF with J inputs: $Y = X_1^{\alpha_1} \ldots X_1^{\alpha_J}$ A. Given this PF and input prices W_j, we can obtain the expression for the corresponding cost function. The marginal condition of optimality for input j implies $W_j - \lambda\, \alpha_j$ $(Y/X_j) = 0$, or equivalently:

$$W_j\, X_j = \lambda\, \alpha_j\, Y \tag{3.6}$$

Therefore, the cost is equal to $\sum_{j=1}^{J} W_j\, X_j = \lambda\, \alpha\, Y$, where the parameter α is defined as $\alpha \equiv \sum_{j=1}^{J} \alpha_j$. Note that α represents the returns to scale in the production function: constant if $\alpha = 1$, decreasing if $\alpha < 1$, and increasing if $\alpha > 1$. To obtain the expression of the cost function, we still need to obtain the (endogenous) value of the Lagrange multiplier λ. For this, we substitute the marginal conditions $X_j = \lambda\, \alpha_j\, Y/W_j$ into the production function:

$$Y = A \left(\frac{\lambda\, \alpha_1 Y}{W_1}\right)^{\alpha_1} \left(\frac{\lambda\, \alpha_2 Y}{W_2}\right)^{\alpha_2} \ldots \left(\frac{\lambda\, \alpha_J Y}{W_J}\right)^{\alpha_J} \tag{3.7}$$

Using this expression to solve for the Lagrange multiplier, we get

$$\lambda = \left(\frac{W_1}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha}} \left(\frac{W_2}{\alpha_2}\right)^{\frac{\alpha_2}{\alpha}} \ldots \left(\frac{W_J}{\alpha_J}\right)^{\frac{\alpha_J}{\alpha}}\ Y^{\frac{1-\alpha}{\alpha}}\ A^{\frac{-1}{\alpha}}. \tag{3.8}$$

And plugging this multiplier into the expression $\lambda\, \alpha\, Y$ for the cost, we obtain the cost function:

$$C(Y) = \alpha\, \left(\frac{Y}{A}\right)^{\frac{1}{\alpha}}\, \left(\frac{W_1}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha}} \left(\frac{W_2}{\alpha_2}\right)^{\frac{\alpha_2}{\alpha}} \ldots \left(\frac{W_J}{\alpha_J}\right)^{\frac{\alpha_J}{\alpha}} \tag{3.9}$$

Looking at the Cobb-Douglas cost function in equation (3.9) we can identify some interesting properties. First, the returns to scale parameter α determines the shape of the cost as a function of output. More specifically, the sign of the second derivative $C''(Y)$ is equal to the sign of $\frac{1}{\alpha} - 1$. If $\alpha = 1$ (*constant returns to scale*, CRS), we have $C''(Y) = 0$ such that the cost function is linear in output. If $\alpha < 1$ (*decreasing returns to scale*, DRS), we have $C''(Y) > 0$ and the cost function is strictly convex in output. Finally, if $\alpha > 1$ (*increasing returns to scale*, IRS), we have $C''(Y) < 0$ such that the cost function is concave in output.

Production functions and the linear regression model

An attractive feature of the Cobb-Douglas PF from the point of view of estimation is that it is linear in logarithms:

$$y = \alpha_L \, \ell + \alpha_K \, k + \omega \tag{3.10}$$

where y is the logarithm of output, ℓ is the logarithm of labor, k is the logarithm of physical capital, and ω is the logarithm of TFP. The simplicity of the Cobb-Douglas PF also comes with some limitations. One of its drawbacks is that it implies an elasticity of substitution between labor and capital (or between any two inputs) that is always equal to one. This implies that all technological changes are neutral for the demand of inputs. For this reason, the Cobb-Douglas PF cannot be used to study topics such as skill-biased technological change. For empirical studies where it is important to have a flexible form for the elasticity of substitution between inputs, the translog PF has been a popular specification:

$$Y = L^{[\alpha_{L0}+\alpha_{LL}\ell+\alpha_{LK}k]} \; K^{[\alpha_{K0}+\alpha_{KL}\ell+\alpha_{KK}k]} \; A \tag{3.11}$$

which in logarithms becomes,

$$y = \alpha_{L0} \, \ell + \alpha_{K0} \, k + \alpha_{LL} \, \ell^2 + \alpha_{KK} \, k^2 + (\alpha_{LK} + \alpha_{KL}) \, \ell \, k + \omega \tag{3.12}$$

3.2.2 Data

The most common type of data that has been used for the estimation of PFs consists of panel data of firms or plants with annual frequency and information on: (i) a measure of output, for instance, units produced, revenue, or value added; (ii) a measure of labor input, such as number of workers; (iii) a measure of capital input. Some datasets also include measures of other inputs such as materials, energy, or R&D, and information on input prices, typically at the industry level but sometimes at the firm level. For the US, the most commonly used datasets in the estimation of PFs are Compustat, and the Longitudinal Research Database from US Census Bureau. In Europe, some countries' Central Banks (for instance, Bank of Italy, Bank of Spain) collect firm level panel data with rich information on output, inputs, and prices.

For the rest of this chapter we consider that the researcher observes a panel dataset of N firms, indexed by i, over several periods of time, indexed by t, with the following information:

$$\textbf{Data} \;=\; \{y_{it}, \, \ell_{it}, \, k_{it}, \, w_{it}, \, r_{it} \, : \; i = 1,2,...N; t = 1,2,...,T_i\} \tag{3.13}$$

where y, ℓ, and k have been defined above, and w and r represent the logarithms of the price of labor and the price of capital for the firm, respectively. T_i is the number of periods that the researcher observes firm i.

Throughout this chapter, we consider that all the observed variables are in mean deviations. Therefore, we omit constant terms in all the equations.

3.3 Econometric issues

We are interested in the estimation of the parameters α_L and α_K in the Cobb-Douglas PF (in logs):

$$y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \omega_{it} + e_{it} \tag{3.14}$$

ω_{it} represents inputs that are known to the firm when it makes its capital and labor decisions, but are unobserved to the econometrician. These include managerial ability, quality of land, materials, etc. We refer to ω_{it} as the logarithm of *total factor productivity (log-TFP),* or *unobserved productivity,* or *productivity shock.* e_{it} represents measurement error in output, or any shock affecting output that is unknown to the firm when it chooses its capital and labor. We assume that the error term e_{it} is independent of inputs and of the productivity shock. We use y_{it}^e to represent the "true" expected value of output for the firm, $y_{it}^e \equiv y_{it} - e_{it}$.

3.3.1 Simultaneity problem

The simultaneity problem in the estimation of a PF establishes that if the unobserved productivity ω_{it} is known to the firm when it decides the amount of inputs to use in production, (k_{it}, ℓ_{it}), then these observed inputs should be correlated with the unobservable ω_{it} and the OLS estimator of α_L and α_K will be biased and inconsistent. This problem was pointed out in the seminal paper by Marschak and Andrews (1944).

Example 3.1. Suppose that firms in our sample operate in the same markets for output and inputs. These markets are competitive. Output and inputs are homogeneous products across firms. For simplicity, consider a PF with only one input, say labor: $Y = L^{\alpha_L}$ $\exp\{\omega + e\}$. The first order condition of optimality for the demand of labor implies that the expected marginal productivity should be equal to the price of labor W_L: that is, α_L $Y^e/L = W_L$, where $Y^e = Y/\exp\{e\}$, because the firm's profit maximization problem does not depend on the measurement error or/and non-anticipated shocks in e_{it}. Note that the price of labor W_L is the same for all the firms because, by assumption, they operate in the same competitive output and input markets. Then, the model can be described in terms of two equations: the production function and the marginal condition of optimality in the demand for labor. In logarithms, and in deviations with respect to mean values (no constant terms), these two equations are:[1]

$$\begin{aligned} y_{it} &= \alpha_L\, \ell_{it} + \omega_{it} + e_{it} \\ y_{it} - \ell_{it} &= e_{it} \end{aligned} \tag{3.15}$$

The reduced form equations of this structural model are:

$$\begin{aligned} y_{it} &= \frac{\omega_{it}}{1-\alpha_L} + e_{it} \\ \ell_{it} &= \frac{\omega_{it}}{1-\alpha_L} \end{aligned} \tag{3.16}$$

Given these expressions for the reduced form equations, it is straightforward to obtain the bias in the OLS estimation of the PF. The OLS estimator of α_L in this simple regression model is a consistent estimator of $Cov(y_{it}, \ell_{it})/Var(\ell_{it})$. But the reduced form equations, together with the condition $Cov(\omega_{it}, e_{it}) = 0$, imply that the covariance between log-output and log-labor should be equal to the variance of log-labor: $Cov(y_{it}, \ell_{it}) = Var(\ell_{it})$.

[1] The firm's profit maximization problem depends on output $\exp\{y_i^e\}$ without the measurement error e_i.

Therefore, under the conditions of this model, the OLS estimator of α_L converges in probability to 1 regardless of the true value of α_L. Even in the hypothetical case that labor has very low productivity and α_L is close to zero, the OLS estimator still converges in probability to 1. It is clear that – at least in this case – ignoring the endogeneity of inputs can generate a serious bias in the estimation of the PF parameters. ■

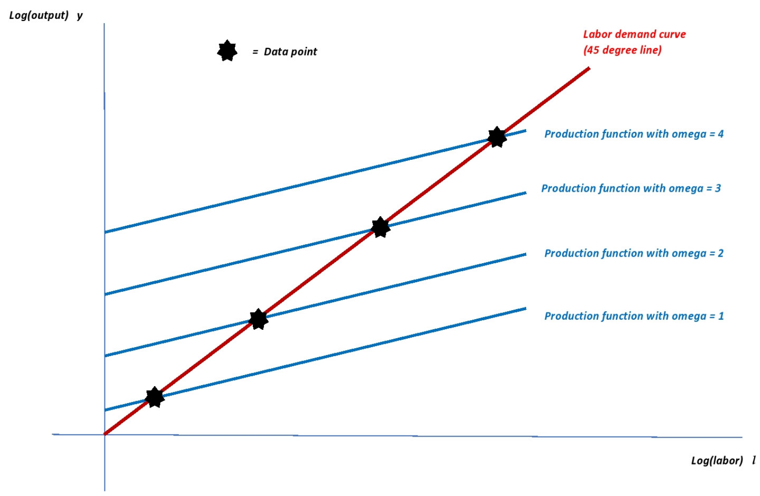

Figure 3.1: Production function and labor demand

Example 3.2: Consider similar conditions as in Example 1, but now firms produce differentiated products and use differentiated labor inputs. In particular, the price of labor R_{it} is an exogenous variable that has variation across firms and over time. Suppose that a firm is a price taker in the market for its labor input, and the price of this input, R_{it}, is independent of the firm's productivity shock, ω_{it}. In this version of the model the system of structural equations is very similar to the one in (3.15), with the only difference being that the labor demand equation now includes the logarithm of the price of labor – denoted by r_{it} — such that we have $y_{it} - \ell_{it} = r_{it} + e_{it}$. The reduced form equations for this model are:

$$\begin{aligned} y_{it} &= \frac{\omega_{it} - r_{it}}{1-\alpha_L} + r_{it} + e_{it} \\ \ell_{it} &= \frac{\omega_{it} - r_{it}}{1-\alpha_L} \end{aligned} \tag{3.17}$$

Again, we can use these reduced form equations to obtain the asymptotic bias in the estimation of α_L if we ignore the endogeneity of labor in the estimation of the PF. The OLS estimator of α_L converges in probability to $Cov(y_{it}, \ell_{it})/Var(\ell_{it})$, and in this case this implies the following expression for the bias:

$$Bias\left(\hat{\alpha}_L^{OLS}\right) = \frac{1-\alpha_L}{1+\sigma_r^2/\sigma_\omega^2} \tag{3.18}$$

where σ_ω^2 and σ_r^2 represent the variances of log-TFP and of the logarithm of labor price, respectively. This bias – of the OLS estimator of α_L – is always upward because the firm's labor demand is always positively correlated with the firm's log-TFP. The ratio between the variance of log-labor-price and the variance of log-TFP, $\sigma_r^2/\sigma_\omega^2$, plays a key role in the determination of the magnitude of this bias. Sample variability in input prices, if it is not correlated with the productivity shock, induces exogenous variability in the labor input. This exogenous sample variability in labor reduces the bias of the OLS estimator. The bias of the OLS estimator declines monotonically with the variance ratio $\sigma_r^2/\sigma_\omega^2$. Nevertheless, the bias can be very significant if the exogenous variability in input prices is not much larger than the variability in unobserved productivity. ■

3.3.2 Endogenous exit

Exit and selection problem

Panel datasets of firms or establishements can contain a significant number of firms/plants that exit from the market. Exiting firms are not randomly chosen from the population of operating firms. For instance, existing firms are typically smaller than surviving firms.

Let V_{it}^1 be the value of firm i at period t if the owners decide to stay active in the market. This value is the expected present value of future profits. Let V_{it}^0 be the value of the assets of firm i if the owners choose to exit from the market at period t. This value includes the scrap value of the assets minus exit costs such as indemnifications to workers and clients. These two values depend on the "installed" inputs of the firm an don the current value of TFP. That is, $V_{it}^1 = V^1(\ell_{it-1}, k_{it}, \omega_{it})$ and be the value of the firm at period staying in the market, $V_{it}^0 = V^0(\ell_{it-1}, k_{it}, \omega_{it})$. Let d_{it} be the indicator of the event "firm i stays in the market at the end of period t". The firm's owners maximize present value. Then, the optimal exit/stay decision is:

$$d_{it} = 1\left\{\, V^1(\ell_{it-1}, k_{it}, \omega_{it}) - V^0(\ell_{it-1}, k_{it}, \omega_{it}) \geq 0 \,\right\} \tag{3.19}$$

where $1\{S\}$ is the indicator function, such that $1\{S\} = 1$ if statement S is true, and $1\{S\} = 0$ otherwise. Under standard conditions, the difference between the value of being in the market and the value of being out, $V^1(\ell_{it-1}, k_{it}, \omega_{it}) - V^0(\ell_{it-1}, k_{it}, \omega_{it})$, is a strictly increasing in all its arguments, that is, all the inputs are more productive in the current firm/industry than in the best alternative use. Therefore, the function is invertible with respect to the productivity shock ω_{it} and we can write the optimal exit/stay decision as a single-threshold condition:

$$d_{it} = 1\left\{\, \omega_{it} \geq v\left(\ell_{it-1}, k_{it}\right) \,\right\} \tag{3.20}$$

where the threshold function $v\left(.,.\right)$ is strictly decreasing in all its arguments.

Consider the PF $y_{it} = \alpha_L\, \ell_{it} + \alpha_K\, k_{it} + \omega_{it} + e_{it}$. In the estimation of this PF, we use the sample of firms that survived at period t: that is, $d_{it} = 1$. Therefore, the error term in the estimation of the PF is $\omega_{it}^{d=1} + e_{it}$, where:

$$\omega_{it}^{d=1} \quad \equiv \quad \{\omega_{it} \mid d_{it} = 1\} \quad = \quad \left\{\omega_{it} \mid \omega_{it} \geq v\left(\ell_{i,t-1}, k_{it}\right)\right\} \tag{3.21}$$

where the notation $\{x|S\}$ represents the random variable x conditional on event S. Even if the productivity shock ω_{it} is independent of the state variables $(\ell_{i,t-1}, k_{it})$, the self-selected productivity shock $\omega_{it}^{d=1}$ will not be mean-independent of $(\ell_{i,t-1}, k_{it})$. That is,

$$\begin{aligned}\mathbb{E}\left(\omega_{it}^{d=1} \mid \ell_{i,t-1}, k_{it}\right) &= \mathbb{E}\left(\omega_{it} \mid \ell_{i,t-1}, k_{it}, d_{it} = 1\right) \\ &= \mathbb{E}\left(\omega_{it} \mid \ell_{i,t-1}, k_{it}, \omega_{it} \geq v\left(\ell_{i,t-1}, k_{it}\right)\right) \\ &= \lambda\left(\ell_{i,t-1}, k_{it}\right)\end{aligned} \tag{3.22}$$

$\lambda\left(\ell_{i,t-1}, k_{it}\right)$ is the selection term. Therefore, the PF can be written as:

$$y_{it} = \alpha_L\ \ell_{it} + \alpha_K\ k_{it} + \lambda\left(\ell_{i,t-1}, k_{it}\right) + \tilde{\omega}_{it} + e_{it} \tag{3.23}$$

where $\tilde{\omega}_{it} \equiv \{\omega_{it}^{d=1} - \lambda\left(\ell_{i,t-1}, k_{it}\right)\}$ is, by construction, mean-independent of $(\ell_{i,t-1}, k_{it})$.

Ignoring the selection term $\lambda\left(\ell_{i,t-1}, k_{it}\right)$ introduces bias in our estimates of the PF parameters. The selection term is an increasing function of the threshold $v\left(\ell_{i,t-1}, k_{it}\right)$, and therefore it is decreasing in $\ell_{i,t-1}$ and k_{it}. Both ℓ_{it} and k_{it} are negatively correlated with the selection term, but the correlation with the capital stock tends to be larger because the value of a firm depends more strongly on its capital stock than on its "stock" of labor. Therefore, this selection problem tends to bias downward the estimate of the capital coefficient.

To provide an intuitive interpretation of this bias, first consider the case of very large firms. Firms with a large capital stock are very likely to survive, even if the firm receives a bad productivity shock. Therefore, for large firms, endogenous exit induces little censoring in the distribution of productivity shocks. Consider now the case of very small firms. Firms with a small capital stock have a large probability of exiting, even if their productivity shocks are not too negative. For small firms, exit induces a very significant left-censoring in the distribution of productivity, that is, we only observe small firms with good productivity shocks and therefore with high levels of output. If we ignore this selection, we will conclude that firms with large capital stocks are not much more productive than firms with small capital stocks. But that conclusion is partly spurious because we do not observe many firms with low capital stocks that would have produced low levels of output if they had stayed.

The relationship between firm size and firm growth

This type of selection problem has been also analyzed by researchers interested in the relationship between firm growth and firm size. This relationship has relevant policy implications. Mansfield (1962), Evans (1987), and Hall (1987) are seminal papers in this literature.

Consider the regression equation:

$$\Delta s_{it} = \alpha + \beta\ s_{i,t-1} + \varepsilon_{it} \tag{3.24}$$

where s_{it} represents the logarithm of a measure of firm size, for instance, the logarithm of capital stock, or the logarithm of the number of workers.

The so called *Gibrat's law* – sometimes described as the *rule of proportionate growth* – is a hypothesis establishing that the rate of growth of a firm is independent of its size.

This "law" was postulated by gibrat (1931) – see the survey by Sutton (1997). Using equation (3.24), we can enunciate Gibrat's hypothesis as the model with $\beta = 0$.

Suppose that the exit decision at period t depends on firm size, $s_{i,t-1}$, and on a shock ε_{it}. More specifically,

$$d_{it} = 1\left\{ \varepsilon_{it} \geq v\left(s_{i,t-1}\right) \right\} \tag{3.25}$$

where $v(.)$ is a decreasing function, that is, smaller firms are more likely to exit. In a regression of Δs_{it} on $s_{i,t-1}$, we can use only observations from surviving firms. Therefore, the regression of Δs_{it} on $s_{i,t-1}$ can be represented using the equation $\Delta s_{it} = \alpha + \beta$ $s_{i,t-1} + \varepsilon_{it}^{d=1}$, where $\varepsilon_{it}^{d=1} \equiv \{\varepsilon_{it} | d_{it} = 1\} = \{\varepsilon_{it} | \varepsilon_{it} \geq v\left(s_{i,t-1}\right)\}$. Thus,

$$\Delta s_{it} = \quad \alpha + \beta s_{i,t-1} + \lambda\left(s_{i,t-1}\right) + \tilde{\varepsilon}_{it} \tag{3.26}$$

where $\lambda\left(s_{i,t-1}\right) \equiv \mathbb{E}(\varepsilon_{it} | \varepsilon_{it} \geq v\left(s_{i,t-1}\right))$, and $\tilde{\varepsilon}_{it} \equiv \{\varepsilon_{it}^{d=1} - \lambda\left(\ell_{i,t-1}, k_{it}\right)\}$ which, by construction, is mean-independent of firm size at $t-1$. The selection term $\lambda\left(s_{i,t-1}\right)$ is an increasing function of the threshold $v\left(s_{i,t-1}\right)$, and therefore it is decreasing in firm size. If the selection term is ignored in the regression of Δs_{it} on $s_{i,t-1}$, then the OLS estimator of β will be downward biased. That is, it seems that smaller firms grow faster just because small firms that would like to grow slowly have exited the industry and they are not observed in the sample.

Mansfield (1962) already pointed out to the possibility of a selection bias due to endogenous exit. He uses panel data from three US industries, steel, petroleum, and tires, over several periods. He tests the null hypothesis of $\beta = 0$, that is, Gibrat's law. Using only the subsample of surviving firms, he can reject Gibrat's Law in 7 of the 10 samples. Including also exiting firms and using the imputed values $\Delta s_{it} = -1$ for these firms, he rejects Gibrat's Law for only for 4 of the 10 samples. An important limitation of Mansfield's approach is that including exiting firms using the imputed values $\Delta s_{it} = -1$ does not correct completely for the selection bias. But Mansfield's paper was written more than a decade before James Heckman's seminal contributions on sample selection in econometrics – Heckman (1974, 1976, 1979). Hall (1987) and Evans (1987) dealt with the selection problem using Heckman's two-step estimator. Both authors find that ignoring endogenous exit induces significant downward bias in β. These two studies find that after controlling for endogenous selection à la Heckman, the estimate of β is significantly smaller than zero such that they reject Gibrat's law. A limitation of their approach is that their models do not have any exclusion restriction and identification is based on functional form assumptions: the assumptions of normal distribution of the error term, and linear (causal) relationship between firm size and firm growth.

3.4 Estimation methods

3.4.1 Input prices as instruments

If input prices, r_i, are observable and uncorrelated with log-TFP ω_i, then we can use these variables as instruments in the estimation of the PF. However, this approach has several important limitations. First, input prices are not always observable in some datasets, or they are only observable at the aggregate level but not at the firm level. Second, if firms in our sample use homogeneous inputs, and operate in the same output and input markets, we should not expect to find any significant cross-sectional variation in input prices. This is a problem because there may not be enough time-series variation for identification, or it can be confounded with any aggregate effect in the error term. Instead, suppose that firms in our sample operate in different input markets, and the researcher observes significant cross-sectional variation in input prices. In this context, a third problem is that this cross-sectional variation in input prices is likely to be endogenous: the different markets where firms operate can be different in the average unobserved productivity of firms, and therefore $cov(\omega_i, r_i) \neq 0$. That is, input prices will not be valid instruments.

3.4.2 Panel data: Fixed-effects

Suppose that we have firm level panel data with information on output, capital and labor for N firms during T time periods. The Cobb-Douglas PF is:

$$y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \omega_{it} + e_{it} \tag{3.27}$$

Mundlak (1961) and Mundlak and Hoch (1965) are pioneer studies in using panel data for the estimation of production functions. They consider the estimation of a production function of an agricultural product. They postulate the following assumptions:

Assumption PD-1: ω_{it} has the following variance-components structure: $\omega_{it} = \eta_i + \delta_t + u_{it}$. The term η_i is a time-invariant, firm-specific effect that may be interpreted as the quality of a fixed input such as managerial ability, or land quality. δ_t is an aggregate shock affecting all firms. And u_{it} is an firm idiosyncratic shock.

Assumption PD-2: The amount of inputs depend on some other exogenous time-varying variables, such that $var\left(\ell_{it} - \bar{\ell}_i\right) > 0$ and $var\left(k_{it} - \bar{k}_i\right) > 0$, where $\bar{\ell}_i \equiv T^{-1} \sum_{t=1}^{T} \ell_{it}$, and $\bar{k}_i \equiv T^{-1} \sum_{t=1}^{T} k_{it}$.

Assumption PD-3: u_{it} is not serially correlated.

Assumption PD-4: The idiosyncratic shock u_{it} is realized after the firm decides the amount of inputs to employ at period t. In the context of an agricultural PF, this shock may be interpreted as weather, or another random and unpredictable shock.

The Within-Groups estimator (WGE) or fixed-effects estimator of the PF is simply the OLS estimator applied to the Within-Groups transformation of the model. The equation that describes the within-groups transformation can be obtained by taking the difference between equation $y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \omega_{it} + e_{it}$ and this equation averaged at the firm level, that is $\bar{y}_i = \alpha_L \bar{\ell}_i + \alpha_K \bar{k}_i + \bar{\omega}_i + \bar{e}_i$. The within-groups equation is:

$$(y_{it} - \bar{y}_i) = \alpha_L \left(\ell_{it} - \bar{\ell}_i\right) + \alpha_K \left(k_{it} - \bar{k}_i\right) + (\omega_{it} - \bar{\omega}_i) + (e_{it} - \bar{e}_i) \tag{3.28}$$

Under assumptions (PD-1) to (PD-4), the WGE is consistent. Under these assumptions, the only endogenous component of the error term is the fixed effect η_i. The transitory shocks u_{it} and e_{it} do not induce any endogeneity problem. The WG transformation removes the fixed effect η_i.

It is important to point out that, for short panels (that is, T fixed), the consistency of the WGE requires the regressors $x_{it} \equiv (\ell_{it}, k_{it})$ to be strictly exogenous. That is, for any (t,s):

$$cov(x_{it}, u_{is}) = cov(x_{it}, e_{is}) = 0 \tag{3.29}$$

Otherwise, the WG-transformed regressors $\left(\ell_{it} - \bar{\ell}_i\right)$ and $\left(k_{it} - \bar{k}_i\right)$ would be correlated with the error $(\omega_{it} - \bar{\omega}_i)$. This is why Assumptions (PD-3) and (PD-4) are necessary for the consistency of the OLS estimator.

However, it is very common to find that the WGE estimator provides very small estimates of α_L and α_K (see Griliches and Mairesse, 1998). There are at least two possible reasons that can explain this empirical regularity. First, though assumptions (PD-2) and (PD-3) may be plausible for estimating PFs of agricultural firms, they are unrealistic for other industries, such as manufacturing. And second, the bias induced by measurement-error in the regressors can be exacerbated by the WG transformation. To see this, consider the model with only one input, such as capital, and suppose that it has measurement error. We observe k_{it}^* where $k_{it}^* = k_{it} + e_{it}^k$, and e_{it}^k represents measurement error in capital and it satisfies the classical assumptions on measurement error.[2] The *noise-to-signal ratio* is the ratio of variances $Var(e^k)/Var(k)$. In the estimation of the PF in levels we have that:

$$Bias(\hat{\alpha}_L^{OLS}) = \frac{Cov(k,\eta)}{Var(k) + Var(e^k)} - \frac{\alpha_L \, Var(e^k)}{Var(k) + Var(e^k)} \tag{3.30}$$

If the *noise-to-signal ratio* $Var(e^k)/Var(k)$ is small, then the (downward) bias introduced by the measurement error is negligible in the estimation in levels. In the estimation in first differences (similar to WGE, in fact equivalent when $T = 2$), we have that:

$$Bias(\hat{\alpha}_L^{WGE}) = -\frac{\alpha_L \, Var(\Delta e^k)}{Var(\Delta k) + Var(\Delta e^k)} \tag{3.31}$$

Suppose that k_{it} is very persistent (that is, $Var(k)$ is much larger than $Var(\Delta k)$) and that e_{it}^k is not serially correlated (that is, $Var(\Delta e^k) = 2 * Var(e^k)$). Under these conditions, the *noise-to-signal ratio* for capital in first differences, $Var(\Delta e^k)/Var(\Delta k)$, can be large even when the ratio $Var(e^k)/Var(k)$ is quite small. Therefore, the WGE may be significantly downward biased.

3.4.3 Dynamic panel data: GMM

In the WGE described in the previous section, the assumption of strictly exogenous regressors is very unrealistic. However, we can relax that assumption and estimate the PF using the GMM method proposed by Arellano and Bond (1991). Consider the PF in first differences:

$$\Delta y_{it} = \alpha_L \, \Delta \ell_{it} + \alpha_K \, \Delta k_{it} + \Delta \delta_t + \Delta u_{it} + \Delta e_{it} \tag{3.32}$$

We maintain assumptions (PD-1), (PD-2), and (PD-3), but we remove assumption (PD-3). Instead, we consider the following assumption.

Assumption PD-5: A firm's demands for labor and capital are dynamic. More formally, the demand equations for labor and capital are $\ell_{it} = f_L(\ell_{i,t-1}, k_{i,t-1}, \omega_{it})$ and $k_{it} = f_K(\ell_{i,t-1}, k_{i,t-1}, \omega_{it})$, respectively, where either $\ell_{i,t-1}$ or $k_{i,t-1}$, or both, have non-zero partial derivatives in f_L and f_K.

[2] Classical measurement error is independent of the true value, independently and identically distributed over observations, and with zero mean.

There are multiple reasons why the demand for capital or and labor are dynamic – that is, depend on the amount of labor and capital at previous period. Hiring and firing cost for labor, irreversibility of capital investments, installation costs, time-to-build, and other forms of adjustment costs are the most common arguments for the existence of dynamics in the demand of these inputs.

Under the conditions in Assumption PD-5, the lagged variables $\{\ell_{i,t-j}, k_{i,t-j}, y_{i,t-j} : j \geq 2\}$ are valid instruments in the PF equation in first differences. Identification comes from the combination of two assumptions: (1) serial correlation of inputs; and (2) no serial correlation in productivity shocks $\{u_{it}\}$. The presence of adjustment costs implies that the marginal cost of labor or capital depends on the firm's amount of the input at previous period. This implies that this shadow price varies across firms even if firms face the same input prices. This variability in shadow prices can be used to identify PF parameters. The assumption of no serial correlation in $\{u_{it}\}$ is key, but it can be tested (see Arellano and Bond ,1991).

This GMM in first-differences approach has also its own limitations. In some applications, it is common to find unrealistically small estimates of α_L and α_K and large standard errors (see Blundell and Bond ,2000). Overidentifying restrictions are typically rejected. Furthermore, the i.i.d. assumption on u_{it} is typically rejected, and this implies that $\{x_{i,t-2}, y_{i,t-2}\}$ are not valid instruments. It is well-known that the Arellano-Bond GMM estimator may suffer from a weak-instruments problem when the serial correlation of the regressors in first differences is weak (see Arellano and Bover ,1995, and Blundell and Bond ,1998). First difference transformation also eliminates the cross-sectional variation in inputs and it is subject to the problem of measurement error in inputs.

The weak-instruments problem deserves further explanation. For simplicity, consider the model with only one input, x_{it}. We are interested in the estimation of the PF:

$$y_{it} = \alpha\, x_{it} + \eta_i + u_{it} + e_{it} \tag{3.33}$$

where u_{it} and e_{it} are not serially correlated. Consider the following dynamic reduced form equation for the input x_{it}:

$$x_{it} = \delta\, x_{i,t-1} + \lambda_1\, \eta_i + \lambda_2\, u_{it} \tag{3.34}$$

where δ, λ_1, and λ_2 are reduced form parameters, and $\delta \in [0,1]$ captures the existence of adjustment costs. The PF in first differences is:

$$\Delta y_{it} = \alpha\, \Delta x_{it} + \Delta u_{it} + \Delta e_{it} \tag{3.35}$$

For simplicity, consider that the number of periods in the panel is $T = 3$. In this context, Arellano-Bond GMM estimator is equivalent to a simple instrumental variables estimator where the instrument is $x_{i,t-2}$. This IV estimator is:

$$\hat{\alpha}_N = \frac{\sum_{i=1}^{N} x_{i,t-2}\, \Delta y_{it}}{\sum_{i=1}^{N} x_{i,t-2}\, \Delta x_{it}} \tag{3.36}$$

Therefore, under the previous assumptions, $\hat{\alpha}_N$ identifies α if the R-square in the auxiliary regression of Δx_{it} on $x_{i,t-2}$ is not zero.

By definition, the R-square coefficient in the auxiliary regression of Δx_{it} on $x_{i,t-2}$ is such that:

$$p\lim R^2 = \frac{Cov\left(\Delta x_{it}, x_{i,t-2}\right)^2}{Var\left(\Delta x_{it}\right)\; Var\left(x_{i,t-2}\right)} = \frac{\left(\gamma_2 - \gamma_1\right)^2}{2\left(\gamma_0 - \gamma_1\right)\gamma_0} \tag{3.37}$$

where $\gamma_j \equiv Cov\left(x_{it}, x_{i,t-j}\right)$ is the autocovariance of order j of $\{x_{it}\}$. Taking into account that $x_{it} = \frac{\lambda_1\,\eta_i}{1-\delta} + \lambda_2(u_{it} + \delta\; u_{i,t-1} + \delta^2\; u_{i,t-2} + ...)$, we can derive the following expressions for the autocovariances:

$$\begin{aligned} \gamma_0 &= \frac{\lambda_1^2\;\sigma_\eta^2}{(1-\delta)^2} + \frac{\lambda_2^2\;\sigma_u^2}{1-\delta^2} \\ \gamma_1 &= \frac{\lambda_1^2\;\sigma_\eta^2}{(1-\delta)^2} + \delta\;\frac{\lambda_2^2\;\sigma_u^2}{1-\delta^2} \\ \gamma_2 &= \frac{\lambda_1^2\;\sigma_\eta^2}{(1-\delta)^2} + \delta^2\;\frac{\lambda_2^2\;\sigma_u^2}{1-\delta^2} \end{aligned} \tag{3.38}$$

Therefore, $\gamma_0 - \gamma_1 = (\lambda_2^2\sigma_u^2)/(1+\delta)$ and $\gamma_1 - \gamma_2 = \delta(\lambda_2^2\sigma_u^2)/(1+\delta)$. The R-square is:

$$\begin{aligned} R^2 &= \frac{\left(\delta\;\dfrac{\lambda_2^2\sigma_u^2}{1+\delta}\right)^2}{2\left(\dfrac{\lambda_2^2\sigma_u^2}{1+\delta}\right)\left(\dfrac{\lambda_1^2\;\sigma_\eta^2}{(1-\delta)^2} + \dfrac{\lambda_2^2\;\sigma_u^2}{1-\delta^2}\right)} \\ &= \frac{\delta^2\left(1-\delta\right)^2}{2\left(1-\delta+\left(1+\delta\right)\rho\right)} \end{aligned} \tag{3.39}$$

with $\rho \equiv \lambda_1^2\sigma_\eta^2/\lambda_2^2\sigma_u^2 \geq 0$. We have a problem of weak instruments and poor identification if this R-square coefficient is very small.

It is simple to verify that this R-square is small both when adjustment costs are small (that is, δ is close to zero) and when adjustment costs are large (that is, δ is close to one). When using this IV estimator, large adjustments costs are bad news for identification because, with delta close to one, the first difference Δx_{it} is almost iid and it is not correlated with lagged input (or output) values. What is the maximum possible value of this R-square? It is clear that this R-square is a decreasing function of ρ. Therefore, the maximum R-square occurs for $\lambda_1^2\sigma_\eta^2 = \rho = 0$ – that is, no fixed effects in the input demand. Under this condition, we have that $R^2 = \delta^2\left(1-\delta\right)/2$, and the maximum value of this R-square is $R^2 = 0.074$ which occurs when $\delta = 2/3$. This upper bound for the R-square is over-optimistic because it is based on the assumption of no fixed effects. For instance, suppose that $\lambda_1^2\sigma_\eta^2 = \lambda_2^2\sigma_u^2$ such that $\rho = 1$. In this case, we have that $R^2 = \delta^2\left(1-\delta\right)^2/4$ and the maximum value of this R-square is $R^2 = 0.016$, which occurs when $\delta = 1/2$.

Arellano and Bover (1995) and Blundell and Bond (1998) have proposed GMM estimators that deal with this weak-instrument problem. Suppose that at some period $t_i^* \leq 0$ (that is, before the first period in the sample, $t = 1$) the shocks u_{it}^* and e_{it} were zero, and input and output were equal to their firm-specific, steady-state mean values:

$$\begin{aligned} x_{it_i^*} &= \frac{\lambda_1 \eta_i}{1-\delta} \\ y_{it_i^*} &= \alpha \, \frac{\lambda_1 \eta_i}{1-\delta} + \eta_i \end{aligned} \tag{3.40}$$

Then, it is straightforward to show that for any period t in the sample:

$$\begin{aligned} x_{it} &= x_{it_i^*} + \lambda_2 \left(u_{it} + \delta u_{it-1} + \delta^2 u_{it-2} + ...\right) \\ y_{it} &= y_{it_i^*} + u_{it} + \alpha\lambda_2 \left(u_{it} + \delta u_{it-1} + \delta^2 u_{it-2} + ...\right) \end{aligned} \tag{3.41}$$

These expressions imply that input and output in first differences depend on the history of the i.i.d. shock $\{u_{it}\}$ between periods t_i^* and t, but they do not depend on the fixed effect η_i. Therefore, $cov(\Delta x_{it}, \eta_i) = cov(\Delta y_{it}, \eta_i) = 0$ and lagged first differences are valid instruments in the equation in levels. That is, for $j > 0$:

$$\begin{aligned} \mathbb{E}\left(\Delta x_{it-j}\left[\eta_i + u_{it} + e_{it}\right]\right) = 0 \;\; &\Rightarrow \;\; \mathbb{E}\left(\Delta x_{it-j}\left[y_{it} - \alpha x_{it}\right]\right) = 0 \\ \mathbb{E}\left(\Delta y_{it-j}\left[\eta_i + u_{it} + e_{it}\right]\right) = 0 \;\; &\Rightarrow \;\; \mathbb{E}\left(\Delta y_{it-j}\left[y_{it} - \alpha x_{it}\right]\right) = 0 \end{aligned} \tag{3.42}$$

These moment conditions can be combined with the "standard" Arellano-Bond moment conditions to obtain a more efficient GMM estimator. The Arellano-Bond moment conditions are, for $j > 1$:

$$\begin{aligned} \mathbb{E}\left(x_{it-j}\left[\Delta u_{it} + \Delta e_{it}\right]\right) = 0 \;\; &\Rightarrow \;\; \mathbb{E}\left(x_{it-j}\left[\Delta y_{it} - \alpha \Delta x_{it}\right]\right) = 0 \\ \mathbb{E}\left(y_{it-j}\left[\Delta u_{it} + \Delta e_{it}\right]\right) = 0 \;\; &\Rightarrow \;\; \mathbb{E}\left(y_{it-j}\left[\Delta y_{it} - \alpha \Delta x_{it}\right]\right) = 0 \end{aligned} \tag{3.43}$$

Based on Monte Carlo experiments and on actual data of UK firms, Blundell and Bond (2000) have obtained very promising results using this GMM estimator. Alonso-Borrego and Sanchez (2001) have obtained similar results using Spanish data. The reason why this estimator works better than Arellano-Bond GMM is that the second set of moment conditions exploit cross-sectional variability in output and input. This has two implications. First, instruments are informative even when adjustment costs are larger and δ is close to one. And second, the problem of large measurement error in the regressors in first-differences is reduced.

Bond and Söderbom (2005) present a very interesting Monte Carlo experiment to study the actual identification power of adjustment costs in inputs. The authors consider a model with a Cobb-Douglas PF and quadratic adjustment cost with both deterministic and stochastic components. They solve numerically the firm's dynamic programming problem, simulate data on inputs and output using the optimal decision rules, and use the Blundell-Bond GMM method to estimate PF parameters. The main results of their experiments are the following. When adjustment costs have only deterministic components, the identification is weak if adjustment costs are too low, or too high, or too similar between the two inputs. With stochastic adjustment costs, identification results improve considerably. Given these results, one might be tempted to "claim victory": if the true model is such that there are stochastic shocks (independent of productivity) in the costs of adjusting inputs, then the panel data GMM approach can identify with precision the PF parameters. However, as Bond and Soderbom explain, there is also a negative interpretation of this result. Deterministic adjustment costs have little identification power in the estimation of PFs. The existence of shocks in adjustment costs that are independent of productivity seems to be a strong identification condition. If these shocks are not present in the "true model", the apparent identification using the GMM approach could be spurious because the identification would be due to the misspecification of the model. As we will see in the next section, we obtain a similar conclusion when using a control function approach.

Table 3.1: Blundell and Bond (2000); Estimation Results

509 manufacturing firms; 1982-89				
Parameter	**OLS-Levels**	**WG**	**AB-GMM**	**SYS-GMM**
β_L	0.538	0.488	0.515	0.479
	(0.025)	(0.030)	(0.099)	(0.098)
β_K	0.266	0.199	0.225	0.492
	(0.032)	(0.033)	(0.126)	(0.074)
ρ	0.964	0.512	0.448	0.565
	(0.006)	(0.022)	(0.073)	(0.078)
Sargan (p-value)	-	-	0.073	0.032
m2	-	-	-0.69	-0.35
Constant RS (p-v)	0.000	0.000	0.006	0.641

3.4.4 Control function methods

Consider a system of simultaneous equations where some unobservables can enter in more than one structural equation. Under some conditions, we can use one of the equations to solve for an unobservable and represent it as a function of observable variables and parameters. Then, we can plug this function into another equation where this unobservable enters, such that we "control for" this unobservable by including observables. This is a particular example of a control function approach and it can be used to deal with endogeneity problems.

More generally, a control function method is an econometric procedure to correct for endogeneity problems by exploiting the structure that the model imposes on its error terms. In general, this approach implies different restrictions than the instrumental variables approach. Heckman and Robb (1985) introduced this term, though the concept had been used before in some empirical applications. An attractive feature of the control function approach is that it can provide consistent estimates of structural parameters in models where unobservables are not additively separable. In those models, instrumental variable estimators are typically inconsistent or at least do not consistently estimate the average causal effect over the whole population.

Olley and Pakes method

In a seminal paper, Olley and Pakes (1996) propose a control function approach to estimate PFs. Levinsohn and Petrin (2003) have extended this method.

Consider the Cobb-Douglas PF in the context of the following model of simultaneous equations:

$$\begin{aligned}(PF) \quad & y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \omega_{it} + e_{it} \\ (LD) \quad & \ell_{it} = f_L \left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it}\right) \\ (ID) \quad & i_{it} = f_K \left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it}\right)\end{aligned} \tag{3.44}$$

where equations (LD) and (ID) represent the firms' optimal decision rules for labor and capital investment, respectively, in a dynamic decision model with state variables $(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it})$. The vector r_{it} represents input prices. Under certain conditions on this system of equations, we can estimate consistently α_L and α_K using a control function method.

Olley and Pakes consider the following assumptions:

Assumption OP-1: $f_K \left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it}\right)$ is invertible in ω_{it}.

Assumption OP-2: There is no cross-sectional variation in input prices. For every firm i, $r_{it} = r_t$.

Assumption OP-3: ω_{it} follows a first order Markov process. That is, at any period $t \geq 0$, the transition probability $\Pr(\omega_{it} \mid \omega_{it-1}, ..., \omega_{i0})$ is equal to $\Pr(\omega_{it} | \omega_{it-1})$.

Assumption OP-4: Time-to-build physical capital. Investment i_{it} is chosen at period t but it is not productive until period $t+1$. And $k_{it+1} = (1-\delta)k_{it} + i_{it}$.

In the Olley and Pakes model, the labor input is assumed to be a static input such that lagged labor, $\ell_{i,t-1}$, is not an explanatory variable in the labor demand function f_L. This is a strong assumption as there may be substantial adjustments costs in hiring and firing workers. Most importantly, this assumption is not necessary for the Olley-Pakes method to provide a consistent estimator of the production function parameters. Therefore, we present here a version of the Olley-Pakes method where both labor and capital are dynamic inputs.

Assumption OP-2 implies that the only unobservable variable in the investment equation that has cross-sectional variation across firms is the productivity shock ω_{it}. This restriction is crucial for the OP method and for the related Levinshon-Petrin method. This imposes restrictions on the underlying model of market competition and inputs demands. This assumption implicitly establishes that firms operate in the same input markets, they do not have any monopsony power in these markets, and there are not internal labor markets within firms. Since a firm's input demand depends also on output price (or on the exogenous demand variables affecting product demand), assumption OP-2 also implies that firms operate in the same output market with either homogeneous goods or completely symmetric product differentiation. Note that these economic restrictions can be relaxed if the researcher has data on inputs prices at the firm level, that is, if r_{it} is observable.

The method proceeds in two-steps. The first step estimates α_L using a control function approach, and it relies on assumptions (*OP-1)* and *(OP-2).* This first step is the same with and without endogenous exit. The second step estimates α_K and it is based on assumptions *(OP-3)* and *(OP-4).* The Olley-Pakes method deals both with the simultaneity problem and with the selection problem due to endogenous exit.

Step 1: Estimation of α_L. Under assumptions *(OP-1)* and *(OP-2)*, we can invert the investment function to obtain a firm's TFP: that is, $\omega_{it} = f_K^{-1}(\ell_{i,t-1}, k_{it}, i_{it}, r_t)$. Solving this equation into the PF we have:

$$\begin{aligned} y_{it} &= \alpha_L\, \ell_{it} + \alpha_K\, k_{it} + f_L^{-1}(\ell_{i,t-1}, k_{it}, i_{it}, r_t) + e_{it} \\ &= \alpha_L\, \ell_{it} + \phi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it} \end{aligned} \tag{3.45}$$

where $\phi_t(\ell_{i,t-1}, k_{it}, i_{it}) \equiv \alpha_K\, k_{it} + f_L^{-1}(\ell_{i,t-1}, k_{it}, i_{it}, r_t)$. Without a parametric assumption on the investment equation f_K, equation (3.45) is a *partially linear model*.[3]

The parameter α_L and the functions ϕ_1, ϕ_2, ..., ϕ_T can be estimated using semiparametric methods. Olley and Pakes use polynomial series approximations for the nonparametric functions ϕ_t. Alternatively, one can use the method in Robinson (1988).

This method is a control function method. Instead of instrumenting the endogenous regressors, we include additional regressors that capture the endogenous part of the error term (that is, proxy for the productivity shock). By including a flexible function in $(\ell_{i,t-1}, k_{it}, i_{it})$, we control for the unobservable ω_{it}. Therefore, α_L is identified if given $(\ell_{i,t-1}, k_{it}, i_{it})$ there is enough cross-sectional variation left in ℓ_{it}.

The key conditions for the identification of α_L are: (a) the invertibility of the labor demand function $f_L(\ell_{i,t-1}, k_{it}, \omega_{it}, r_t)$ with respect to ω_{it}; (b) $r_{it} = r_t$, that is, no cross-sectional variability in unobservables, other than ω_{it}, affecting investment; and (c) given $(\ell_{i,t-1}, k_{it}, i_{it}, r_t)$, current labor ℓ_{it} still has enough sample variability. Assumption (c) is key, and it forms the basis for Ackerberg, Caves, and Frazer (2015) criticism (and extension) of the Olley-Pakes approach.

Example 3.3. Consider the Olley-Pakes model but with a parametric specification of the optimal investment equation (*ID*). More specifically, the inverse function f_K^{-1} has the following linear form:

$$\omega_{it} = \gamma_1\, i_{it} + \gamma_2\, \ell_{i,t-1} + \gamma_3\, k_{it} + r_{it} \tag{3.46}$$

Solving this equation into the PF, we have that:

$$y_{it} = \alpha_L\, \ell_{it} + (\alpha_K + \gamma_3)\, k_{it} + \gamma_1\, i_{it} + \gamma_2\, \ell_{i,t-1} + (r_{it} + e_{it}) \tag{3.47}$$

Note that current labor ℓ_{it} is correlated with current input prices r_{it}. That is the reason why we need Assumption OP-2, that is, $r_{it} = r_t$. Given that assumption we can control for the unobserved r_t by including time-dummies. Furthermore, to identify α_L with enough precision, there should not be high collinearity between current labor ℓ_{it} and the other regressors $(k_{it}, i_{it}, \ell_{i,t-1})$. ■

[3] The *partially linear model* is a regression model with two sets of regressors. One set of regressors enters linearly according to the linear index $x\beta$, and the other set of regressorts enters in a nonparametric function $\phi(z)$. That is, the regression model is $y = x\beta + g(z) + u$. The partially linear model is a class of semiparametric model that has received substantial attention in econometrics. See Li and Racine (2007).

In this first step, the control function approach deals also with the selection problem due to endogenous exit. This is because the control function controls for the value of the unobserved productivity ω_{it} such that there is not a selection problem associated with this nobservable.

Step 2: Estimation of α_K. For the sake of clarity, we first describe a version of the method that does not deal with the selection problem. We will discuss later the approach to deal with endogenous exit.

Given the estimate of α_L in step 1, the estimation of α_K is based on Assumptions *(OP-3)* and *(OP-4)*, that is, the Markov structure of the productivity shock, and the assumption of time-to-build productive capital. Since ω_{it} is first order Markov, we can write:

$$\omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{i,t-1}] + \xi_{it} = h\left(\omega_{i,t-1}\right) + \xi_{it} \tag{3.48}$$

where ξ_{it} is an innovation which is mean independent of any information at $t-1$ or before. Function $h(.)$ is unknown to the researcher and it has nonparametric form. Define $\phi_{it} \equiv \phi_t(\ell_{i,t-1}, k_{it}, i_{it})$, and remember that $\phi_t(\ell_{i,t-1}, k_{it}, i_{it}) = \alpha_K\ k_{it} + \omega_{it}$. Therefore, we have that:

$$\begin{aligned} \phi_{it} &= \alpha_K\ k_{it} + h\left(\omega_{i,t-1}\right) + \xi_{it} \\ &= \alpha_K\ k_{it} + h\left(\phi_{i,t-1} - \alpha_K\ k_{i,t-1}\right) + \xi_{it} \end{aligned} \tag{3.49}$$

Though we do not know the true value of ϕ_{it}, we have consistent estimates of these values from step 1: that is, $\hat{\phi}_{it} = y_{it} - \hat{\alpha}_L\ \ell_{it}$.[4]

If function $h(.)$ is nonparametrically specified, equation (3.49) is a partially linear model. However, it is not a standard partially linear model because the argument in function $h(.)$ is not observable. That is, though $\phi_{i,t-1}$ and $k_{i,t-1}$ are observable to the researcher (after the first step), the argument $\phi_{i,t-1} - \alpha_K k_{i,t-1}$ is unobservable because parameter α_K is unknown.

To estimate function $h(.)$ and parameter α_K, Olley and Pakes propose a recursive method. For the sake of illustration, suppose that we consider a quadratic function for $h(.)$: that is, $h(\omega) = \pi_1 \omega + \pi_2 \omega^2$. We start with an initial value for the parameter α_K, say $\hat{\alpha}_K^0$. Given this value, we construct the regressor $\hat{\omega}_{it}^0 = \hat{\phi}_{it} - \hat{\alpha}_K^0 k_{it}$, and estimate parameters (α_K, π_1, π_2) by applying OLS to the regression equation $\hat{\phi}_{it} = \alpha_K k_{it} + \pi_1 \hat{\omega}_{it-1}^0 + \pi_2 (\hat{\omega}_{it-1}^0)^2 + \xi_{it}$. Let $\hat{\alpha}_K^1$ be the OLS estimate of α_K. Then, we construct new values $\hat{\omega}_{it}^1 = \hat{\phi}_{it} - \hat{\alpha}_K^1 k_{it}$ and estimate again α_K, π_1, and π_2 by OLS. We apply this method repeatedly until convergence: that is, until the distance between the estimates of α_K in the last two iterations is smaller than a small constant: until $|\hat{\alpha}_K^n - \hat{\alpha}_K^{n-1}| < 10^{-6}$.

An alternative to this recursive procedure is the following Minimum Distance method. Again for concreteness, suppose that the specification of function $h(\omega)$ is quadratic. We have the regression model:

$$\hat{\phi}_{it} = \beta_1\ k_{it} + \beta_2\ \hat{\phi}_{i,t-1} + \beta_3\ \hat{\phi}_{i,t-1}^2 + \beta_4\ k_{i,t-1} + \beta_5\ k_{i,t-1}^2 + \beta_6\ \hat{\phi}_{i,t-1} k_{i,t-1} + \xi_{it} \tag{3.50}$$

where, according to the model, the parameters β in this regression satisfy the following restrictions: $\beta_1 = \alpha_K$; $\beta_2 = \pi_1$; $\beta_3 = \pi_2$; $\beta_4 = -\pi_1 \alpha_K$; $\beta_5 = \pi_2 \alpha_K^2$; and $\beta_6 = -2\pi_2 \alpha_K$.

[4] In fact, $\hat{\phi}_{it}$ is an estimator of $\phi_{it} + e_{it}$, but this does not have any incidence on the consistency of the estimator.

We can estimate the six β parameters by OLS. Then, in a second step, we use the OLS estimate of β and its variance-covariance matrix to estimate (α_K,π_1,π_2) by minimum distance imposing the six restrictions that relate the vector β with (α_K,π_1,π_2). More precisely, this minimum distance estimator is:

$$(\widehat{\alpha}_K, \widehat{\pi}_1, \widehat{\pi}_2) = \arg \min_{(\alpha_K,\pi_1,\pi_2)} \left[\widehat{\beta} - f(\alpha_K,\pi_1,\pi_2)\right]' \left[\widehat{V}(\widehat{\beta})\right]^{-1} \left[\widehat{\beta} - f(\alpha_K,\pi_1,\pi_2)\right] \tag{3.51}$$

where: $\widehat{\beta}$ is the column vector of OLS estimates; $\widehat{V}(\widehat{\beta})$ is its estimated variance matrix; and $f(\alpha_K,\pi_1,\pi_2)$ is the column vector with the functions $(\alpha_K,\ \pi_1,\ \pi_2,\ -\pi_1\alpha_K,\ \pi_2\alpha_K^2,\ -2\pi_2\alpha_K)$.

Example 3.4: Suppose that ω_{it} follows the AR(1) process $\omega_{it} = \rho\ \omega_{i,t-1} + \xi_{it}$, where $\rho \in [0,1)$ is a parameter. Then, $h(\omega_{i,t-1}) = \rho\,\omega_{i,t-1} = \rho(\phi_{i,t-1} - \alpha_K\ k_{i,t-1})$, and we can write:

$$\phi_{it} = \beta_1\ k_{it} + \beta_2\ \phi_{i,t-1} + \beta_3\ k_{i,t-1} + \xi_{it} \tag{3.52}$$

where $\beta_1 = \alpha_K$, $\beta_2 = \rho$, and $\beta_3 = -\rho\alpha_K$. In this regression, parameters α_K and ρ are over-identified. There is a testable over-identifying restriction: $\beta_3 = -\beta_1\beta_2$. ■

Time-to build is a key assumption for the consistency of this method. If new investment at period t is productive in the same period t, then we have that: $\phi_{it} = \alpha_K\ k_{i,t+1} + h(\phi_{i,t-1} - \alpha_K\ k_{it}) + \xi_{it}$. Now, the regressor $k_{i,t+1}$ depends on investment at period t and therefore it is correlated with the innovation in productivity ξ_{it}.

Empirical application. Olley and Pakes (1996) study the US telecommunication equipment industry during the period 1974-1987. During this period, the industry experienced substantial technological change and deregulation. There were elimination of barriers to entry. The 1984 Consent Decree was a antitrust decision to divest the industry leader, AT&T. There was substantial entry/exit of plants in the industry.

The authors use annual firm level data on output, capital, labor, and investment from the US Census of manufacturers. They estimate the production function for this industry. Table 3.2 presents their estimates using different estimation methods: OLS, Within-Groups, and the Olley-Pakes method described above. We can see that going from the OLS balanced panel to OLS full sample almost doubles β_K and reduces β_L by 20%. This result provides supportive evidence on the importance of selection bias due to endogenous exit. Controlling for simultaneity further increases β_K and reduces β_L.

Levinsohn and Petrin method

Levinsohn and Petrin (2003) propose an alternative control function method. A main difference between the models and methods by OP and the ones by Levinsohn and Petrin (LP) is that the latter use a control function for the unobserved productivity that comes from inverting the demand materials, instead of inverting the investment equation as in OP method. There are two main motivations for using this alternative control function. First, investment can be responsive only to persistent shocks in TFP; materials is responsive to every shock in TFP. Second, in some datasets there is a substantial fraction of observations with zero investment. At $i_{it} = 0$ (corner solution / extensive margin) there is not invertibility between i_{it} and ω_{it}. This has two implications: loss of efficiency because of the smaller number of observations, and, after estimation of

TABLE VI

ALTERNATIVE ESTIMATES OF PRODUCTION FUNCTION PARAMETERS[a]
(STANDARD ERRORS IN PARENTHESES)

Sample:	Balanced Panel		Full Sample[c, d]						
								Nonparametric F_ω	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation Procedure	Total	Within	Total	Within	OLS	Only P	Only h	Series	Kernel
Labor	.851	.728	.693	.629	.628			.608	
	(.039)	(.049)	(.019)	(.026)	(.020)			(.027)	
Capital	.173	.067	.304	.150	.219	.355	.339	.342	.355
	(.034)	(.049)	(.018)	(.026)	(.018)	(.02)	(.03)	(.035)	(.058)
Age	.002	−.006	−.0046	−.008	−.001	−.003	.000	−.001	.010
	(.003)	(.016)	(.0026)	(.017)	(.002)	(.002)	(.004)	(.004)	(.013)
Time	.024	.042	.016	.026	.012	.034	.011	.044	.020
	(.006)	(.017)	(.004)	(.017)	(.004)	(.005)	(.01)	(.019)	(.046)
Investment	—	—	—	—	.13	—	—	—	—
					(.01)				
Other Variables	—	—	—	—	—	Powers of P	Powers of h	Full Polynomial in P and h	Kernel in P and h
# Obs.[b]	896	896	2592	2592	2592	1758	1758	1758	1758

Figure 3.2: Olley and Pakes (1996): Production Function Estimation

the model, no possibility of recovering the value of TFP for observations with zero investment.

LP consider a Cobb-Douglas production function in terms of labor, capital, and intermediate inputs (materials):

$$y_{it} = \alpha_L \, \ell_{it} + \alpha_K \, k_{it} + \alpha_M \, m_{it} + \omega_{it} + e_{it} \tag{3.53}$$

The investment equation is replaced with the intermediate input demand:

$$m_{it} = f_M \left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right) \tag{3.54}$$

Note that this demand for intermediate inputs is static in the sense that the lagged value m_{it-1} is not an argument in this demand function.

TABLE XI

DECOMPOSITION OF PRODUCTIVITY[a]

(EQUATION (16))

Year	p_t	$\bar{p}_t$	$\Sigma_i \Delta s_{it} \Delta p_{it}$	$\rho(p_t, k_t)$
1974	1.00	0.90	0.01	−0.07
1975	0.72	0.66	0.06	−0.11
1976	0.77	0.69	0.07	−0.12
1977	0.75	0.72	0.03	−0.09
1978	0.92	0.80	0.12	−0.05
1979	0.95	0.84	0.12	−0.05
1980	1.12	0.84	0.28	−0.02
1981	1.11	0.76	0.35	0.02
1982	1.08	0.77	0.31	−0.01
1983	0.84	0.76	0.08	−0.07
1984	0.90	0.83	0.07	−0.09
1985	0.99	0.72	0.26	0.02
1986	0.92	0.72	0.20	0.03
1987	0.97	0.66	0.32	0.10

Figure 3.3: Olley and Pakes (1996): Productivity estimates

Levinsohn and Petrin maintain assumptions OP-2 to OP-4, but replace the assumption of invertibility of the investment function in OP-1 with the following assumption of invertibility of the demand for intermediate inputs:

Assumption LP-1: $f_M(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it})$ is invertible in ω_{it}.

Similarly to the Olley-Pakes method, the key identification restriction in Levinsohn-Petrin method is that the only unobservable variable in the intermediate input demand equation that has cross-sectional variation across firms is the productivity shock ω_{it}. This is *assumption OP-2:* there is no cross-sectional variation in input prices such that $r_{it} = r_t$ for every firm i.

The LP method also proceeds in two steps. The first step consists of the least squares estimation of the parameter α_L and the nonparametric functions $\{\phi_t : t = 1, 2, ..., T\}$ in the semiparametric regression equation:

$$y_{it} \quad = \quad \alpha_L \, \ell_{it} + \phi_t(\ell_{i,t-1}, k_{it}, m_{it}) + e_{it} \tag{3.55}$$

where $\phi_t(\ell_{i,t-1}, k_{it}, m_{it}) = \alpha_K \, k_{it} + f_M^{-1}(\ell_{i,t-1}, k_{it}, m_{it}, r_t)$ and f_M^{-1} represents the inverse function of the demand for intermediate inputs with respect to productivity.

The second step is also in the spirit of OP's second step, but it is substantially different because it requires instrumental variables or GMM estimation. More specifically, the estimates of α_L and ϕ_t are plugged-in, such that we have the regression equation:

$$\phi_{it} \quad = \quad \alpha_K \, k_{it} + \alpha_M \, m_{it} + h\left(\phi_{i,t-1} - \alpha_K \, k_{i,t-1} - \alpha_M \, m_{i,t-1}\right) + \xi_{it} \tag{3.56}$$

The main difference with respect to the OP method is that now the regressor m_{it} is correlated with the error term ξ_{it}. LP propose two alternative approaches to deal with this endogeneity problem. The first approach – described as "unrestricted method" – consists in applying instrumental variables, using lagged values to instrument m_{it} [see GNR (2013) criticism]. The second approach – described as "restricted method" – consists in using the first order condition for profit maximization with respect to materials. Under the assumptions that materials is an static input the firm is a price taker, the first oder condition implies that parameters β_M is equal to the ratio between the firm's cost of materials and its revenue.

Example 3.5: As in equation 3.4 above, suppose that ω_{it} follows the AR(1) process $\omega_{it} = \rho \; \omega_{i,t-1} + \xi_{it}$. Then, $h\left(\omega_{i,t-1}\right) = \rho \, \omega_{i,t-1} = \rho(\phi_{i,t-1} - \alpha_K \, k_{i,t-1} - \alpha_M \, m_{i,t-1})$, and we have that:

$$\phi_{it} = \beta_1 \, k_{it} + \beta_2 \, m_{it} + \beta_3 \, \phi_{i,t-1} + \beta_4 \, k_{i,t-1} + \beta_5 \, m_{i,t-1} + \xi_{it} \tag{3.57}$$

where: $\beta_1 = \alpha_K$, $\beta_2 = \alpha_M$, $\beta_3 = \rho$, $\beta_4 = -\rho\,\alpha_K$, and $\beta_5 = -\rho\,\alpha_M$. We have only three free parameters – α_K, α_M, and ρ – and the model implies four moment conditions: $\mathbb{E}(k_{it} \; \xi_{it}) = 0$; $\mathbb{E}(\phi_{i,t-1} \; \xi_{it}) = 0$; $\mathbb{E}(k_{i,t-1} \; \xi_{it}) = 0$; and $\mathbb{E}(m_{i,t-1} \; \xi_{it}) = 0$. These four moment conditions over-identify the three parameters. ■

Empirical application. LP use plant-level data from 8 different Chilean manufacturing industries during the period 1979-1985.

Ackerberg-Caves-Frazer critique

This critique applies both to Olley-Pakes and Levinsohn-Petrin methods. For the sake of concreteness, we focus here on Olley-Pakes method.

Under Assumptions (*OP-1)* and *(OP-2)*, we can invert the investment equation to obtain the productivity shock $\omega_{it} = f_K^{-1}(\ell_{i,t-1}, k_{it}, i_{it}, r_t)$. Then, we can solve the expression into the labor demand equation, $\ell_{it} = f_L\left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_t\right)$, to obtain the following relationship:

$$\ell_{it} = f_L\left(\ell_{i,t-1}, \; k_{it}, \; f_K^{-1}(\ell_{i,t-1}, k_{it}, i_{it}, r_t), \; r_t\right) = G_t\left(\ell_{i,t-1}, k_{it}, i_{it}\right) \tag{3.58}$$

This expression shows an important implication of Assumptions (*OP-1)* and *(OP-2).* For any cross-section t, there should be a deterministic relationship between employment at period t and the observable state variables $(\ell_{i,t-1}, k_{it}, i_{it})$. In other words, once we condition on the observable variables $(\ell_{i,t-1}, k_{it}, i_{it})$, employment at period t should

TABLE 3

Average Nominal Revenue Shares (Percentage), 1979-85

Industry	Unskilled	Skilled	Materials	Fuels	Electricity
Metals	15.2	8.3	44.9	1.6	1.7
Textiles	13.8	6.0	48.2	1.0	1.6
Food Products	12.1	3.5	60.3	2.1	1.3
Beverages	11.3	6.8	45.6	1.8	1.5
Other Chemicals	18.9	10.1	37.8	1.7	0.7
Printing & Pub.	19.8	10.7	40.1	0.5	1.3
Wood Products	20.6	5.3	47.0	3.0	2.4
Apparel	14.0	4.9	52.4	0.9	0.3

Figure 3.4: Levinsohn and Petrin (2003): Input shares

not have any cross-sectional variability. It should be constant. This implies that in the regression in step 1, $y_{it} = \alpha_L \; \ell_{it} + \phi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it}$, it should not be possible to identify α_L because the regressor ℓ_{it} does not have any sample variability that is independent of the other regressors $(\ell_{i,t-1}, k_{it}, i_{it})$.

Example 3.6: The problem can be simply illustrated using linear functions for the optimal investment and labor demand. Suppose that the inverse function f_K^{-1} is $\omega_{it} = \gamma_1 \; i_{it} + \gamma_2 \; \ell_{i,t-1} + \gamma_3 \; k_{it} + \gamma_4 r_t$; and the labor demand equation is $\ell_{it} = \delta_1 \ell_{i,t-1} + \delta_2 k_{it} + \delta_3 \omega_{it} + \delta_4 r_t$. Then, solving the inverse function f_K^{-1} into the production function, we get:

$$y_{it} = \alpha_L \; \ell_{it} + (\alpha_K + \gamma_3) \; k_{it} + \gamma_1 \; i_{it} + \gamma_2 \; \ell_{i,t-1} + (\gamma_4 r_t + e_{it}) \tag{3.59}$$

TABLE 2
Percent of Usable Observations, 1979-85

Industry	Investment	Fuels	Materials	Electricity
Metals	44.8	63.1	99.9	96.5
Textiles	41.2	51.2	99.9	97.0
Food Products	42.7	78.0	99.8	88.3
Beverages	44.0	73.9	99.8	94.1
Other Chemicals	65.3	78.4	100	96.5
Printing & Pub.	39.0	46.4	99.9	96.8
Wood Products	35.9	59.3	99.7	93.8
Apparel	35.2	34.5	99.9	97.2

Figure 3.5: Levinsohn and Petrin (2003): Frequency of nonzeroes

And solving the inverse function f_K^{-1} into the labor demand, we have that:

$$\ell_{it} = (\delta_1 + \delta_3 \gamma_2)\ell_{i,t-1} + (\delta_2 + \delta_3 \gamma_3) k_{it} + \delta_3 \gamma_1 i_{it} + (\delta_4 + \delta_3 \gamma_4) r_t \tag{3.60}$$

Equation (3.60) shows that, using one year of data (say year t) such that r_t is constant over this cross-sectional sample, there is perfect collinearity between ℓ_{it} and $(\ell_{i,t-1}, k_{it}, i_{it})$. This perfect multi-collinearity implies that it should not be possible to estimate α_L in equation (3.59). In most datasets, we find that this is not the case. That is, we find that ℓ_{it} has cross-sectional variation that is independent of $(\ell_{i,t-1}, k_{it}, i_{it})$. The presence of this independent variation contradicts the model. According to equation (3.60), a simple and plausible way to explain this independent variation is that input prices r_{it} have cross-sectional variation. However, this variation in input prices introduces an endogeneith problem in the estimation of equation (3.59) because the unobservable r_{it} is

TABLE 4
Unrestricted and Restricted Parameter Estimates for 8 Industries
(Bootstrapped Standard Errors in Parentheses)

Input	Industry (ISIC Code) 311	381	321	331	352	322	342	313
Unskilled labor								
	0.138	0.164	0.138	0.206	0.137	0.163	0.192	0.087
	(0.010)	(0.032)	(0.027)	(0.035)	(0.039)	(0.044)	(0.048)	(0.082)
Skilled labor								
	0.053	0.185	0.139	0.136	0.254	0.125	0.161	0.164
	(0.008)	(0.017)	(0.030)	(0.032)	(0.036)	(0.038)	(0.036)	(0.087)
Materials								
	0.703	0.587	0.679	0.617	0.567	0.621	0.483	0.626
	(0.013)	(0.017)	(0.019)	(0.022)	(0.045)	(0.020)	(0.028)	(0.075)
Fuels								
	0.023	0.024	0.041	0.018	0.004	0.0162	0.053	0.087
	(0.004)	(0.008)	(0.012)	(0.018)	(0.020)	(0.016)	(0.014)	(0.027)
Capital								
unrestricted	0.13	0.09	0.08	0.18	0.17	0.10	0.21	0.08
	(0.032)	(0.027)	(0.054)	(0.029)	(0.034)	(0.024)	(0.042)	(0.050)
restricted	0.14	0.09	0.06	0.11	0.15	0.09	0.21	0.07
	(0.011)	(0.02)	(0.019)	(0.025)	(0.034)	(0.039)	(0.045)	(0.11)
Electricity								
unrestricted	0.038	0.020	0.017	0.032	0.017	0.022	0.020	0.012
	(0.021)	(0.010)	(0.024)	(0.028)	(0.032)	(0.014)	(0.024)	(0.022)
restricted	0.011	0.015	0.014	0.021	0.005	0.008	0.011	0.012
No. Obs.	6051	1394	1129	1032	758	674	507	465

Figure 3.6: Levinsohn and Petrin (2003): PF estimate

part of the error term. That is, though there is apparent identification, it seems that this identification is spurious. ■

After pointing out this important problem in the Olley-Pakes model and method, Ackerberg, Caves, and Frazer discuss additional conditions in the model under which the Olley-Pakes estimator is consistent – that is, conditions under which there is no perfect collinearity problem, and the control function approach still solves the endogeneity problem.

For identification, we need some source of exogenous variability in labor demand that is independent of productivity and does not affect capital investment. Ackerberg-Caves-Frazer discuss several possible arguments/assumptions that incorporate this kind of exogenous variability in the model.

Consider a model with the same Cobb-Douglas PF as in the OP model but with the following specification of labor demand and optimal capital investment:

$$
\begin{array}{ll}
(LD') & \ell_{it} = f_L\left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it}^L\right) \\
(ID') & i_{it} = f_K\left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it}^K\right)
\end{array}
\tag{3.61}
$$

Ackerberg-Caves-Frazer propose to maintain Assumptions (OP-1), (OP-3), and (OP-4), and to replace Assumption (OP-2) by the following assumption:

Assumption ACF: Unobserved input prices r_{it}^L and r_{it}^K are such that conditional on $(t,\, i_{it},\, \ell_{i,t-1},\, k_{it})$: (a) r_{it}^L has cross-sectional variation, that is, $var(r_{it}^L\, |t, i_{it}, \ell_{i,t-1}, k_{it}) > 0$; and (b) r_{it}^L and r_{it}^K are independently distributed.

There are different possible interpretations of Assumption ACF. The following list of conditions (a) to (d) is a group of economic assumptions that generate Assumption ACF: (a) the capital market is perfectly competitive and the price of capital is the same for every firm ($r_{it}^K = r_t^K$); (b) there are internal labor markets such that the price of labor has cross-sectional variation; (c) the realization of the cost of labor r_{it}^L occurs after the investment decision takes place, and therefore r_{it}^L does not affect investment; and (d) the idiosyncratic labor cost shock r_{it}^L is not serially correlated such that lagged values of this shock are not state variables for the optimal investment decision. Aguirregabiria and Alonso-Borrego (2014) consider similar assumptions for the estimation of a production function with physical capital, permanent employment, and temporary employment.

Other identifying conditions: Quasi-fixed inputs

Consider a Cobb-Douglas PF with labor and capital as the only inputs. Suppose that the OP assumptions hold such that ℓ_{it} is perfectly collinear with $\phi_t(\ell_{i,t-1}, k_{it}, i_{it})$. If both capital and labor are quasi-fixed inputs, then it is possible to use a control function method in the spirit of OP or LP to identify/estimate β_L and β_K. Or in other words, this model has moment conditions that identify β_L and β_K (Wooldridge, 2009).

In the first step we have:

$$
\begin{aligned}
y_{it} &= \beta_L\, \ell_{it} + \phi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it} \\
&= \beta_L\, g_t(\ell_{i,t-1}, k_{it}, i_{it}) + \phi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it} \\
&= \psi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it}
\end{aligned}
$$

In this first step, we estimate $\psi_t(\ell_{i,t-1}, k_{it}, i_{it})$ nonparametrically. In the second step, given ψ_{it}, and taking into account that $\psi_{it} = \beta_L\, \ell_{it} + \beta_K\, k_{it} + \omega_{it}$, and $\omega_{it} = h(\omega_{i,t-1}) + \xi_{it}$, we have that:

$$
\psi_{it} \;=\; \beta_L\, \ell_{it} + \beta_K\, k_{it} + h\left(\psi_{it} - \beta_L\, \ell_{it-1} + \beta_K\, k_{it-1}\right) + \xi_{it}
$$

In this second step, ℓ_{it} is correlated with ξ_{it}, but $(k_{it}, \psi_{it}, \ell_{it-1},\, k_{it-1})$ are not, and (ℓ_{it-2}, k_{it-2}) can be used to instrument ℓ_{it}. This approach is in the same spirit as the Dynamic Panel Data (DPD) methods of Arellano-Bond and Blundell-Bond. This approach cannot be applied if some inputs (for instance, materials) are perfectly flexible. The PF coefficient parameter of the flexible inputs cannot be identified from the moment conditions in the second step.

Other identifying conditions: F.O.C. for flexible inputs

Klette and Griliches (1996), Doraszelski and Jaumandreu (2013), and Gandhi, Navarro, and Rivers (2017) propose combining conditions from the PF with conditions from the demand of variable inputs. This approach requires the price of the variable input to be observable to the researcher, though this price may not have cross-sectional variation across firms.

Note that in the LP method, the function that relates m_{it} with the state variables is just the condition "VMP of materials equal to price of materials". The parameters in this condition are the same as in the PF. This approach takes these restrictions into account.

For the CD-PF, with materials as flexible input, we have that:

$$\begin{aligned} (PF) \quad & y_{it} = \beta_L\, \ell_{it} + \beta_K\, k_{it} + \beta_M\, m_{it} + \omega_{it} + e_{it} \\ (FOC) \quad & p_t - p_t^M = \ln(\beta_M) + \beta_L\, \ell_{it} + \beta_K\, k_{it} + (\beta_M - 1) m_{it} + \omega_{it} \end{aligned} \tag{3.62}$$

The difference between these two equations is:

$$\ln(s_{it}^M) \equiv \ln\left(\frac{P_t^M M_{it}}{P_t Y_{it}}\right) = \ln(\beta_M) + e_{it}$$

where s_{it}^M is the ratio between materials expenditure and revenue. The parameter(s) of the flexible inputs are identified from the expenditure-share equations. The parameter(s) of the quasi-fixed inputs are identified using the dynamic conditions described above.

Gandhi, Navarro, and Rivers (2017) show that this approach can be extended in two important ways: (1) to a nonparametric specification of the production function: $y_{it} = f(\ell_{it}, k_{it}, m_{it}) + \omega_{it} + e_{it}$; and (2) to a model with monopolistic competition – instead of perfect competition – with an isoelastic product demand. Their approach to get extension (2) relies on an important assumption: there is not any bias or missing parameter in the marginal cost of the flexible input. For instance, suppose that the marginal cost of material were $MC_{Mt} = P_t^M\, \tau$, then our estimate of β_M would actually estimate $\beta_M\, \tau$.

3.4.5 Endogenous exit

Semiparametric selection models

The estimator in Olley and Pakes (1996) controls for selection bias due to endogenous exit of firms. Before describing their approach, it can be helpful to describe some general features of semiparametric selection models.

Consider a selection model with outcome equation,

$$y_i = \begin{cases} x_i\, \beta + \varepsilon_i & \text{if} \quad d_i = 1 \\ \text{unobserved} & \text{if} \quad d_i = 0 \end{cases} \tag{3.63}$$

and selection equation

$$d_i = \begin{cases} 1 & \text{if} \quad h(z_i) - u_i \geq 0 \\ 0 & \text{if} \quad h(z_i) - u_i < 0 \end{cases} \tag{3.64}$$

where x_i and z_i are exogenous regressors; (u_i, ε_i) are unobservable variables independently distributed of (x_i, z_i); and $h(.)$ is a real-valued function. We are interested in the consistent estimation of the vector of parameters β. We would like to have an estimator that does not rely on parametric assumptions on the function h or on the distribution of the unobservables.

The outcome equation can be represented as a regression equation: $y_i = x_i\ \beta + \varepsilon_i^{d=1}$, where $\varepsilon_i^{d=1} \equiv \{\varepsilon_i | d_i = 1\} = \{\varepsilon_i | u_i \leq h(z_i)\}$. Or similarly,

$$y_i = x_i\beta + \mathbb{E}(\varepsilon_i^{d=1} | x_i, z_i) + \tilde{\varepsilon}_i \tag{3.65}$$

where $\mathbb{E}(\varepsilon_i^{d=1} | x_i, z_i)$ is the selection term. The new error term, $\tilde{\varepsilon}_i$, is equal to $\varepsilon_i^{d=1} - \mathbb{E}(\varepsilon_i^{d=1} | x_i, z_i)$ and, by construction, it has mean zero and it is mean-independent of (x_i, z_i). The selection term is equal to $\mathbb{E}\left(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i)\right)$. Given that u_i and ε_i are independent of (x_i, z_i), it is simple to show that the selection term depends on the regressors only through the function $h(z_i)$: that is, $\mathbb{E}\left(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i)\right) = g(h(z_i))$. The form of the function g depends on the distribution of the unobservables, and it is unknown if we adopt a nonparametric specification of that distribution. Therefore, we have the following partially linear model: $y_i = x_i\beta + g(h(z_i)) + \tilde{\varepsilon}_i$.

Define the *propensity score* P_i as:

$$P_i \equiv \Pr\left(d_i = 1 \mid z_i\right) = F_u\left(h(z_i)\right) \tag{3.66}$$

where F_u is the CDF of u. Note that $P_i = \mathbb{E}\left(d_i \mid z_i\right)$, and therefore we can estimate propensity scores nonparametrically using a Nadaraya-Watson kernel estimator or other nonparametric methods for conditional means. If u_i has unbounded support and a strictly increasing CDF, then there is a one-to-one invertible relationship between the propensity score P_i and $h(z_i)$. Therefore, the selection term $g(h(z_i))$ can be represented as $\lambda(P_i)$, where the function λ is unknown. The selection model can be represented using the partially linear model:

$$y_i = x_i\beta + \lambda(P_i) + \tilde{\varepsilon}_i. \tag{3.67}$$

A sufficient condition for the identification of β (without a parametric assumption on λ) is that $\mathbb{E}\left(x_i\ x_i' \mid P_i\right)$ has full rank. Given equation (3.67) and nonparametric estimates of propensity scores, we can estimate β and the function λ using standard estimators for partially linear model such as sieve methods, kernel-based methods like Robinson (1988), or differencing methods like Yatchew (2003).

Olley and Pakes method to control for endogenous exit

Now, we describe the Olley-Pakes procedure for the estimation of the production function taking into account endogenous exit. The first step of the method (that is, the estimation of α_L) is not affected by the selection problem because we are controlling for ω_{it} using a control function approach. However, there is endogenous selection in the second step of the method.

For simplicity consider that the productivity shock follows an AR(1) process: $\omega_{it} = \rho\ \omega_{i,t-1} - \xi_{it}$. Then, the "outcome" equation is:

$$\phi_{it} = \begin{cases} \alpha_K\ k_{it} + \rho\ \phi_{i,t-1} + (-\rho\,\alpha_K)\ k_{i,t-1} + \xi_{it} & \text{if} \quad d_{it} = 1 \\ \\ \text{unobserved} & \text{if} \quad d_{it} = 0 \end{cases} \tag{3.68}$$

The exit/stay decision is: $\{d_{it} = 1\}$ iff $\{\omega_{it} \geq v(\ell_{it-1}, k_{it})\}$. Taking into account that $\omega_{it} = \rho\,\omega_{i,t-1} + \xi_{it}$, and that $\omega_{i,t-1} = \phi_{i,t-1} - \alpha_K\ k_{it-1}$, we have that the condition $\{\omega_{it} \geq v(\ell_{it-1}, k_{it})\}$ is equivalent to:

$$d_{it} = \begin{cases} 1 & \text{if} \quad \xi_{it} \leq v(\ell_{it-1}, k_{it}) - \rho(\phi_{i,t-1} - \alpha_K k_{it-1}) \\ \\ 0 & \text{if} \quad \xi_{it} > v(\ell_{it-1}, k_{it}) - \rho(\phi_{i,t-1} - \alpha_K k_{it-1}) \end{cases} \tag{3.69}$$

The propensity score is $P_{it} \equiv \mathbb{E}\,(d_{it} \mid \ell_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1})$ such that P_{it} is a function of $(\ell_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1})$. The equation controlling for selection is:

$$\phi_{it} = \beta_1\ k_{it} + \beta_2\ \phi_{i,t-1} + \beta_3\ k_{i,t-1} + \lambda\ (P_{it}) + \tilde{\xi}_{it} \tag{3.70}$$

where $\beta_1 = \alpha_K$, $\beta_2 = \rho$, and $\beta_2 = -\rho\,\alpha_K$. By construction, $\tilde{\xi}_{it}$ is mean independent of k_{it}, k_{it-1}, $\phi_{i,t-1}$, and P_{it}. We can estimate parameters β_1, β_2, and β_3 and function $\lambda\ (.)$ in the regression equation (3.70) by using standard methods for semiparametric partially linear models.

In reality, the method to control for selection in Olley and Pakes (1996) is a bit more involved because the stochastic process for the productivity shock is nonparametrically specified: $\omega_{it} = h(\omega_{i,t-1}) - \xi_{it}$. Therefore, the regression model is:

$$\phi_{it} = \alpha_K\ k_{it} + h\,(\phi_{i,t-1} - \alpha_K\ k_{i,t-1}) + \lambda\ (P_{it}) + \tilde{\xi}_{it} \tag{3.71}$$

such that we have two nonparametric functions, $h(.)$ and $\lambda\ (.)$. However, the identification and estimation of the model proceeds in a very similar way. For instance, we can consider a polynomial approximation to these nonparametric functions and estimate the parameters by least squares.

3.5 Determinants of productivity

3.5.1 What determines productivity?

There are large and persistent differences in TFP across firms. This evidence is ubiquitous even within narrowly defined industries and products.

Large TFP differences. A commonly used measure of the heterogeneity in TFP across firms is the ratio between the 90th to 10th percentile in the (cross-sectional) distribution. Using data from U.S. manufacturing industries – 4-digit SIC industries – Syverson (2004) reports that the ratio between the 90th to 10th percentile is on average equal 1.92. For industries in Denmark, Fox and Smeets (2011) report an average ratio of 3.75. This ratio is even larger in developing countries. For instance, Hsieh and Klenow (2009) report average ratios above 5 for China and India.

Persistent TFP differences. A statistic that is commonly used to measure this persistence is the slope parameter in the simple regression of the log-TFP of a firm on its log-TFP in the previous year. Most studies report estimates of this autoregressive coefficient between 0.6 to 0.8.

Relevant TFP differences. Studies show that these differences in productivity have an important impact on different decisions such as market exit, exporting, or investing in R&D.

Why do firms differ in their productivity levels? What mechanism can support such large differences in productivity in market equilibrium? Can producers control the factors that influence productivity or are they purely external effects of the environment? If firms can partly control their TFP, what type of choices increase it?

3.5.2 TFP dispersion in equilibrium

Following Syverson (2004), we present here a very stylized model to illustrate how dispersion in TFP within the same industry is perfectly possible in equilibrium, and that it can be driven by very common forces that exist in most markets. Consider a homogeneous product industry and let the profit of a firm be $\pi_i = R(q_i, d) - C(q_i, A_i, w) - F$, where: $R(q_i, d)$ is the revenue function; $C(q_i, A_i, w)$ is the variable cost function; q_i is output; A_i is TFP; d is the state of the demand; w represents input prices; and F is the fixed cost. Firms with different TFPs coexist in the same market if it is not optimal for the firm with the largest TFP to produce all the quantity demanded in the market. The key necessary and sufficient condition for this to occur is that the profit function of a firm must be strictly concave in output q_i. That is, either the revenue function $R(.)$ is strictly concave in q_i (because market power and oligopoly competition), or the cost function $C(.)$ is strictly convex in q_i (because diseconomies of scale or fixed inputs).

For instance, consider a perfectly competitive industry such that the revenue function is $R(q_i, d) = P(d)\ q_i$, that is, it is linear in output q_i. Suppose that there are decreasing returns to scale such that the cost function $C(q_i, A_i, w)$ is strictly convex in q_i. Then, even in this perfectly competitive industry we have that the firm with the highest TFP does not produce all the output demanded in the market.

Consider the – somehow – opposite case. The industry has a constant returns to scale technology such that the cost function is $C(q_i, A_i, w) = c(A_i, w)\ q_i$, that is, it is linear in output. This industry is characterized by Cournot competition. This implies that the revenue function is $R(q_i, d) = P(q_i + Q_{-i}, d)\ q_i$, where $P(.)$ is the inverse demand function. This revenue function is strictly concave in q_i – provided the demand curve is downward sloping.

More formally, the equilibrium in the industry can be described by two types of conditions. At the intensive margin, optimal $q_i^* = q^*[A_i, d, w]$ is such that:

$$MR_i \equiv \frac{\partial R(q_i, A_i, d)}{\partial q_i} = \frac{\partial C(q_i, A_i, w)}{\partial q_i} \equiv MC_i \tag{3.72}$$

At the extensive margin, a firm is active in the market if:

$$R(q^*[A_i, d, w], A_i, d) - C(q^*[A_i, d, w], A_i, w) - F \geq 0 \tag{3.73}$$

If variable profit is strictly concave, this equilibrium can support firms with different TFPs. It is not optimal for the firm with the highest TFP to provide all the output in the industry. Firms with different TFPs – above a certain threshold value – coexist and compete in the same market.

3.5.3 How can firms improve their TFP?

There are multiple ways in which firms can affect their TFP. The following is a list of practices – non exhaustive – that firms can follow to increase their TFP, as well as empirical papers that have found evidence for these effects.

(i) Human resources and managerial practices: Bloom and VanReenen (2007); Ichniowski and Shaw (2003).
(ii) Learning-by-doing: Benkard (2000).
(iii) Organizational structure such as outsourcing or (the opposite) vertical integration.
(iv) Adoption of new technologies: Bresnahan, Brynjolfsson, and Hitt (2002).
(v) Investment in R&D and product and process innovation: Griliches (1979); Doraszelski and Jaumandreu (2013).

There is a long literature linking R&D investment and innovation to productivity. Multiple studies show evidence that R&D and innovation are important factors to explain firm heterogeneity in the level and growth of TFP. As usual, the main difficulty in these studies comes from separating causation from correlation. In section 3.6, we review models, methods, and datasets in different empirical applications dealing with the causal effect of R&D and/or innovation on TFP.

3.6 R&D and productivity

Investment in R&D and innovation is expensive. Investors – firms and governments – are interested in measuring its private and social returns. Process R&D is directed towards invention of new methods of production. Product R&D tries to create new and improved goods. Both process and product R&D can increase a firm's TFP. It can have also spillover effects in other firms: competition spillovers, and/or knowledge spillovers.

3.6.1 Knowledge capital model

In an influential paper, Griliches (1979) proposes a model and method to measure knowledge capital, that is, the capital generated by investments in R&D that is intangible and different from physical capital. This model is often referred to as the *knowledge capital model*, and many studies have used it to measure the returns to R&D.

The model is based on the estimation of a production function. Consider a Cobb-Douglas PF in logs:

$$y_{it} = \beta_L \, \ell_{it} + \beta_K \, k_{it} + \beta_M \, m_{it} + \beta_R \, k_{it}^R + \omega_{it} + e_{it}$$

where k_{it} is the logarithm of the stock of physical capital, and k_{it}^R is logarithm of the of stock of knowledge capital. A major difficulty is the measurement of the stock of knowledge capital. Let R_{it} be the investment in R&D of firm i at period t, and let K_{it}^R be the firm's stock of knowledge capital: that is, $K_{it}^R = \exp\{k_{it}^R\}$. Suppose that the researcher observes R_{it} for $t = 1, 2, ... T_i$. However, the researcher does not observe the stock of knowledge capital. Griliches (1979) proposes the following *perpetual inventory method* to obtain the sequence of stocks K_{it}^R for $t = 1, 2, ... T_i$. Suppose that the stock follows the transition rule:

$$K_{it}^R = (1 - \delta_R) \; K_{i,t-1}^R + R_{it}$$

where δ_R is the depreciation rate of knowledge capita. Given values for δ_R and for the the initial condition K_{i0}^R, we can use the data on R&D investments to construct the sequence $\{K_{it}^R : t = 1, 2, ... T_i\}$.

How to choose δ_R and K_{i0}^R? It is difficult to know the true value of the rate of technological obsolescence, δ_R: it can be endogenous, and vary across industries and firms. Researchers have considered different approaches to estimate this depreciation rate: using patent renewal data (Pakes and Schankerman ,1984; Pakes ,1986); or using Tobin's Q model (Hall ,2007). The estimates of this depreciation rate in the literature range between 10% and 35%. Different authors have performed sensitivity analysis on the estimates of β_R for different value of δ_R. They report small differences, if any, in the estimate of β_R when δ_R varies between 8% and 25%.

3.6.2 An application

Doraszelski and Jaumandreu (2013) propose and estimate a model that extends the knowledge capital model in important ways. In their model, TFP and Knowledge capital (KC) are unobservables to the researcher. They follow stochastic processes that are endogenous and depend on (observable) R&D investments. The model accounts for uncertainty and heterogeneity across firms in the relationship between R&D and TFP. The model takes into account that the outcome of R&D investments is subject to a high degree of uncertainty.

For the estimation of the structural parameters in the PF the and stochastic process of KC, the authors exploit first order conditions for variable inputs.

Model

Consider the production function in logs:

$$y_{it} = \beta_L\, \ell_{it} + \beta_K\, k_{it} + \beta_M\, m_{it} + \omega_{it} + e_{it} \tag{3.74}$$

log-TFP ω_{it} follows a stochastic process with transition probability $p(\omega_{it+1} \mid \omega_{it},\, r_{it})$, where r_{it} is log-R&D expenditure. Every period t a firm chooses static inputs (ℓ_{it}, m_{it}) and investment in physical capital and R&D (i_{it}, r_{it}) to maximize its value.

$$V(s_{it}) = \max_{i_{it}, r_{it}} \left\{ \pi(s_{it}) - c^{(1)}(i_{it}) - c^{(2)}(r_{it}) + \rho\, \mathbb{E}\left[V(s_{it+1}) | s_{it}, i_{it}, r_{it}\right] \right\} \tag{3.75}$$

with $s_{it} = (k_{it}, \omega_{it}$, input prices $[w_{it}]$, demand shifters $[d_{it}])$.

The Markov structure of log-TFP implies:

$$\omega_{it} = \mathbb{E}\left[\omega_{it} \mid \omega_{it-1},\, r_{it-1}\right] + \xi_{it} = g(\omega_{it-1},\, r_{it-1}) + \xi_{it}$$

where $\mathbb{E}\left[\xi_{it} \mid \omega_{it-1},\, r_{it-1}\right] = 0$. The *productivity innovation* ξ_{it} captures two sources of uncertainty for the firm: the uncertainty linked to the evolution of TFP; and the uncertainty inherent to R&D – for instance, chances of making a new discovery, its degree of applicability, successful implementation, etc.

The authors' identification approach exploits static marginal conditions of optimality. Obtaining these conditions requires an assumption about competition. The authors assume monopolistic competition. More precisely, they assume the following form for the marginal revenue:

$$MR_{it} = P_{it} \left(1 - \frac{1}{\eta(p_{it}, d_{it})}\right) \tag{3.76}$$

where $\eta(p_{it}, d_{it})$ is the price elasticity of demand for firm i, that is, monopolisitc competition.

The marginal condition of optimality for labor provides a closed-form expression for labor demand. Solving for log-TFP in the labor demand equation, we get:

$$\begin{aligned}\omega_{it} &= \lambda - \beta_K\, k_{it} + (1-\beta_L-\beta_M)\, \ell_{it} + (1-\beta_M)\, (w_{it}-p_{it}) \\ &+ \beta_M\, (p_{Mit}-p_{it}) - \ln\left(1-\frac{1}{\eta(p_{it},d_{it})}\right)\end{aligned} \tag{3.77}$$

We represent the RHS as $h(x_{it},\beta)$, such that $\omega_{it} = h(x_{it},\beta)$, with $x_{it} = (k_{it},\, \ell_{it},\, w_{it},\, p_{Mit},\, p_{it},\, d_{it})$.

Identification and estimation

Combining the PF equation with the stochastic process for TFP, and the marginal condition for optimal labor, we have the equation:

$$y_{it} = \beta_L\, \ell_{it} + \beta_K\, k_{it} + \beta_M\, m_{it} + g\,[h(x_{it-1},\beta),\ r_{it-1}] + \xi_{it} + e_{it} \tag{3.78}$$

From the marginal condition for labor we have:

$$h(x_{it},\beta) = g\,[h(x_{it-1},\beta),\ r_{it-1}] + \xi_{it} \tag{3.79}$$

The "parameters" in this system of equations are: β_L, β_K, β_M, g, and η. The unobservables ξ_{it} and e_{it} are mean independent of any observable variable at period $t-1$ or before. Therefore, x_{it-1} and r_{it-1} are exogenous w.r.t. $\xi_{it}+e_{it}$. Capital stock k_{it} is also uncorrelated to the error term because of time-to-build. However, we need to instrument the regressors ℓ_{it} and m_{it}.

To see that the parameters of the model are identified, it is convenient to consider a simplified version with: $\beta_K = \beta_M = 1/\eta = 0$ and $g\,[\omega_{t-1},\ r_{t-1}] = \rho_\omega\ \omega_{t-1} + \rho_r\ r_{t-1}$. Then, we have:

$$y_{it} = \beta_L\, \ell_{it} + \rho_\omega\,[(1-\beta_L)\,\ell_{it-1} + w_{it-1} - p_{it-1}] + \rho_r\, r_{it-1} + \xi_{it} + e_{it} \tag{3.80}$$

By using the vector of instruments $Z_{it} = (y_{it-1},\, \ell_{it-1},\, w_{it-1} - p_{it-1}\ ,\, r_{it-1})$, we have that the moment conditions $\mathbb{E}\,[Z_{it}\ (\xi_{it}+e_{it})] = 0$ identify β_L, ρ_ω, ρ_r. Given the identification of these parameters, we know $\omega_{it} = h(x_{it},\beta) = (1-\beta_L)\ell_{it} + (w_{it}-p_{it})$. The model implies, that:

$$\xi_{it} = h(x_{it},\beta) - \rho_\omega\ h(x_{it},\beta) - \rho_r\ r_{it-1} \tag{3.81}$$

such that ξ_{it} is identified, and so is its variance $Var(\xi_{it})$ that represents uncertainty in the link between R&D and TFP.

The instrument $w_{it-1} - p_{it-1}$ plays a very important role in the identification of the model. Without variation in lagged (real) input prices the model is not identified. But note that the model does not use contemporaneous input prices as instruments because they can be correlated with the innovation ξ_{it}.

Data

The papers uses panel data of Spanish manufacturing firms ($N = 1,870$ firms) from ten industries (SIC 2-digits). The dataset has annual frequency and it covers the period 1990 – 1999 (max $T_i = 10$). This was a period of rapid growth in output and physical capital, coupled with stagnant employment. Table 3.7 presents some descriptive statistics. R&D intensity = R&D expenditure / Sales: the average among all firms is 0.6% (smaller than in France, Germany, or UK, $> 2\%$). R&D intensity among performers (column 13) is between 1% and 3.5%.

TABLE 1
Descriptive statistics

					Rates of growth[a]					With R&D[b]			
Industry	Obs.[a]	Firms[a]	Entry[a] (%)	Exit[a] (%)	Output (s. d.)	Capital (s. d.)	Labour (s. d.)	Materials (s. d.)	Price (s. d.)	Obs. (%)	Stable (%)	Occas. (%)	R&D inten. (s. d.)
	(1)	(2)	(3)	(4)	(5)	(7)	(6)	(8)	(9)	(10)	(11)	(12)	(13)
1. Metals and metal products	1235	289	88	17	0.050	0.086	0.010	0.038	0.012	420	63	72	0.0126
			(30.4)	(5.9)	(0.238)	(0.278)	(0.183)	(0.346)	(0.055)	(34.0)	(21.8)	(24.9)	(0.0144)
2. Non-metallic minerals	621	131	20	15	0.037	0.062	−0.001	0.039	0.010	186	16	41	0.0100
			(15.3)	(11.5)	(0.208)	(0.238)	(0.141)	(0.308)	(0.059)	(30.0)	(12.2)	(31.3)	(0.0211)
3. Chemical products	1218	275	64	15	0.068	0.093	0.007	0.054	0.007	672	124	55	0.0268
			(23.3)	(5.5)	(0.196)	(0.238)	(0.146)	(0.254)	(0.061)	(55.2)	(45.1)	(20.0)	(0.0353)
4. Agric. and ind. machinery	576	132	36	6	0.059	0.078	0.010	0.046	0.013	322	52	35	0.0219
			(27.3)	(4.5)	(0.275)	(0.247)	(0.170)	(0.371)	(0.032)	(55.9)	(39.4)	(26.5)	(0.0275)
6. Transport equipment	637	148	39	10	0.087	0.114	0.011	0.087	0.007	361	62	35	0.0224
			(26.4)	(6.8)	(0.354)	(0.255)	(0.207)	(0.431)	(0.037)	(56.7)	(41.9)	(23.6)	(0.0345)
7. Food, drink, and tobacco	1408	304	47	22	0.025	0.094	−0.003	0.019	0.022	386	56	64	0.0071
			(15.5)	(7.2)	(0.224)	(0.271)	(0.186)	(0.305)	(0.065)	(27.4)	(18.4)	(21.1)	(0.0281)
8. Textile, leather, and shoes	1278	293	77	49	0.020	0.059	−0.007	0.012	0.016	378	39	66	0.0152
			(26.3)	(16.7)	(0.233)	(0.235)	(0.192)	(0.356)	(0.040)	(29.6)	(13.3)	(22.5)	(0.0219)
9. Timber and furniture	569	138	52	18	0.038	0.077	0.014	0.029	0.020	66	7	18	0.0138
			(37.7)	(13.0)	(0.278)	(0.257)	(0.210)	(0.379)	(0.035)	(12.6)	(5.1)	(13.8)	(0.0326)
10. Paper and printing products	665	160	42	10	0.035	0.099	−0.000	0.026	0.019	113	21	25	0.0143
			(26.3)	(6.3)	(0.183)	(0.303)	(0.140)	(0.265)	(0.089)	(17.0)	(13.1)	(13.8)	(0.0250)

Figure 3.7: Doraszelski and Jaumandreu (2013): Descriptive statistics

Estimates

Figure 3.8 presents parameter estimates. Comparing GMM and OLS estimates, we can see that correcting for endogeneity has the expected implications. For instance, β_L and β_M decline, and β_K increases. There are not big differences in the β estimates across industries. The test of of over-identifying restrictions (OIR) cannot reject the validity of the instruments with a 5% confidence level. The test of parameter restrictions (in the two equations) can reject these restrictions at 5% level only in 2 out of 10 industries.

As for the stochastic process for TFP, the model where TFP doesn't depend on R&D is clearly rejected. Models with linear effects or without complementarity between ω_{t-1} and r_{t-1} are rejected. $Var(e)$ is approx. equal to $Var(\omega)$ in most industries. $Var(\xi)/Var(\omega)$ is between 30% and 75%. The authors find significant evidence supporting that the effect of R&D on TFP is stochastic and uncertain to forms. There are significant differences across industries in the magnitude of this uncertainty.

The authors test three different versions of the knowledge capital (KC) model. For the basic KC model (where $\omega_{it} + e_{it} = \beta_R \, k_{it}^R + e_{it}$), the authors can reject this model for all industries. The second model is Hall and Hayashi (1989) and Klette (1996) KC model, where $\omega_{it} = \sigma \, \omega_{it-1} + (1-\sigma) \, r_{it-1} + \xi_{it}$. This model is rejected at 5% level in 8 industries, and at 7% level in all the industries. The third KC model is characterized by the equation $k_{it}^R + \omega_{it} + e_{it}$, and ω_{it} follows an exogenous Markov process. This model is ejected at 5% levle in 2 industries, and at 10% in 6 industries.

TABLE 2
Production function estimates and specification tests

Industry	OLS[a]			GMM[a]			Overidentifying restrictions test		Parameter restrictions test	
	Capital (std. err.)	Labour (std. err.)	Materials (std. err.)	Capital (std. err.)	Labour (std. err.)	Materials (std. err.)	$\chi^2(df)$	p val.	$\chi^2(3)$	p val.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. Metals and metal products	0.109 (0.013)	0.252 (0.022)	0.642 (0.020)	0.106 (0.014)	0.111 (0.031)	0.684 (0.011)	62.553 (51)	0.129	11.666	0.009
2. Non-metallic minerals	0.096 (0.021)	0.275 (0.034)	0.655 (0.028)	0.227 (0.014)	0.137 (0.016)	0.633 (0.014)	50.730 (47)	0.329	6.047	0.109
3. Chemical products	0.060 (0.010)	0.239 (0.021)	0.730 (0.020)	0.132 (0.015)	0.122 (0.026)	0.713 (0.011)	48.754 (47)	0.402	0.105	0.991
4. Agric. and ind. machinery	0.051 (0.017)	0.284 (0.038)	0.671 (0.027)	0.079 (0.015)	0.281 (0.029)	0.642 (0.013)	45.833 (44)	0.396	1.798	0.615
6. Transport equipment	0.080 (0.023)	0.289 (0.033)	0.636 (0.046)	0.117 (0.015)	0.158 (0.023)	0.675 (0.016)	40.296 (47)	0.745	0.414	0.937
7. Food, drink, and tobacco	0.094 (0.014)	0.177 (0.016)	0.739 (0.016)	0.068 (0.014)	0.129 (0.024)	0.766 (0.008)	61.070 (46)	0.068	8.866	0.031
8. Textile, leather, and shoes	0.059 (0.010)	0.335 (0.024)	0.605 (0.019)	0.057 (0.011)	0.313 (0.016)	0.593 (0.013)	66.143 (51)	0.075	4.749	0.191
9. Timber and furniture	0.079 (0.019)	0.283 (0.029)	0.670 (0.029)	0.131 (0.009)	0.176 (0.017)	0.697 (0.011)	44.951 (43)	0.390	0.618	0.892
10. Paper and printing products	0.092 (0.016)	0.321 (0.029)	0.621 (0.025)	0.121 (0.013)	0.249 (0.025)	0.617 (0.014)	51.371 (42)	0.152	5.920	0.118

Figure 3.8: Doraszelski and Jaumandreu (2013): PF estimates

Counterfactuals on R&D and TFP.

The distribution of TFP with R&D stochastically dominates distribution without R&D. Differences in means are between 3% and 5% for all industries and firm sizes, except for small firms in industries with low observed R&D intensity.

The magnitude of the elasticity of TFP with respect to R&D has considerable variation between and within industries. Its average is 0.015. The elasticity of TFP with respect to lagged TFP shows substantial persistence, but there is also considerable heterogeneity between and within industries. Non-performers have a higher degree of persistence than performers. The degreee of persistence is negatively related to the degree of uncertainty.

In summary, the authors model TFP growth as the consequence of R&D expenditures with uncertain outcomes. Results show that this model can explain better the relationship between TFP and R&D than standard Knowledge Capital models without uncertainty and non-linearity. R&D is a major determinant of the differences in TFP across firms and of their evolution. They also find that firm-level uncertainty in the outcome of R&D is considerable. Their estimates suggest that engaging in R&D roughly doubles the degree of uncertainty in the evolution of a producer's TFP.

3.7 Exercises

3.7.1 Exercise 1

Consider an industry for an homogeneous product. Firms use capital and labor to produce output according to a Cobb-Douglas technology with parameters α_L and α_K and Total Factor Productivity (TFP) A.

Industry	Exogeneity test $\chi^2(10)$	p val.	Separability test $\chi^2(3)$	p val.	$\frac{Var(e_{jt})}{Var(\omega_{jt})}$	$\frac{Var(\xi_{jt})}{Var(\omega_{jt})}$
	(1)	(2)	(3)	(4)	(5)	(6)
1. Metals and metal products	65.55	0.000	16.360	0.001	0.735	0.407
2. Non-metallic minerals	92.65	0.000	13.027	0.005	0.842	0.410
3. Chemical products	40.79	0.000	8.647	0.034	0.749	0.244
4. Agric. and ind. machinery	51.88	0.000	11.605	0.009	1.410	0.505
6. Transport equipment	56.85	0.000	18.940	0.000	1.626	0.524
7. Food, drink, and tobacco	38.29	0.000	7.186	0.066	1.526	0.300
8. Textile, leather, and shoes	29.91	0.001	18.417	0.000	1.121	0.750
9. Timber and furniture	118.17	0.000	32.260	0.000	1.417	0.515
10. Paper and printing products	59.73	0.000	23.249	0.000	0.713	0.433

Figure 3.9: Doraszelski and Jaumandreu (2013): Stochastic process TFP

Question 1.1. Write the expression for this Cobb-Douglas production function (PF).

Suppose that firms are price takers in the input markets for labor and capital. Let W_L and W_K be the price of labor and capital, respectively. Capital is a fixed input such that the fixed cost for a firm, say i, is $FC_i = W_K\ K_i$. The *variable cost function*, $VC(Y)$, is defined as the minimum cost of labor to produce an amount of output Y.

Question 1.2. Derive the expression for the variable cost function of a firm in this industry. Explain your derivation. [Hint: Given that capital is fixed and there is only one variable input, the minimization problem is trivial. The PF implies that there is only one possible amount of labor that give us a certain amount of output].

Question 1.3. Using the expressions for the fixed cost and for the variable cost function in Q1.2:

(a) Explain how an increase in the amount of capital affects the fixed cost and the variable cost of a firm.

(b) Explain how an increase in TFP affects the fixed cost and the variable cost.

Suppose that the output market in this industry is competitive: firms are price takers. The demand function is linear with the following form: $P = 100 - Q$, where P and Q are the industry price and total output, respectively. Suppose that $\alpha_L = \alpha_K = 1/2$, and the value of input prices are $W_L = 1/2$ and $W_K = 2$. Remember that firms' capital stocks are fixed (exogenous), and for simplicity suppose that all the firms have the same capital stock $K = 1$.

Question 1.4. Using these primitives, write the expression for the profit function of a firm (revenue, minus variable cost, minus fixed cost) as a function of the market price, P, the firm's output, Y_i, and its TFP, A_i.

Knowledge capital model tests

Basic		Generalization 1		Generalization 2	
$N(0,1)$	p val.	$N(0,1)$	p val.	$N(0,1)$	p val.
(7)	(8)	(9)	(10)	(11)	(12)
−2.815	0.002	−2.431	0.008	−1.987	0.023
−2.041	0.021	−1.541	0.062	−0.784	0.216
−3.239	0.001	−2.090	0.018	−1.400	0.081
−2.693	0.004	−1.588	0.056	−1.493	0.068
−2.317	0.010	−2.042	0.021	−1.821	0.034
−3.263	0.001	−2.499	0.006	−0.901	0.184
−2.770	0.003	−1.788	0.037	−1.488	0.068
−2.510	0.006	−2.097	0.018	−1.028	0.152
−3.076	0.001	−2.210	0.014	−1.595	0.055

Figure 3.10: Doraszelski and Jaumandreu (2013): Testing Knowledge capital

Question 1.5. Using the condition "price equal to marginal cost", obtain the optimal amount of output of a firm as a function of the market price, P, and the firm's TFP, A_i. Explain your derivation.

Question 1.6. A firm is active in the market (that is, it finds optimal to produce a positive amount of output) only if its profit is greater or equal than zero. Using this condition show that a firm is active in this industry only if its TFP satisfies the condition $A_i \geq 2/P$. Explain your derivation.

Let $(P^*, Q^*, Y_1^*, Y_2^*, ..., Y_N^*)$ the equilibrium price, total output, and individual firms' outputs. Based on the previous results, the market equilibrium can be characterized by the following conditions: (i) the demand equation holds; (ii) total output is equal to the sum of firms' individual outputs; (iii) firm i is active ($Y_i^* > 0$) if and only if its total profit is greater than zero; and (iv) for firms with $Y_i^* > 0$, the optimal amount of output is given by the condition price is equal to marginal cost.

Question 1.7. Write conditions (i) to (iv) for this particular industry.

Question 1.8. Combine conditions (i) to (iv) to show that the equilibrium price can be written as the solution to this equation:

$$P^* = 100 - P^* \left[\sum_{i=1}^{N} A_i^2 \; 1\{A_i \geq 2/P^*\} \right]$$

where $1\{x\}$ is the indicator function that is defined as $1\{x\} = 1$ if condition x is true, and $1\{x\} = 0$ if condition x is false. Explain your derivation.

Suppose that the subindex i sorts firms by their TFP such that firm 1 is the most efficient, then firm 2, etc. That is, $A_1 > A_2 > A_3 > .....$

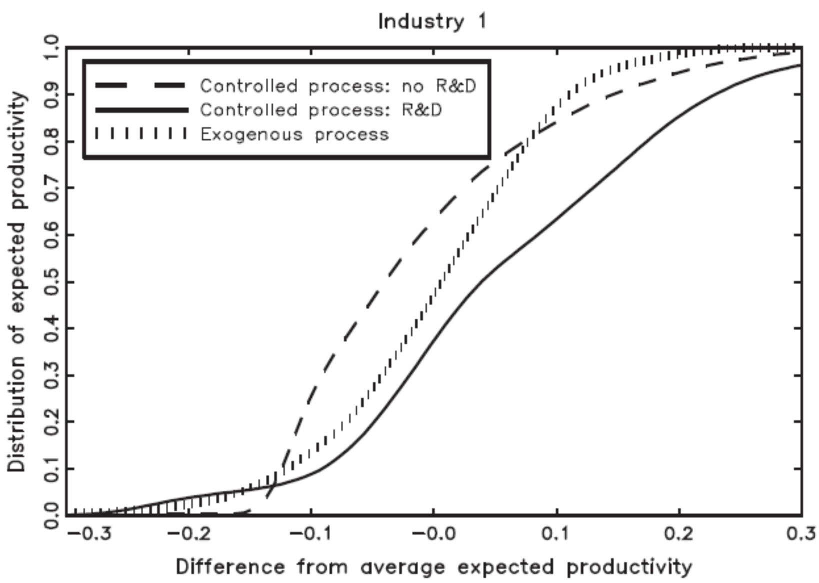

Figure 3.11: Doraszelski and Jaumandreu (2013): R&D and productivity

Question 1.9. Suppose that $A_1 = 7$, $A_2 = 5$, and $A_3 = 1$. Obtain the equilibrium price, total output, and output of each individual firm in this industry. [Hint: Start with the conjecture that only firms 1 and 2 produce in equilibrium. Then, confirm this conjecture. Note that we do not need to know the values of A_4, A_5, etc].

Question 1.10. Explain why the most efficient firm, with the largest TFP, does not produce all the output of the industry.

3.7.2 Exercise 2

The Stata datafile `blundell_bond_2000_production_function.dta` contains annual information on sales, labor, and capital for 509 firms for the period 1982-1989 (8 years). Consider a Cobb-Douglas production function in terms of labor and capital. Use this dataset to implement the following estimators.

Question 2.1. OLS with time dummies. Test the null hypothesis $\alpha_L + \alpha_K = 1$. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.2. Fixed Effects estimator with time dummies. Test the null hypothesis of no time-invariant unobserved heterogeneity: $\eta_i = \eta_i$ for every firm i. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.3. Fixed Effects - Cochrane Orcutt estimator with time dummies. Test the two over-identifying restrictions of the model. Provide the code in Stata and the table of estimation results. Comment the results.

Industry	Elasticity wrt. R_{jt-1}[a] Q_1	Q_2	Q_3	Mean
	(1)	(2)	(3)	(4)
1. Metals and metal products	−0.013	0.007	0.021	0.022
2. Non-metallic minerals	−0.018	−0.012	0.000	−0.006
3. Chemical products	0.009	0.011	0.014	0.013
4. Agric. and ind. machinery	−0.017	−0.009	0.021	0.005
6. Transport equipment	−0.034	−0.008	0.010	0.020
7. Food, drink, and tobacco	−0.008	0.010	0.026	0.020
8. Textile, leather, and shoes	−0.003	0.014	0.051	0.046
9. Timber and furniture	−0.031	0.005	0.048	0.004
10. Paper and printing products	−0.036	0.022	0.049	0.013

Figure 3.12: Doraszelski and Jaumandreu (2013): Elasticity TFP lagged R&D

Question 2.4. Arellano-Bond estimator with time dummies and non-serially correlated transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.5. Arellano-Bond estimator with time dummies and AR(1) transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.6. Blundell-Bond system estimator with time dummies and non-serially correlated transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.7. Blundell-Bond system estimator with time dummies and AR(1) transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.8. Based on the previous results, select your preferred estimates of the production function. Explain your choice.

3.7.3 Exercise 3

The Stata datafile `data_mines_eco2901_2017.dta` contains annual information on output and inputs from 330 copper mines for the period 1992-2010 (19 years). The following is a description of the variables.

Elasticity wrt. ω_{jt-1}[b]					
Performers			Non-performers		
Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
(5)	(6)	(7)	(8)	(9)	(10)
0.504	0.619	0.755	0.441	0.759	0.901
0.433	0.477	0.575	0.377	0.646	0.878
0.459	0.523	0.634	0.547	0.815	0.947
0.434	0.721	0.791	0.729	0.894	0.979
0.404	0.615	0.727	0.423	0.513	0.646
0.445	0.705	0.867	0.822	0.930	0.965
0.090	0.325	0.626	0.491	0.605	0.689
0.458	0.585	0.814	0.303	0.430	0.641
0.405	0.676	0.812	0.569	0.644	0.670

Figure 3.13: Doraszelski and Jaumandreu (2013): Elasticity TFP lagged TFP

Variable name		**Description**
`id`	:	Mine identification number
`year`	:	Year [from 1992 to 2010]
`active`	:	Binary indicator of the event "mine is active during the year"
`prod_tot`	:	Annual production of pure copper of the mine [in thousands of tonnes]
`reserves`	:	Estimated mine reserves [in thousands of ore]
`grade`	:	Average ore grade (in %) of mined ore during the year (% copper / ore)
`labor_n_tot`	:	Total number of workers per year (annual equivalent)
`cap_tot`	:	Measure of capital [maximum production capacity of the mine]
`fuel_cons_tot`	:	Consumption of fuel (in physical units)
`elec_cons_tot`	:	Consumption of electricity (in physical units)
`materials_tot`	:	Consumption of intermediate inputs / materials (in $ value)

Note that some variables have a few missing values even at years when the mine is actively producing.

Question 3.1. Consider a Cobb-Douglas production function in terms of labor, capital, fuel, electricity, and ore grade. Use this dataset to implement the following estimators:

- OLS
- Fixed-Effects
- Arellano-Bond estimator with non-serially correlated transitory shock
- Arellano-Bond estimator with AR(1) transitory shock
- Blundell-Bond estimator with non-serially correlated transitory shock

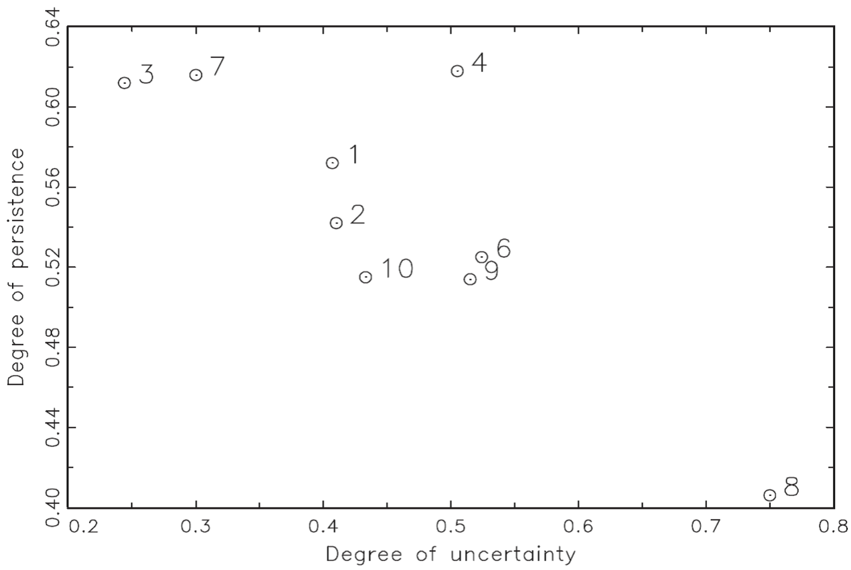

Figure 3.14: Doraszelski and Jaumandreu (2013): Uncertainty and persistence TFP

- Blundell-Bond estimator with AR(1) transitory shock
- Olley-Pakes (Using the first difference in cap_tot as investment)
- Levinshon-Petrin

Question 3.2. Suppose that these mines are price takers in the input markets. Consider that the variable inputs are labor, fuel, and electricity.

(a) Derive the expression for the Variable Cost function for a mine (that is, the minimum cost to produce an amount of output given input prices).
(b) Let $\ln MC_{it}$ be the logarithm of the realized Marginal Cost of mine i at year t. I have not included data on input prices in this dataset, so we will assume that mines face the same prices for variable inputs, and normalize to zero the contribution of these input prices to $\ln MC_{it}$. Calculate the quantiles 5%, 25%, 50%, 75%, and 95% in the cross-sectional distributions of $\ln MC_{it}$ at each year in the sample. Present a figure with the time-series of these five quantiles over the sample period. Comment the results.
(c) For a particular sample year, say 2005, calculate the contribution of each component of $\ln MC_{it}$ (that is, total factor productivity, capital, ore grade, and output) to the cross-sectional variance of $\ln MC_{it}$. Present it in a table. Comment your results.
[Note: To measure the contribution of each component, use the following approach. Consider $y = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K$. A measure of the contribution of x_j to $var(y)$ is $\rho_j \equiv \dfrac{var(y) - var(y \mid x_j = constant)}{var(y)}$. Note that $\rho_j \in (0,1)$ for any variable x_j. However, in general, $\sum_{j=1}^{K} \rho_j$ can be either

smaller or greater than one, depending the sign of the covariances between the components.]

(d) Consider the balance panel of mines that are active in the industry every year during the sample period. Repeat exercises (b) and (c) for this balanced panel. Compared your results with those in (c) and (d). Comment the results.