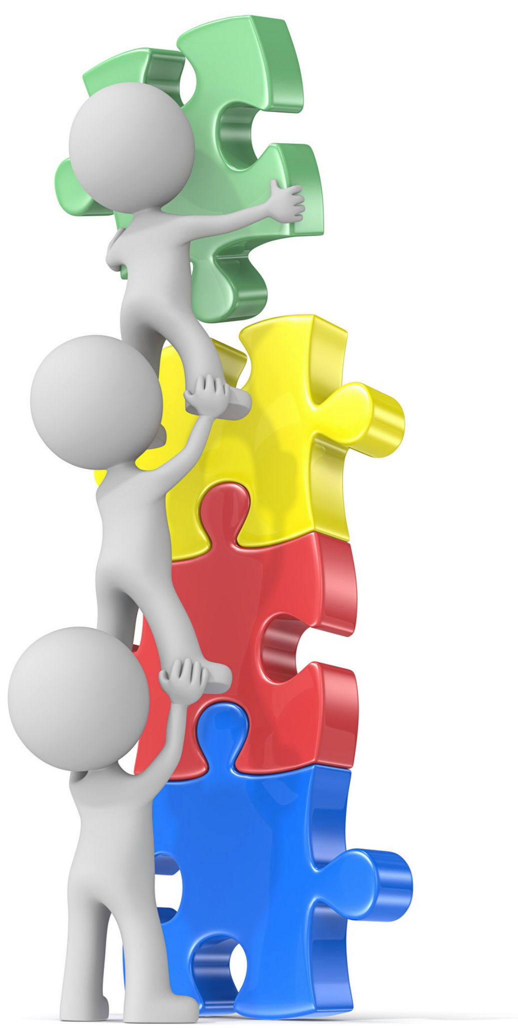
Empirical Industrial Organization: Models, Methods, and Applications

Victor Aguirregabiria

University of Toronto

This version: March 8, 2019



Contents

CHAPTER 3

Estimation of Production Functions

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 costs 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 productions functions.

(a) Measurement issues. There are important measurement problems such as measurement error in output (for instance, observeing revenue instead of output in physical units), and inputs (for instance, construction of the economic value of capital using information of book value; differences in the quality of labor).

(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. It 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 than 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.

2. Model and Data

2.1. Model. 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:

$$Y = F(X_1, X_2, ..., X_J, A) \tag{2.1}$$

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 technological efficiency. The marginal productivity of input j is $MP_j = \partial F / \partial X_j$.

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

$$Y = L^{\alpha_L} \ K^{\alpha_K} \ A \tag{2.2}$$

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 explicitly 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 (blue collar and white collar labor), and capital (equipment, information technology). For the Cobb-Douglas PF, the productivity term A is denoted the *Total Factor Productivity* (TFP). For this PF 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:

$$\frac{\partial MP_j}{\partial X_k} = \frac{\alpha_j}{X_j} \frac{\alpha_k}{X_k} Y > 0 \tag{2.3}$$

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

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:

$$\begin{aligned} C(Y) &= \min_{\{X_1, X_2, ..., X_J\}} W_1 X_1 + W_2 X_2 + ... + W_J X_J \\ & \text{subject to: } Y \geq F(X_1, X_2, ..., X_J, A) \end{aligned} \tag{2.4}$$

The marginal conditions of optimality imply that for every input j, $W_j - \lambda\ F_j(\mathbf{X}, A) = 0$, where λ is the Lagrange multiplier of the restriction.

Exercise 3.1. Given the Cobb-Douglas PF $Y = X_1^{\alpha_1} ...\ X_1^{\alpha_J}\ A$ and input prices W_j, obtain the corresponding cost function. The marginal condition of optimality for input j implies $W_j - \lambda\alpha_j\ (Y/X_j) = 0$, or equivalently:

$$W_j\ X_j = \lambda\alpha_j\ Y \tag{2.5}$$

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 solve the marginal conditions $X_j = \lambda\alpha_j\ Y/W_j$ into the production function:

$$Y = A\left(\frac{\lambda\alpha_1 Y}{W_1}\right)^{\alpha_1}\left(\frac{\lambda\alpha_2 Y}{W_2}\right)^{\alpha_2}...\left(\frac{\lambda\alpha_J Y}{W_J}\right)^{\alpha_J} \tag{2.6}$$

Solving in this expression for the Lagrange multiplier we get

$$\lambda = \left(\frac{W_1}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha}}\left(\frac{W_2}{\alpha_2}\right)^{\frac{\alpha_2}{\alpha}}...\left(\frac{W_J}{\alpha_J}\right)^{\frac{\alpha_J}{\alpha}}\ Y^{\frac{1-\alpha}{\alpha}}\ A^{\frac{-1}{\alpha}} \tag{2.7}$$

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

$$C(Y) = \alpha\ \left(\frac{Y}{A}\right)^{\frac{1}{\alpha}}\ \left(\frac{W_1}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha}}\left(\frac{W_2}{\alpha_2}\right)^{\frac{\alpha_2}{\alpha}}...\left(\frac{W_J}{\alpha_J}\right)^{\frac{\alpha_J}{\alpha}} \qquad \blacksquare \tag{2.8}$$

Looking at the Cobb-Douglas cost function in equation (2.8) 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.

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

$$y = \alpha_L \; l + \alpha_K \; k + \omega \tag{2.9}$$

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 the residual term U. The simplicity of the Cobb-Douglas PF comes also with a price. One of its drawbacks is that it implies that the elasticity of substitution between labor and capital (or between any two inputs) is always 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:

$$Y = L^{[\alpha_{L0}+\alpha_{LL} l+\alpha_{LK} k]} \; K^{[\alpha_{K0}+\alpha_{KL} l+\alpha_{KK} k]} \; U \tag{2.10}$$

that in logarithms becomes,

$$y = \alpha_{L0} \; l + \alpha_{K0} \; k + \alpha_{LL} \; l^2 + \alpha_{KK} \; k^2 + (\alpha_{LK} + \alpha_{KL}) \; l \; k + \omega \tag{2.11}$$

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: a measure of output, for instance, number of units, or revenue, or valued added; input measures such as labor, capital, R&D, materials, and energy; and some measures of output and 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 has been Compustat, and the Longitudinal Research Database from US Census Bureau. In Europe, some country 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 researcher observes a panel dataset of N firms, indexed by i, over several periods of time, indexed by t, with the following information:

$$\textbf{Data} \;=\; \{y_{it},\, l_{it},\, k_{it},\, w_{it},\, r_{it} : \; i = 1, 2, ...N; \; t = 1, 2, ..., T_i\} \tag{2.12}$$

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. Econometric Issues

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

$$y_{it} = \alpha_L \; l_{it} + \alpha_K \; k_{it} + \omega_{it} + e_{it} \tag{3.1}$$

ω_{it} represents unobserved (for the econometrician) inputs such as managerial ability, quality of land, materials, etc, which are known to the firm when it decides capital and labor. 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 decides 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, $y_{it}^e \equiv y_{it} - e_{it}$.

3.1. 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 already pointed out in the seminal paper by Marshak and Andrews (1944).

Example 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, $\alpha_L \; Y^e/L = W_L$, 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]

$$\begin{aligned} y_{it} &= \alpha_L \; l_{it} + \omega_{it} + e_{it} \\ y_{it} - l_{it} &= e_{it} \end{aligned} \tag{3.2}$$

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

The reduced form equations of this structural model are:

$$\begin{aligned} y_{it} &= \frac{\omega_{it}}{1-\alpha_L} + e_{it} \\ l_{it} &= \frac{\omega_{it}}{1-\alpha_L} \end{aligned} \tag{3.3}$$

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 $Cov(y_{it}, l_{it})/Var(l_{it})$. 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: $Cov(y_{it}, l_i) = Var(l_{it})$. Therefore, under the conditions of this model the OLS estimator of α_L converges in probability to 1 regardless 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 converges in probability to 1. It is clear that in this case ignoring the endogeneity of inputs can generate a serious in the estimation of the PF parameters.

[Graphical representation of structural equations in space (ℓ, y) and graphical interpretation of the bias.]

Example 2: Consider the similar conditions as in Example 1, but now firms in our sample 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 the type labor input that it demands to produce its product and that the market price 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 (3.2) with the only difference that the labor demand equation now includes the logarithm of the price of labor: $y_{it} - l_{it} = r_{it} + e_{it}$. The reduced form equations for this model are:

$$\begin{aligned} y_{it} &= \frac{\omega_{it} - r_{it}}{1-\alpha_L} + r_{it} + e_{it} \\ l_{it} &= \frac{\omega_{it} - r_{it}}{1-\alpha_L} \end{aligned} \tag{3.4}$$

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 $Cov(y_{it}, l_{it})/Var(l_{it})$ and in this case this implies the following expression for the bias:

$$Bias\left(\hat{\alpha}_L^{OLS}\right) = \frac{1-\alpha_L}{1 + \dfrac{\sigma_r^2}{\sigma_\omega^2}} \tag{3.5}$$

where σ_ω^2 and σ_r^2 represent the variance of the productivity shock and the logarithm of the price of labor, respectively. The bias is always upward because the firm's labor demand is always positively correlated with the firm's productivity shock. The ratio between the variance of the price of labor and the variance of productivity, $\sigma_r^2/\sigma_\omega^2$, 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.

3.2. Endogenous Exit. Firm or plant panel datasets are unbalanced, with significant amount of firm exits. Exiting firms are not randomly chosen from the population of operating firms. For instance, existing firms are typically smaller than surviving firms.

Let d_{it} be the indicator of the event "firm i stays in the market at the end of period t". Let $V^1(l_{it-1}, k_{it}, \omega_{it})$ be the value of staying in the market, and let $V^0(l_{it-1}, k_{it}, \omega_{it})$ be the value of exiting (that is, the scrapping value of the firm). Then, the optimal exit/stay decision is:

$$d_{it} = I\left\{\ V^1(l_{it-1}, k_{it}, \omega_{it}) - V^0(l_{it-1}, k_{it}, \omega_{it}) \geq 0\ \right\} \tag{3.6}$$

Under standard conditions, the function $V^1(l_{it-1}, k_{it}, \omega_{it}) - V^0(l_{it-1}, k_{it}, \omega_{it})$ is 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:

$$d_{it} = I\{\ \omega_{it} \geq \omega^*(l_{it-1}, k_{it})\ \} \tag{3.7}$$

where the threshold function $\omega^*(.,.)$ is strictly decreasing in all its arguments.

Consider the PF $y_{it} = \alpha_L\ l_{it} + \alpha_K\ k_{it} + \omega_{it} + e_{it}$. 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:

$$\omega_{it}^{d=1} \quad \equiv \quad \{\omega_{it} \mid d_{it} = 1\} \quad = \quad \{\omega_{it} \mid \omega_{it} \geq \omega^*(l_{i,t-1}, k_{it})\} \tag{3.8}$$

Even if the productivity shock ω_{it} is independent of the state variables $(l_{i,t-1}, k_{it})$, the self-selected productivity shock $\omega_{it}^{d=1}$ will not be mean-independent of $(l_{i,t-1}, k_{it})$. That is,

$$\begin{aligned} E\left(\omega_{it}^{d=1} \mid l_{i,t-1}, k_{it}\right) &= E\left(\omega_{it} \mid l_{i,t-1}, k_{it}, d_{it} = 1\right) \\ &= E\left(\omega_{it} \mid l_{i,t-1}, k_{it}, \omega_{it} \geq \omega^*(l_{i,t-1}, k_{it})\right) \\ &= \lambda\left(l_{i,t-1}, k_{it}\right) \end{aligned} \tag{3.9}$$

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

$$y_{it} = \quad \alpha_L \; l_{it} + \alpha_K \; k_{it} + \lambda\left(l_{i,t-1}, k_{it}\right) + \tilde{\omega}_{it} + e_{it} \tag{3.10}$$

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

Ignoring the selection term $\lambda\left(l_{i,t-1}, k_{it}\right)$ introduces bias in our estimates of the PF parameters. The selection term is an increasing function of the threshold $\omega^*\left(l_{i,t-1}, k_{it}\right)$, and therefore it is decreasing in $l_{i,t-1}$ and k_{it}. Both l_{it} and k_{it} are negatively correlated with the selection term, but the correlation with the capital stock tend to be larger because the value of a firm depends 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.

This type of selection problem has been pointed out also by different authors who have studied empirically the relationship between firm growth and firm size. The relationship between firm size and firm growth has important policy implications. Mansfield (1962), Evans (1987), and Hall (1987) are seminal papers in that literature. Consider the regression equation:

$$\Delta s_{it} = \alpha + \beta \; s_{i,t-1} + \varepsilon_{it} \tag{3.11}$$

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

$$d_{it} = I\left\{\; \varepsilon_{it} \geq \varepsilon^*\left(s_{i,t-1}\right) \;\right\} \tag{3.12}$$

where $\varepsilon^*(.)$ 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 $\varepsilon_{it}^{d=1} \equiv \{\varepsilon_{it} | d_{it} = 1\} = \{\varepsilon_{it} | \varepsilon_{it} \geq \varepsilon^*\left(s_{i,t-1}\right)\}$. Thus,

$$\Delta s_{it} = \quad \alpha + \beta s_{i,t-1} + \lambda\left(s_{i,t-1}\right) + \tilde{\varepsilon}_{it} \tag{3.13}$$

where $\lambda\left(s_{i,t-1}\right) \equiv E(\varepsilon_{it}|\varepsilon_{it} \geq \varepsilon^{*}\left(s_{i,t-1}\right))$, and $\tilde{\varepsilon}_{it} \equiv \{\varepsilon_{it}^{d=1} - \lambda\left(l_{i,t-1}, k_{it}\right)\}$ that, 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 $\varepsilon^{*}\left(s_{i,t-1}\right)$, 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 used 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. Of course, the main limitation of Mansfield's approach is that including exiting firms using the imputed values $\Delta s_{it} = -1$ does not correct completely for selection bias. But Mansfield's paper was written almost twenty years before Heckman's seminal contributions on sample selection in econometrics. 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. However, they also find that after controlling for endogenous selection a la Heckman, the estimate of β is significantly lower than zero. 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, that is, normality of the error term, and linear relationship between firm size and firm growth.

4. Estimation Methods

4.1. Using Input Prices as Instruments. If input prices, r_i, are observable, and they are not correlated with the productivity shock 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. Time-series variation is not enough for identification. Third, if firms in our sample operate in different input markets, we may observe significant cross-sectional variation in input prices. However, this variation is suspicious of being endogenous. The different markets where firms operate can be also different in the average unobserved productivity of firms, and therefore $cov\left(\omega_i, r_i\right) \neq 0$, that is, input prices not a valid instruments. In general, when there is cross-sectional variability in input prices, can one say that input prices are valid instruments for inputs in a PF? Is $cov(\omega_i, r_i) = 0$? When inputs are firm-specific, it is commonly the case that input prices depend on the firm's productivity.

4.2. Panel Data: Fixed-Effects Estimators. 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:

$$y_{it} = \alpha_L \; l_{it} + \alpha_K \; k_{it} + \omega_{it} + e_{it} \tag{4.1}$$

Mundlak (1961) and Mundlak and Hoch (1965) are seminal studies in the use of 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: $\omega_{it} = \eta_i + \delta_t + \omega^*_{it}$. 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 ω^*_{it} is an firm idiosyncratic shock.

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

Assumption PD-3: ω^*_{it} is not serially correlated.

Assumption PD-4: The idiosyncratic shock ω^*_{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 other random and unpredictable shock.

The Within-Groups estimator (WGE) or fixed-effects estimator of the PF is just the OLS estimator in the Within-Groups transformed equation:

$$(y_{it} - \bar{y}_i) = \alpha_L \left(l_{it} - \bar{l}_i\right) + \alpha_K \left(k_{it} - \bar{k}_i\right) + (\omega_{it} - \bar{\omega}_i) + (e_{it} - \bar{e}_i) \tag{4.2}$$

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 ω^*_{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 $x_{it} \equiv (l_{it}, k_{it})$ to be strictly exogenous. That is, for any (t, s):

$$cov\left(x_{it}, \omega^*_{is}\right) = cov\left(x_{it}, e_{is}\right) = 0 \tag{4.3}$$

Otherwise, the WG-transformed regressors $\left(l_{it} - \bar{l}_i\right)$ 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 Grilliches and Mairesse, 1998). There are at least two factors that can explain this empirical regularity. First, though Assumptions (PD-2) and (PD-3) may be plausible for the estimation of agricultural PFs, they are very unrealistic for manufacturing firms. And second, the bias induced by measurement-error in the regressors can be exacerbated by the WG transformation. That is, the noise-to-signal ratio can be much larger for the WG transformed inputs than for the variables in levels. To see this, consider the model with only one input, say capital, and suppose that it is measured with 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. In the estimation of the PF in levels we have that:

$$Bias(\hat{\alpha}_L^{OLS}) = \frac{Cov(k, \eta)}{Var(k) + Var(e^k)} - \frac{\alpha_L \; Var(e^k)}{Var(k) + Var(e^k)} \tag{4.4}$$

If $Var(e^k)$ is small relative to $Var(k)$, 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:

$$Bias(\hat{\alpha}_L^{WGE}) = -\frac{\alpha_L \; Var(\Delta e^k)}{Var(\Delta k) + Var(\Delta e^k)} \tag{4.5}$$

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 ratio $Var(\Delta e^k)/Var(\Delta k)$ can be large even when the ratio $Var(e^k)/Var(k)$ is quite small. Therefore, the WGE may be significantly downward biased.

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

$$\Delta y_{it} = \alpha_L \; \Delta l_{it} + \alpha_K \; \Delta k_{it} + \Delta\delta_t + \Delta\omega_{it}^* + \Delta e_{it} \tag{4.6}$$

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: There are adjustment costs in inputs (at least in one input). More formally, the reduced form equations for labor and capital are $l_{it} = f_L(l_{i,t-1}, k_{i,t-1}, \omega_{it})$ and

$k_{it} = f_K(l_{i,t-1}, k_{i,t-1}, \omega_{it})$, respectively, where either $l_{i,t-1}$ or $k_{i,t-1}$, or both, have non-zero partial derivatives in f_L and f_K.

Under these assumptions $\{l_{i,t-j}, k_{i,t-j}, y_{i,t-j} : j \geq 2\}$ are valid instruments in the PD in first differences. Identification comes from the combination of two assumptions: (1) serial correlation of inputs; and (2) no serial correlation in productivity shocks $\{\omega_{it}^*\}$. The presence of adjustment costs implies that the shadow prices of inputs vary 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 $\{\omega_{it}^*\}$ is key, but it can be tested using an LM test (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 ω_{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 of 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:

$$y_{it} = \alpha\ x_{it} + \eta_i + \omega_{it}^* + e_{it} \tag{4.7}$$

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

$$x_{it} = \delta\ x_{i,t-1} + \lambda_1\ \eta_i + \lambda_2\ \omega_{it}^* \tag{4.8}$$

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:

$$\Delta y_{it} = \alpha\ \Delta x_{it} + \Delta\omega_{it}^* + \Delta e_{it} \tag{4.9}$$

For simplicity, consider that the number of periods in the panel is $T = 3$. In this context, Arellano-Bond GMM estimator is equivalent to Anderson-Hsiao IV estimator (Anderson and Hsiao, 1981, 1982) where the endogenous regressor Δx_{it} is instrumented using $x_{i,t-2}$. This IV estimator is:

$$\hat{\alpha}_N = \frac{\sum_{i=1}^{N} x_{i,t-2}\ \Delta y_{it}}{\sum_{i=1}^{N} x_{i,t-2}\ \Delta x_{it}} \tag{4.10}$$

Under the assumptions of the model, we have that $x_{i,t-2}$ is orthogonal to the error $(\Delta\omega^*_{it} + \Delta e_{it})$. Therefore, $\hat{\alpha}_N$ identifies α if the (asymptotic) 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:

$$p\lim R^2 = \frac{Cov\left(\Delta x_{it}, x_{i,t-2}\right)^2}{Var\left(\Delta x_{it}\right)\ Var\left(x_{i,t-2}\right)} = \frac{\left(\gamma_2 - \gamma_1\right)^2}{2\left(\gamma_0 - \gamma_1\right)\gamma_0} \tag{4.11}$$

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 $x_{it} = \frac{\lambda_1\ \eta_i}{1-\delta} + \lambda_2(\omega_{it} + \delta\ \omega_{i,t-1} + \delta^2\ \omega_{i,t-2} + ...)$, we can derive the following expressions for the autocovariances:

$$\begin{aligned}
\gamma_0 &= \frac{\lambda_1^2\ \sigma_\eta^2}{\left(1-\delta\right)^2} + \frac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2} \\
\gamma_1 &= \frac{\lambda_1^2\ \sigma_\eta^2}{\left(1-\delta\right)^2} + \delta\ \frac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2} \\
\gamma_2 &= \frac{\lambda_1^2\ \sigma_\eta^2}{\left(1-\delta\right)^2} + \delta^2\ \frac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2}
\end{aligned} \tag{4.12}$$

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

$$\begin{aligned}
R^2 &= \frac{\left(\delta\ \dfrac{\lambda_2^2\sigma_\omega^2}{1+\delta}\right)^2}{2\left(\dfrac{\lambda_2^2\sigma_\omega^2}{1+\delta}\right)\left(\dfrac{\lambda_1^2\ \sigma_\eta^2}{\left(1-\delta\right)^2} + \dfrac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2}\right)} \\
&= \frac{\delta^2\left(1-\delta\right)^2}{2\left(1-\delta+\left(1+\delta\right)\rho\right)}
\end{aligned} \tag{4.13}$$

with $\rho \equiv \lambda_1^2\sigma_\eta^2/\lambda_2^2\sigma_\omega^2 \geq 0$. 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 δ 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). Then, $R^2 = \delta^2\left(1-\delta\right)/2$. The maximum value of this R-square is $R^2 = 0.074$ that occurs when $\delta = 2/3$. This is the upper bound for the R-square, but it is a too optimistic upper bound because it is based on the assumption of no fixed effects. For instance, a more realistic case for ρ is $\lambda_1^2\sigma_\eta^2 = \lambda_2^2\sigma_\omega^2$ and therefore $\rho = 1$. Then, $R^2 = \delta^2\left(1-\delta\right)^2/4$. The maximum value of this R-square is $R^2 = 0.016$ that 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 ω_{it}^* and e_{it} were zero, and input and output were equal to their firm-specific, steady-state mean values:

$$\begin{aligned} x_{it_i^*} &= \frac{\lambda_1 \eta_i}{1-\delta} \\ y_{it_i^*} &= \alpha \ \frac{\lambda_1 \eta_i}{1-\delta} + \eta_i \end{aligned} \tag{4.14}$$

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

$$\begin{aligned} x_{it} &= x_{it_i^*} + \lambda_2 \left(\omega_{it}^* + \delta \omega_{it-1}^* + \delta^2 \omega_{it-2}^* + ... \right) \\ y_{it} &= y_{it_i^*} + \omega_{it}^* + \alpha \lambda_2 \left(\omega_{it}^* + \delta \omega_{it-1}^* + \delta^2 \omega_{it-2}^* + ... \right) \end{aligned} \tag{4.15}$$

These expressions imply that input and output in first differences depend on the history of the i.i.d. shock $\{\omega_{it}^*\}$ 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 $j > 0$:

$$\begin{aligned} E\left(\Delta x_{it-j}\left[\eta_i + \omega_{it}^* + e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(\Delta x_{it-j}\left[y_{it} - \alpha x_{it}\right]\right) = 0 \\ E\left(\Delta y_{it-j}\left[\eta_i + \omega_{it}^* + e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(\Delta y_{it-j}\left[y_{it} - \alpha x_{it}\right]\right) = 0 \end{aligned} \tag{4.16}$$

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 $j > 1$:

$$\begin{aligned} E\left(x_{it-j}\left[\Delta\omega_{it}^* + \Delta e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(x_{it-j}\left[\Delta y_{it} - \alpha \Delta x_{it}\right]\right) = 0 \\ E\left(y_{it-j}\left[\Delta\omega_{it}^* + \Delta e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(y_{it-j}\left[\Delta y_{it} - \alpha \Delta x_{it}\right]\right) = 0 \end{aligned} \tag{4.17}$$

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-Mangas (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 Soderbom (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 firms' dynamic programming problem, simulate data of inputs and output using the optimal decision rules, and use simulated data and 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 two 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 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 which are independent of productivity seems 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 & Bond (2001); 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

4.4. Control Function Approaches. In a seminal paper, Olley and Pakes (1996) propose a control function approach to estimate PFs. Levinsohn and Petrin (2003) have extended Olley-Pakes approach to contexts where data on capital investment presents significant censoring at zero investment.

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

$$
\begin{aligned}
&(PF) \quad y_{it} = \alpha_L \; l_{it} + \alpha_K \; k_{it} + \omega_{it} + e_{it} \\
&(LD) \quad l_{it} = f_L \left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right) \\
&(ID) \quad i_{it} = f_K \left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right)
\end{aligned}
\tag{4.18}
$$

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

Assumption OP-2: There is not 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.

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 Olley and Pakes model, lagged labor, $l_{i,t-1}$, is not a state variable, that is, there a not labor adjustment costs, and labor is a perfectly flexible input. However, that assumption is not necessary for Olley-Pakes estimator. Here we discuss the method in the context of a model with labor adjustment costs.

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 OP method and for the related Levinshon-Petrin method. This imposes restrictions on the underlying model of market competition and inputs demands. For instance, this assumption implicitly establishes that firms operate in the same input markets and they do not have any monopsony power in these markets, for instance, no internal labor markets. 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 economics restrictions can be relaxed if the researcher has data on inputs prices at the firm level, that is, r_{it} is observable.

Olley-Pakes method deals both with the simultaneity problem and with the selection problem due to endogenous exit. For the sake of clarity, we start describing here a version of the method that does not deal with the selection problem. We will discuss later their approach to deal with endogenous exit.

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).* This second step is different when we deal with endogenous exit.

Step 1: Estimation of α_L. Assumptions (*OP-1*) and *(OP-2)* imply that $\omega_{it} = f_K^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t)$. Solving this equation into the PF we have:

$$\begin{aligned} y_{it} &= \alpha_L \ l_{it} + \alpha_K \ k_{it} + f_L^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t) + e_{it} \\ &= \alpha_L \ l_{it} + \phi_t(l_{i,t-1}, k_{it}, i_{it}) + e_{it} \end{aligned} \tag{4.19}$$

where $\phi_t(l_{i,t-1}, k_{it}, i_{it}) \equiv \alpha_K \ k_{it} + f_L^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t)$. Without a parametric assumption on the investment equation f_K, equation (4.19) is a semiparametric partially linear model. The parameter α_L and the functions ϕ_1, ϕ_2, ..., $\phi_T$can be estimated using semiparametric methods. A possible semiparametric method is the kernel method in Robinson (1988). Instead, Olley and Pakes use polynomial series approximations for the nonparametric functions ϕ_t.

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 $(l_{i,t-1}, k_{it}, i_{it})$, we control for the unobservable ω_{it}. Therefore, α_L is identified if given $(l_{i,t-1}, k_{it}, i_{it})$ there is enough cross-sectional variation left in l_{it}. The key conditions for the identification of α_L are: (a) invertibility of $f_L(l_{i,t-1}, k_{it}, \omega_{it}, r_t)$ 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 $(l_{i,t-1}, k_{it}, i_{it}, r_t)$, current labor l_{it} still has enough sample variability. Assumption (c) is key, and it is the base for Ackerberg, Caves, and Frazer (2006) criticism (and extension) of Olley-Pakes approach.

Example 3: Consider 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:

$$\omega_{it} = \ \gamma_1 \ i_{it} + \gamma_2 \ l_{i,t-1} + \gamma_3 \ k_{it} + r_{it} \tag{4.20}$$

Solving this equation into the PF, we have that:

$$y_{it} = \alpha_L \ l_{it} + (\alpha_K + \gamma_3) \ \ k_{it} + \gamma_1 \ i_{it} + \gamma_2 \ l_{i,t-1} + (r_{it} + e_{it}) \tag{4.21}$$

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 $(k_{it}, i_{it}, l_{i,t-1})$.

Step 2: Estimation of α_K. 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:

$$\omega_{it} = E[\omega_{it} \mid \omega_{i,t-1}] + \xi_{it} = h\left(\omega_{i,t-1}\right) + \xi_{it} \tag{4.22}$$

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

$$\begin{aligned} \phi_{it} &= \alpha_K\ k_{it} + h\left(\omega_{i,t-1}\right) + \xi_{it} \\ &= \alpha_K\ k_{it} + h\left(\phi_{i,t-1} - \alpha_K\ k_{i,t-1}\right) + \xi_{it} \end{aligned} \tag{4.23}$$

Though we do not know the true value of ϕ_{it}, we have consistent estimates of these values from step 1: that is, $\hat{\phi}_{it} = y_{it} - \hat{\alpha}_L\ l_{it}$.[2]

If function $h$is nonparametrically specified, equation (4.23) is a partially linear model. However, it is not a "standard" partially linear model because the argument of the h function, $\phi_{i,t-1} - \alpha_K k_{i,t-1}$, is not observable, that is, it depends on the unknown parameter α_K. To estimate $h$and α_K, Olley and Pakes propose a recursive version of the semiparametric method in the first step. Suppose that we consider a quadratic function for h: that is, $h(\omega) = \pi_1\omega + \pi_2\omega^2$. Then, given an initial value of α_K, we construct the variable $\hat{\omega}_{it}^{\alpha_K} = \hat{\phi}_{it} - \alpha_K k_{it}$, and estimate by OLS the equation $\hat{\phi}_{it} = \alpha_K k_{it} + \pi_1 \hat{\omega}_{it-1}^{\alpha_K} + \pi_2 (\hat{\omega}_{it-1}^{\alpha_K})^2 + \xi_{it}$. Given the OLS estimate of α_K, we construct new values $\hat{\omega}_{it}^{\alpha_K} = \hat{\phi}_{it} - \alpha_K k_{it}$ and estimate again α_K, π_1, and π_2 by OLS. We proceed until convergence. An alternative to this recursive procedure is the following Minimum Distance method. For instance, if the specification of $h(\omega)$ is quadratic, we have the regression model:

$$\begin{aligned} \hat{\phi}_{it} &= \alpha_K k_{it} + \pi_1 \hat{\phi}_{i,t-1} + \pi_2 \hat{\phi}_{i,t-1}^2 + \left(-\pi_1 \alpha_K\right) k_{i,t-1} + (\pi_2 \alpha_K^2) k_{i,t-1}^2 \\ &+ \left(-2\pi_2 \alpha_K\right) \hat{\phi}_{i,t-1} k_{i,t-1} + \xi_{it} \end{aligned} \tag{4.24}$$

We can estimate the parameters α_K, π_1, π_2, $(-\pi_1\alpha_K)$, $(\pi_2\alpha_K^2)$, and $(-2\pi_2\alpha_K)$ by OLS. This estimate of α_K can be very imprecise because the collinearity between the regressors. However, given the estimated vector of $\{\alpha_K,\ \pi_1,\ \pi_2,\ (-\pi_1\alpha_K),\ (\pi_2\alpha_K^2),\ (-2\pi_2\alpha_K)\}$ and its

[2]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.

variance-covariance matrix, we can obtain a more precise estimate of $(\alpha_K,\ \pi_1,\ \pi_2)$ by using minimum distance.

Example 4: Suppose that we consider a parametric specification for the stochastic process of $\{\omega_{it}\}$. More specifically, consider the AR(1) process $\omega_{it} = \rho\ \omega_{i,t-1} + \xi_{it}$, where $\rho \in [0,1)$ is a parameter. Then, $h\left(\omega_{i,t-1}\right) = \rho\omega_{i,t-1} = \rho(\phi_{i,t-1} - \alpha_K\ k_{i,t-1})$, and we can write:

$$\phi_{it} = \alpha_K\ k_{it} + \rho\ \phi_{i,t-1} + (-\rho\alpha_K)\ \ k_{i,t-1} + \xi_{it} \tag{4.25}$$

we can see that a regression of ϕ_{it} on k_{it}, $\phi_{i,t-1}$ and $k_{i,t-1}$ identifies (in fact, over-identifies) α_K and ρ.

Time-to build is a key assumption for the consistency of this method. If new investment at period t is productive at the same period, then we have that: $\phi_{it} = \alpha_K\ k_{i,t+1} + h\left(\phi_{i,t-1} - \alpha_K\ k_{it}\right) + \xi_{it}$. 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}.

Levinshon and Petrin (2003) propose a related control function method. The main difference with OP method is that Levinshon and Petrin (LP) use the demand function for intermediate inputs instead of the investment equation to invert out unobserved productivity. The consider a Cobb-Douglas production function in terms of labor, capital, and intermediate inputs (materials):

$$y_{it} = \alpha_L\ l_{it} + \alpha_K\ k_{it} + \alpha_M\ m_{it} + \omega_{it} + e_{it} \tag{4.26}$$

The investment equation is replaced with the intermediate input demand:

$$m_{it} = f_M\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}\right) \tag{4.27}$$

Levinshon and Petrin maintain Assumptions OP-2 to OP-4, and replace Assumption OP-1 of monotonicity (invertibility) of the investment equation with a similar assumption for the intermediate input demand.

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

Similarly as for the Olley-Pakes method, the key identification restriction in Levinshon-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}, that is, *Assumption OP-2:* There is not cross-sectional variation in input prices such that $r_{it} = r_t$ for every firm i.

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

$$y_{it} \quad = \quad \alpha_L\ l_{it} + \phi_t(l_{i,t-1}, k_{it}, m_{it}) + e_{it} \tag{4.28}$$

where $\phi_t(l_{i,t-1}, k_{it}, m_{it}) = \alpha_K\ k_{it} + f_M^{-1}(l_{i,t-1}, k_{it}, m_{it}, r_t)$ and f_M^{-1} represents the inverse function of the intermediate input demand with respect to productivity. The second step is also similar to OP's second step but in the model with the intermediate input. More specifically, the estimates of α_L and ϕ_t are plugged-in, and a least squares is applied to the estimation of the parameters α_K and α_M and function $h$in the regression equation:

$$\phi_{it} \quad = \quad \alpha_K\ k_{it} + \alpha_M\ m_{it} + h\left(\phi_{i,t-1} - \alpha_K\ k_{i,t-1} - \alpha_M\ m_{i,t-1}\right) + \xi_{it} \tag{4.29}$$

There are several advantages of using the intermediate input

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

$$l_{it} = f_L\left(l_{i,t-1},\ k_{it},\ f_K^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t),\ r_t\right) = G_t\left(l_{i,t-1}, k_{it}, i_{it}\right) \tag{4.30}$$

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 $(l_{i,t-1}, k_{it}, i_{it})$. In other words, once we condition on the observable variables $(l_{i,t-1}, k_{it}, i_{it})$, employment at period t should not have any cross-sectional variability. It should be constant. This implies that in the regression in step 1, $y_{it} = \alpha_L\ l_{it} + \phi_t(l_{i,t-1}, k_{it}, i_{it}) + e_{it}$, 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 $(l_{i,t-1}, k_{it}, i_{it})$.

Example 5: The problem can be illustrated more clearly by using linear functions for the optimal investment and labor demand. Suppose that the inverse function f_K^{-1} is $\omega_{it} = \gamma_1\ i_{it} + \gamma_2\ l_{i,t-1} + \gamma_3\ k_{it} + \gamma_4 r_t$; and the labor demand equation is $l_{it} = \delta_1 l_{i,t-1} + \delta_2 k_{it} + \delta_3 \omega_{it} + \delta_4 r_t$. Then, solving the inverse function f_K^{-1} into the production function, we get:

$$y_{it} = \alpha_L\ l_{it} + (\alpha_K + \gamma_3)\ \ k_{it} + \gamma_1\ i_{it} + \gamma_2\ l_{i,t-1} + (\gamma_4 r_t + e_{it}) \tag{4.31}$$

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

$$l_{it} = (\delta_1 + \delta_3\gamma_2) l_{i,t-1} + (\delta_2 + \delta_3\gamma_3) k_{it} + \delta_3\gamma_1 i_{it} + (\delta_4 + \delta_3\gamma_4) r_t \tag{4.32}$$

Equation (4.32) shows that there is perfect collinearity between l_{it} and $(l_{i,t-1}, k_{it}, i_{it})$ and therefore it should not be possible to estimate α_L in equation (4.31). Of course, in the data we will find that l_{it} has some cross-sectional variation independent of $(l_{i,t-1}, k_{it}, i_{it})$. Equation (4.32) shows that if that variation is present it is because input prices r_{it} have cross-sectional variation. However, that variation is endogenous in the estimation of equation (4.31) because the unobservable r_{it} is part of the error term. That is, if there is apparent identification, that identification is spurious.

After pointing out this important problem in Olley-Pakes model and method, Ackerberg-Caves-Frazer study different that could be combined with Olley-Pakes control function approach to identify the parameters of the PF. For identification, we need some source of exogenous variability in labor demand that is independent of productivity and that does not affect capital investment. Ackerberg-Caves-Frazer discuss several possible arguments/assumptions that could incorporate in the model this kind of exogenous variability.

Consider a model with same specification of the PF, but with the following specification of labor demand and optimal capital investment:

$$\begin{aligned} (LD') \quad & l_{it} = f_L\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}^L\right) \\ (ID') \quad & i_{it} = f_K\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}^K\right) \end{aligned} \tag{4.33}$$

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 $(t, i_{it}, l_{i,t-1}, k_{it})$: (a) r_{it}^L has cross-sectional variation, that is, $var(r_{it}^L \,|t, i_{it}, l_{i,t-1}, k_{it}) > 0$; 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 variability; (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 (2008) consider similar assumptions for the estimation of a production function with physical capital, permanent employment, and temporary employment.

4.5.1. *Other identifying conditions: Quasi-fixed inputs.* Consider a CD-PF with labor and capital as only inputs. Suppose that 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, EL 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:

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

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.

4.5.2. *Other identifying conditions: F.O.C. for flexible inputs.* Klette & Grilliches (1996), Doraszelski & Jaumandreu (2013), and Gandhi, Navarro, & Rivers (2013) 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:

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

The difference between these two equations is:

$$\ln(s_{it}^M) = \ln(\beta_M) + e_{it}$$

where $s_{it}^M \equiv \dfrac{P_t^M M_{it}}{P_t Y_{it}}$ is the ratio between material expenditures and revenue. The parameter(s) of the flexible inputs are identified from the expensiture-share equations. The parameter(s) of the quasi-fixed inputs are identified using the dynamic conditions described above.

Gandhi, Navarro, & Rivers (2013) show that this approach can be extended in two important way: (1) To a nonparametric specification of the production function: $y_{it} =$

$f(\ell_{it}, k_{it}, m_{it}) + \omega_{it} + e_{it}$; (2) - To a model with monopolistic competition (instead of perfect competition) with and isoelastic product demand. This approach relies on two strong assumptions. There is not any bias or missing parameter in the MC of the flexible input. Suppose that $MC_{Mt} = P_t^M \tau$, then our estimate of β_M will actually estimate $\beta_M \tau$. In its current version, this method cannot incorporate oligopoly competition in the product market, only a limited form of monopolistic competition.

4.6. Olley and Pakes on Endogenous Selection. Olley and Pakes (1996) show that there is a structure that permits to control for selection bias without a parametric assumption on the distribution of the unobservables. Before describing the approach proposed by Olley and Pakes, it will be helpful to describe some general features of semiparametric selection models.

Consider a selection model with outcome equation,

$$y_i = \begin{cases} x_i \beta + \varepsilon_i & \text{if} \quad d_i = 1 \\ \\ \text{unobserved} & \text{if} \quad d_i = 0 \end{cases} \tag{4.34}$$

and selection equation

$$d_i = \begin{cases} 1 & \text{if} \quad h(z_i) - u_i \geq 0 \\ \\ 0 & \text{if} \quad h(z_i) - u_i < 0 \end{cases} \tag{4.35}$$

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,

$$y_i = x_i\beta + E(\varepsilon_i^{d=1} | x_i, z_i) + \tilde{\varepsilon}_i \tag{4.36}$$

where $E(\varepsilon_i^{d=1} | x_i, z_i)$ is the selection term. The new error term, $\tilde{\varepsilon}_i$, is equal to $\varepsilon_i^{d=1} - E(\varepsilon_i^{d=1} | x_i, z_i)$ and, by construction, is mean independent of (x_i, z_i). The selection term is equal to $E(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i))$. 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, $E(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i)) = g(h(z_i))$. 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:

$$P_i \equiv \Pr(d_i = 1 \mid z_i) = F_u(h(z_i)) \tag{4.37}$$

where F_u is the CDF of u. Note that $P_i = E(d_i \mid z_i)$, 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:

$$y_i = x_i\beta + \lambda(P_i) + \tilde{\varepsilon}_i. \tag{4.38}$$

A sufficient condition for the identification of β (without a parametric assumption on λ) is that $E(x_i\ x_i' \mid P_i)$ has full rank. Given equation (4.38) and nonparametric estimates of propensity scores, we can estimate β and the function λ using standard estimators for partially linear model such as the kernel estimator in Robinson (1988), or alternative estimators as discussed in Yatchew (2003).

Now, we describe 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:

$$\phi_{it} = \begin{cases} \alpha_K\ k_{it} + \rho\ \phi_{i,t-1} + (-\rho\alpha_K)\ \ k_{i,t-1} + \xi_{it} & \text{if} \quad d_{it} = 1 \\ \\ \text{unobserved} & \text{if} \quad d_i = 0 \end{cases} \tag{4.39}$$

The exit/stay decision is: $\{d_{it} = 1\}$ iff $\{\omega_{it} \geq \omega^*(l_{it-1}, k_{it})\}$. 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 $\{\omega_{it} \geq \omega^*(l_{it-1}, k_{it})\}$ is equivalent to $\{\xi_{it} \leq \omega^*(l_{it-1}, k_{it}) - \rho(\phi_{i,t-1} - \alpha_K k_{it-1})\}$. Then, it is convenient to represent the exit/stay equation as:

$$d_{it} = \begin{cases} 1 & \text{if} \quad \xi_{it} \leq h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \\ \\ 0 & \text{if} \quad \xi_{it} > h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \end{cases} \tag{4.40}$$

where $h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \equiv \omega^*(l_{it-1}, k_{it}) - \rho(\phi_{i,t-1} - \alpha_K k_{it-1})$. The propensity score is $P_{it} \equiv E\left(d_{it} \mid l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}\right)$. And the equation controlling for selection is:

$$\phi_{it} = \alpha_K k_{it} + \rho\phi_{i,t-1} + (-\rho\alpha_K)\, k_{i,t-1} + \lambda(P_{it}) + \tilde{\xi}_{it} \tag{4.41}$$

where, by construction, $\tilde{\xi}_{it}$ is mean independent of k_{it}, k_{it-1}, $\phi_{i,t-1}$, and P_{it}. And we can estimation equation (4.41) using standard methods for partially linear models.

5. Innovation and productivity growth: Production functions

5.1. What determines productivity? Large and persistent differences in TFP across firms. Ubiquitous, even within narrowly defined industries and products. **Large:** 90th to 10th percentile TFP ratios: $\frac{A_{90th}}{A_{10th}}$. U.S. manufacturing, **average** within 4-digit SIC industries = **1.92** (Syverson, 2004). Denmark: average = **3.75** (Fox and Smeets, 2011). China or India, **average** > **5** (Hsieh & Klenow, 2009). **Persistent:** AR(1) of log-TFP with annual frequency: autoregressive coefficients between 0.6 to 0.8. **It matters:** Higher productivity producers are more likely to survive.

Why firms differ in their productivity levels? What supports such large productivity differences in 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? Why dispersion is possible in equilibrium? Let the profit of a firm be: $\pi_i = R(q_i, A_i, d) - C(q_i, A_i, w) - F$, where $R(q_i, A_i, d)$ is the revenue function, d is the state of the industry, $C(q_i, A_i, w)$ is the cost function, w represents input prices, and F is the fixed cost. **Key condition:** either $R$is strictly concave in q_i, or $C$is strictly convex in q_i. [The variable profit function is strictly concave].

Example: Perfect competition [or Bertrand competition with homogeneous product]. $R = P\ q_i$ is linear in q_i. We need $C$to be strictly convex. that is, DRS in variable inputs.

Example: Cournot competition or Bertrand competition with differentiated product. R is strictly concave in q_i (downward sloping demand). So we can have either CRS or DRS.

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

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

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

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

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

How can a firm affect its TFP? (HR) Managerial Practices. (Bloom & Van Reenen, 2007; Ichniowski and Shaw, 2003). Learning-by-Doing (Benkard, 2000). Organizational structure (vertical integration vs outsourcing). Higher-Quality (Labor and Capital) inputs. **Adoption of new (IT) technologies.**(Brynjolfsson et al., 2008). **Investment in R&D.** Long literature linking R&D investment and productivity. **Innovation**. Many firms undertake both process and product innovation without formally reporting R&D spending.

Innovation, R&D, and TFP. Multiple studies show evidence that R&D and innovation are very (the most?) important factors to explain firm heterogeneity in TFP level and growth. As usual, the main difficulty in these studies comes from separating causation from correlation. For the rest of this lecture, we review models, methods, and datasets in different empirical applications dealing with the causal effect of R&D and innovations on TFP.

But still we will have to address the question **"why firms have different propensities to innovate / invest in R&D?"**, for instance, managerial talent, competition, spillovers, ... What factors determine how large innovative activity will be? Can we predict whether product or process innovation will dominate, based on market features?

5.2. Application: Olley & Pakes on the US Telecom industry. US Telecom. equipment industry: 1974-1987. Technological change and deregulation. Elimination of barriers to entry. Antitrust decisions against AT&T: The Consent Decree (implemented in 1984) –> divestiture of AT&T. Substantial entry/exit of plants. Data from the US Census of manufacturers.

Table 3.2: Olley & Pakes (1995); Estimation Results

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

Going from OLS balanced panel to OLS full sample almost doubles β_K and reduces β_L by 20%. [Importance of endogenous exit]. Controlling for simultaneity further increases β_K and reduces β_L.

Table 3.3: Olley & Pakes (1995); Productivity estimates

TABLE XI

DECOMPOSITION OF PRODUCTIVITY[a]
(EQUATION (16))

Year	p_t	$\bar{p}_t$	$\Sigma_t \Delta s_{it} \Delta p_{it}$	$\rho(p_t, k_t)$
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

5.3. Levinshon & Petrin on Chilean manufacturing. The main difference with OP method is that LP use the demand function for intermediate inputs instead of the investment equation to invert out unobserved productivity. Two main motivations. (1) Investment can be responsive to more persistent shocks in TFP; materials is responsive to every shock in TFP. (2) In some datasets **Zero Investment** accounts for a large fraction of the data. At

$i_{it} = 0$ (corner solution / extensive margin) there is not invertibility between i_{it} and ω_{it}. Problems: loss of efficiency; missing estimates of TFP for many observations.

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

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

Investment equation is replaced with demand for materials:

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

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

They maintain OP-2 [No other unobservables; $r_{it} = r_t$], OP-3 [Markov TFP], and OP-4 [Time-to-build]. Least squares estimation of parameter β_L and the nonparametric functions $\{\phi_t : t = 1, 2, ..., T\}$ in regression equation:

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

$\phi_t(\ell_{i,t-1}, k_{it}, m_{it}) = \beta_K \ k_{it} + \beta_M m_{it} + f_M^{-1}(\ell_{i,t-1}, k_{it}, m_{it}, r_t)$ and f_M^{-1} is the inverse function of f_M with respect to ω_{it}. The second step is also similar to OP's second step but in the model with the intermediate input. ϕ_{it} is estimated in 1st step; and $\phi_{it} = \beta_K \ k_{it} + \beta_M \ m_{it} + \omega_{it}$. Then,

$$\phi_{it} \quad = \quad \beta_K \ k_{it} + \beta_M \ m_{it} + h\left(\phi_{i,t-1} - \beta_K \ k_{i,t-1} - \beta_M \ m_{i,t-1}\right) + \xi_{it}$$

Important difference with OP: In this second step $E(m_{it} \ \xi_{it}) \neq 0$, that is, materials m_{it} is endogenous. LP propose two approaches: **"unrestricted method"**: instrument m_{it} with its lagged values [see GNR (2013) criticism]; **"restricted method"**: under statis input, price-taking: $\beta_M =$ Cost of materials/Revenue.

Empirical application. Plant-level data from 8 different Chilean manufacturing industries: 1979-1985.

Table 3.4: Levinsohn & Petrin (2003): Input shares

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

Table 3.5: Levinsohn & Petrin (2003): Frequency of nonzeroes

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

Table 3.6: Levinsohn & Petrin (2003): PF estimates

TABLE 4

Unrestricted and Restricted Parameter Estimates for 8 Industries

(Bootstrapped Standard Errors in Parentheses)

	Industry (ISIC Code)							
Input	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

6. Measuring the productivity effects of R&D

Investment in R&D and innovation is expensive. Investors (for instance, firms, policy makers) are interested in measuring its returns, private and social. **Process R&D:** Directed towards invention of new methods of production. **Product R&D:** Directed towards creation of new and improved goods. Both can increase the firm's TFP It can have also **spillover effects in other firms**: competition spillovers, and/or knowledge spillovers.

6.1. Knowledge capital model. Knowledge capital model (Grilliches, 1979). Most studies measuring returns to R&D are based on the estimation production function, that is, effect of R&D on TFP. Cobb-Douglas in logs:

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

k_{it} = log of **stock of physical capital;** k_{it}^R = log of **stock of knowledge capital**. A major difficulty is the **measurement of the stock of knowledge capital**.

Measurement of knowledge capital. We observe firms' R&D expenses, R_{it}. How to construct K_{it}^R? **Perpetual inventory method.** Given $\{R_{it} : t = 1, 2, ...T_i\}$, the transition rule:

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

and values for δ_R and K_{i0} we can construct $\{K_{it}^R : t = 1, 2, ...T_i\}$. How to choose δ_R and K_{i0}? It is very difficult to know the true value of the rate of technological obsolescence, δ_R: it can be endogenous, vary across industries and firms, ...

Different studies using patent renewal data (Pakes and Schankerman, 1984; Pakes, 1986) or Tobin's Q model (Hall, 2005) estimate depreciation rates ranging between 10% and 35%.

Different authors (for instance, Grilliches and Mairesse, 1984) 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%.

6.2. Extending Knowledge capital model: Doraszelski & Jaumandreu (2013). In their model, TFP and Knowledge capital (KC) are unobservables to the researcher. They follow a stochastic process that is endogenous and depends on (observable) R&D investments. The model accounts for **uncertainty and heterogeneity** across firms in the link between R&D and TFP.

It 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 PF and stochastic process of KC, they exploit first order conditions for variable inputs.

Model. The PF (in logs) is:

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

log-TFP ω_{it} follows a stochastic process with transition probability $p\left(\omega_{it+1} \mid \omega_{it}, \; r_{it}\right)$ 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.

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

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

TFP stochastic process. The Markov structure of log-TFP implies:

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

where $\mathbb{E}\left[\xi_{it} \mid \omega_{it-1}, \; r_{it-1}\right] = 0$. The *productivity innovation* ξ_{it} captures to sources of uncertainty for the firm: the naturally linked to the evolution of TFP; the uncertainty inherent to R&D (for instance, chance of discovery, degree of applicability, success in implementation).

Marginal Revenue (Market Power). D&J identification approach exploits marginal conditions (MR = MC) for variable inputs. This requires an assumption about competition/market power. They assume:

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

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

VMP of labor = wage. This 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:

$$\omega_{it} = \lambda - \beta_K\ k_{it} + (1 - \beta_L - \beta_M)\ \ell_{it} + (1 - \beta_M)\ (w_{it} - p_{it})$$

$$+\beta_M\ (p_{Mit} - p_{it}) - \ln\left(1 - \frac{1}{\eta(p_{it}, d_{it})}\right)$$

We represent the RHS as $h(x_{it}, \beta)$, such that $\omega_{it} = h(x_{it}, \beta)$, with:

$$x_{it} = (k_{it},\ \ell_{it},\ w_{it},\ p_{Mit},\ p_{it},\ d_{it})$$

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

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

And form the marginal condition for labor we have:

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

The "parameters" in this system of equations are: β_L, β_K, β_M, g, and η. The unobservables ξ_{it} and e_{it} is 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 exogenous because time-to-build. **But we need to instrument ℓ_{it} and m_{it}.**

Identification. 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 $g\left[\omega_{t-1},\ r_{t-1}\right] = \rho_\omega\ \omega_{t-1} + \rho_r\ r_{t-1}$. Them we have:

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

Using as instruments $Z_{it} = (y_{it-1}, \ell_{it-1}, w_{it-1}{-}p_{it-1}, r_{it-1})$, moment conditions $\mathbb{E}\left[Z_{it}\ (\xi_{it} + e_{it})\right] = 0$ 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:

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

such that ξ_{it} is identified, and so 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. Panel of Spanish manufacturing firms ($N = 1,870$). Annual data for period $1990 - 1999$ (max $T_i = 10$). 10 industries (SIC 2-digits). Period of rapid growth in output and physical capital, coupled with stagnant employment. **R&D intensity = R&D expenditure / Sales.** 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 3.7: Doraszelski & Jaumandreu (2013): Descriptive statistics

TABLE 1
Descriptive statistics

Industry	Obs.[a]	Firms[a]	Entry[a] (%)	Exit[a] (%)	Rates of growth[a]: Output (s. d.)	Capital (s. d.)	Labour (s. d.)	Materials (s. d.)	Price (s. d.)	With R&D[b]: 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)

Production Function Estimates. Comparing GMM and OLS estimates, correcting for endogeneity has the expected implications, for instance, β_L and β_M decline, and $\beta_{K.}$ increases. There are not big differences in the β estimates across industries. Test of OIR from instruments: Cannot be rejected at 5% level. Test of parameter restrictions (in the two equations): Rejected at 5% level only in 2 out of 10 industries.

Table 3.8: Doraszelski & Jaumandreu (2013): PF estimates

TABLE 2
Production function estimates and specification tests

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

Stochastic Process for TFP. The model where TFP is exogenous (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%. Very significant uncertainty of the effect of R&D on TFP. Significant differences across industries.

Table 3.9: Doraszelski & Jaumandreu (2013): Stochastic process TFP

Industry	Exogeneity test		Separability test		$\frac{Var(e_{jt})}{Var(\omega_{jt})}$	$\frac{Var(\xi_{jt})}{Var(\omega_{jt})}$
	$\chi^2(10)$	p val.	$\chi^2(3)$	p val.		
	(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

Testing three versions of the Knowledge capital model. **Basic model:** $\omega_{it} + e_{it} = \beta_R$ $k_{it}^R + e_{it}$. Rejected for all industries. **Hall & Hayashi (1989) and Klette (1996) KC model.** $K_{it}^R = \left[K_{it-1}^R\right]^{\sigma} \left[1 + R_{it-1}^R\right]^{1-\sigma} \exp\{\xi_{it}\}$. Using D&J notation: $\omega_{it} = \sigma\, \omega_{it-1} + (1-\sigma)$ $r_{it-1} + \xi_{it}$. Rejected at 5% in 8 industries, and at 7% in all industries. **Model with:** β_R $k_{it}^R + \omega_{it} + e_{it}$, and ω_{it} with exogenous Markov process. Rejected at 5% in 2 industries, and at 10% in 6 industries.

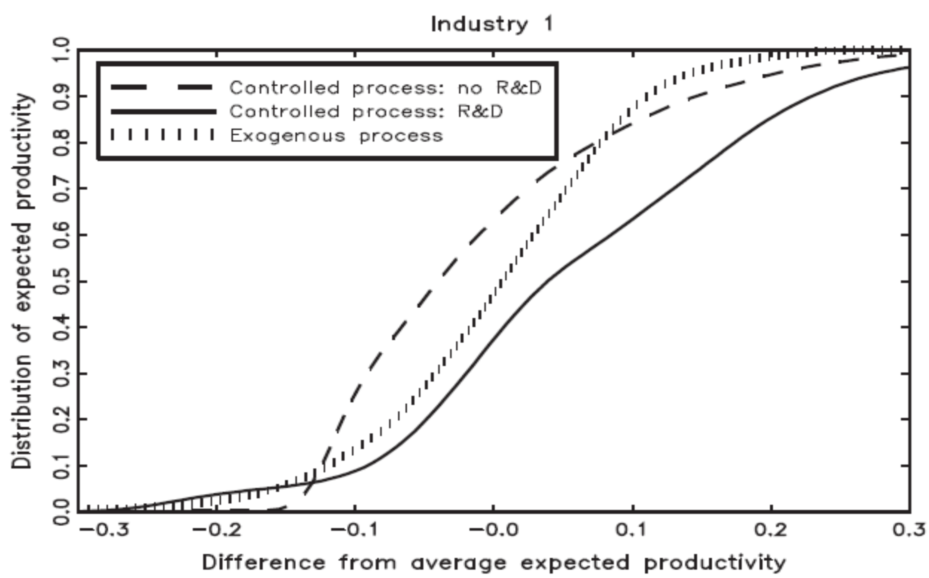
Table 3.10: Doraszelski & Jaumandreu (2013): Testing Knowledge capital

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

R&D and TFP (Counterfactuals). 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.

Figure 3.1: Doraszelski & Jaumandreu (2013): R&D and productivity



Elasticities of TFP w.r.t. R&D and lagged TFP. **Elasticity w.r.t. R&D:** Considerable variation between and within industries. Average is 0.015. **Degree of persistence:** Considerable between and within industries. Non-performers have a higher degree of persistence than performers. Persistence is negatively related to the degree of uncertainty.

Table 3.11: Doraszelski & Jaumandreu (2013): Elasticity TFP lagged R&D

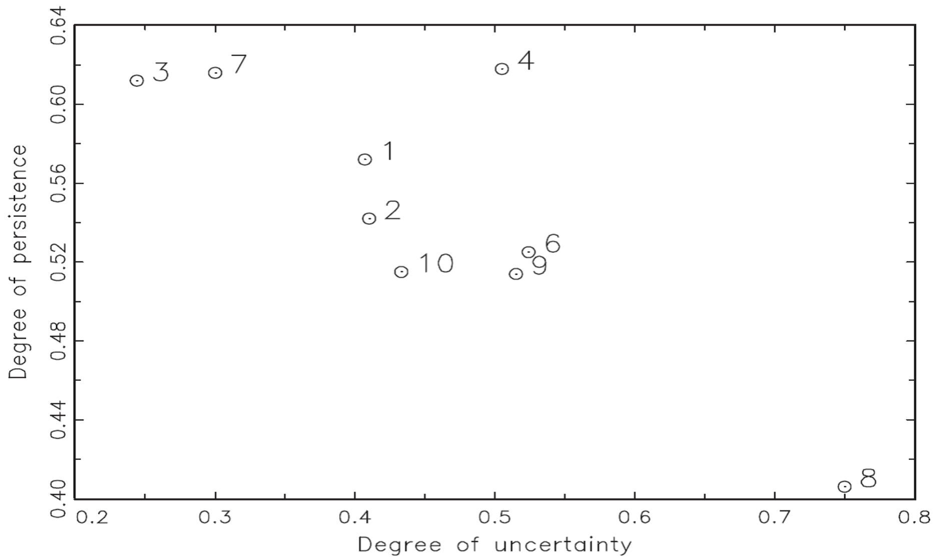
Industry	Elasticity wrt. R_{jt-1}[a] Q_1	Q_2	Q_3	Mean
	(1)	(2)	(3)	(4)
1. Metals and metal products	−0.013	0.007	0.021	0.022
2. Non-metallic minerals	−0.018	−0.012	0.000	−0.006
3. Chemical products	0.009	0.011	0.014	0.013
4. Agric. and ind. machinery	−0.017	−0.009	0.021	0.005
6. Transport equipment	−0.034	−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

Table 3.9: Doraszelski & Jaumandreu (2013): Elasticity TFP lagged TFP

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

Table 3.10: Doraszelski & Jaumandreu (2013): Uncertainty & persistence TFP



Summary of results. They 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 nonlinearity. 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.

7. Exercises

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.

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.

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, $A_1 > A_2 > A_3 > .....$

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.

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.

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.

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.

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
- 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.

Bibliography

[1] Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," Chapter 63 in Handbook of Econometrics, vol. 6A, James J. Heckman and Ed Leamer, eds. North-Holland Press.

[2] Ackerberg, D., K. Caves and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators, " *Econometrica*, 83(6), 2411-2451.

[3] Aguirregabiria, V. and Alonso-Borrego, C. (2014): "Labor Contracts and Flexibility: Evidence from a Labor Market Reform in Spain," manuscript. Department of Economics. University of Toronto.

[4] Alonso-Borrego, C., and R. Sanchez-Mangas (2001): "GMM Estimation of a Production Function with Panel Data: An Application to Spanish Manufacturing Firms," Statistics and Econometrics Working Papers #015527. Universidad Carlos III.

[5] Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, 277-297.

[6] Arellano, M., and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," Journal of Econometrics, 68, 29-51.

[7] Aw, B., Roberts, M. and Xu, D. (2011), "R&D Investment, Exporting, and Productivity Dynamics", American Economic Review, 101, 1312-1344.

[8] Blundell, R., and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," Journal of Econometrics, 87, 115-143.

[9] Blundell, R., and S. Bond (2000): "GMM estimation with persistent panel data: an application to production functions," Econometric Reviews, 19(3), 321-340.

[10] Bond, S., and M. Söderbom (2005): "Adjustment costs and the identification of Cobb Douglas production functions," IFS Working Papers W05/04, Institute for Fiscal Studies.

[11] Bond, S. and J. Van Reenen (2007): "Microeconometric Models of Investment and Employment," in J. Heckman and E. Leamer (editors) *Handbook of Econometrics, Vol. 6A*. North Holland. Amsterdam.

[12] Cobb, C. and P. Douglas (1928): "A Theory of Production," American Economic Review, 18(1), 139-165.

[13] Crepon, B., E. Guguet, and J. Mairesse (1998): "Research, Innovation and Productivity: An Econometric Analysis at the Firm Level," Economics of Innovation and new Technology, 7(2), 115-158.

[14] Doraszelski, U., and J. Jaumandreu (2013): "R&D and Productivity" Estimating Endogenous Productivity," Review of Economic Studies, , 80, 1338 - 1383.

[15] Gandhi, A., S. Navarro, and D. Rivers (2017): "On the identification of production functions: How heterogeneous is productivity?" Journal of Political Economy, forthcoming,

[16] Griliches, Z., and J. Mairesse (1998): "Production Functions: The Search for Identification," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*. S. Strøm (editor). Cambridge University Press. Cambridge, UK.

[17] Hall, B., J. Mairesse, and P. Mohnen (2010): "Measuring the Returns to R&D", in Hall, B. and Rosenberg, N. (eds), Handbook of the Economics of Innovation. North-Holland.

[18] Kasahara, H. (2009): "Temporary Increases in Tariffs and Investment: The Chilean Case," Journal of Business and Economic Statistics, 27(1), 113-127.

[19] Levinshon, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," Review of Economic Studies , 70, 317-342.

[20] Marschak, J. (1953): "Economic measurements for policy and prediction," in *Studies in Econometric Method*, eds. W. Hood and T. Koopmans. New York. Wiley.

[21] Marshak, J., and W. Andrews (1944): "Random simultaneous equation and the theory of production," Econometrica, 12, 143–205.

[22] Mundlak, Y. (1961): "Empirical Production Function Free of Management Bias," Journal of Farm Economics, 43, 44-56.

[23] Mundlak, Y., and I. Hoch (1965): "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," Econometrica, 33, 814-828.

[24] Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry", *Econometrica,* 64, 1263-97.

[25] Pakes, A. (1994): "Dynamic structural models, problems and prospects," in C. Sims (ed.) Advances in Econometrics. Sixth World Congress, Cambridge University Press.

[26] Syverson, C. (2011): "What Determines Productivity?" Journal of Economic Literature, 49:2, 326–365

[27] Wooldridge, J. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," Economics Letters, 104, 112-114.