Econometric Issues and Methods in the Estimation of Production Functions

Aguirregabiria, Victor

Department of Economics. University of Toronto

25 June 2009

Econometric Issues and Methods in the Estimation of Production Functions

Victor Aguirregabiria
University of Toronto

June 25, 2009

Abstract

This paper discusses the main econometric issues in the identification and estimation of production functions, and reviews recent methods. The main emphasis of the paper is in explaining the role of different identifying assumptions used in alternative estimation methods.

Keywords: Production Function Estimation. Dynamic Panel Data Models. Endogeneity. Sample Selection.
JEL codes: C10, C35, C63.

Corresponding Author: Victor Aguirregabiria. Address: 150 St. George Street. Toronto, ON, M5S 3G7. Phone: (416) 978-4358. E-mail: victor.aguirregabiria@utoronto.ca

1 Introduction

Production functions (PF) are important components of many economic models. The estimation of PFs plays a key role in the empirical analysis of issues such as the contribution of different factors to productivity growth; complementarity and substitutability of inputs; 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.

There are some important econometric issues in the estimation of productions functions. (a) Data problems: measurement error in output (typically we observe revenue but not output, and we do not have prices at the firm level); measurement error in capital (we observe the book value of capital, but not the economic value of capital); differences in the quality of labor; etc.

(b) Specification problems: Functional form assumptions, particularly when we have different types of labor and capital inputs such that there may be both complementarity and substitutability.

(c) Simultaneity: Observed inputs (e.g., labor, capital) may be correlated with unobserved inputs or productivity shocks (e.g., managerial ability, quality of land, materials, capacity utilization). This correlation introduces biases in some estimators of PF parameters.

(d) Multicollinearity: Typically, labor and capital inputs are highly correlated with each other. This collinearity may be an important problem for the precise estimation of PF parameters.

(e) Endogenous Exit/Selection: 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 introduces selection-biases in some estimators of PF parameters.

In this paper, I concentrate on the problems of simultaneity and endogenous exit, and on different solutions that have been proposed to deal with these issues. For the sake of simplicity, I 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 fact, some of the estimation approaches could be generalized to estimate nonparametric specifications of PF.

It is important to emphasize that different estimation approaches are based on different identification assumptions. Some assumptions can be more plausible for some applications (industries, markets) than for others. One of the main goals of this paper is to explain the role of different identifying assumptions used in alternative estimation methods.

The rest of the paper is organized as follows. Section 2 discusses the simultaneity problem and different approaches to deal with this issues. Section 3 concentrates on the problem of endogenous exit. Section 4 summarizes and concludes.

2 Simultaneity Problem

Consider a random sample of N firms, indexed by i, with information on the logarithm of output (y_i), the logarithm of labor (l_i), and the logarithm of physical capital (k_i): $\{y_i,\ l_i,\ k_i : i = 1, 2, ..., N\}$. Throughout the paper, I consider that all the observed variables are in mean deviations. Therefore, I omit constant terms in all the equations. We are interested in the estimation of the Cobb-Douglas PF (in logs):

$$y_i = \alpha_L\ l_i + \alpha_K\ k_i + \omega_i + e_i \tag{1}$$

α_L and α_K are technological parameters. ω_i 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. I refer to ω_i as *total factor productivity (TFP),* or *unobserved productivity,* or *productivity shock.* e_i represents measurement error in output, or any shock affecting output that is unknown to the firm when it decides capital and labor. Throughout the paper, the error term e_i is assumed to be independent of inputs and of the productivity shock. I use the variable y_i^e to represent the "true" value of output, $y_i^e \equiv y_i - e_i$.

The seminal paper by Marshak and Andrews (Econometrica, 1944) presented what probably is the first discussion of the simultaneity problem in the estimation of PF. If ω_i is known to the firm when it decides (k_i, l_i), then observed inputs will be correlated with the unobserved ω_i and the OLS estimator of α_L and α_K will be biased.

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. The model can be described in terms of two equations. The production function:

$$y_i \;=\; \alpha_L\ l_i + \omega_i + e_i \tag{2}$$

and the condition for profit maximization, i.e., marginal product is equal to the real wage:[1]

$$\alpha_L \frac{\exp\{y_i^e\}}{\exp\{l_i\}} = W \tag{3}$$

where W represents the real wage. Note that W is the same for all the firms because, by assumption, they operate in the same competitive output and input markets. The reduced form equations of this structural model are:

$$\begin{aligned} y_i &= \frac{\omega_i}{1-\alpha_L} + e_i \\ l_i &= \frac{\omega_i}{1-\alpha_L} \end{aligned} \tag{4}$$

Note that, $Cov(y_i, l_i) = E\left(\left[\frac{\omega_i}{1-\alpha_L} + e_i\right]\frac{\omega_i}{1-\alpha_L}\right) = Var(l_i)$. Therefore, the OLS estimator of α_L is such that:

$$p\lim_{N\to\infty} \hat{\alpha}_L = p\lim_{N\to\infty} \frac{\sum_{i=1}^N y_i l_i}{\sum_{i=1}^N {l_i}^2} = \frac{Cov(y_i, l_i)}{Var(l_i)} = 1 \tag{5}$$

That is, the OLS estimator of α_L converges in probability to 1 regardless the true value of α_L. Even if the hypothetical case that labor is not productive and $\alpha_L = 0$, the OLS estimator converges in probability to 1. It is clear that the OLS estimator can be seriously biased.

Example 2: Consider the similar conditions as in Example 1, but now firms in our sample produce differentiated products and use differentiated labor inputs. The model can be described in terms of two equations: the production function (2), and the profit maximization equation $\alpha_L \exp\{y_i^e\}/\exp\{l_i\} = W_i$. The key difference with respect to Example 1 is that now the real wage W_i has sample variation across firms. The reduced form equations for this model are:

$$\begin{aligned} y_i &= \frac{\omega_i - r_i}{1-\alpha_L} + r_i + e_i \\ l_i &= \frac{\omega_i - r_i}{1-\alpha_L} \end{aligned} \tag{6}$$

where $r_i = \ln(Wage_i)$. Therefore, the OLS estimator of α_L is such that:

$$p\lim_{N\to\infty} \hat{\alpha}_L = \frac{Cov(y_i, l_i)}{Var(l_i)} = 1 + \frac{Cov(l_i, r_i)}{Var(l_i)} \tag{7}$$

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

For instance, suppose that $Cov(\omega_i, r_i) = 0$, then:

$$Bias(\hat{\alpha}_L) = (1 - \alpha_L) \frac{Var(\omega_i)}{Var(\omega_i) + Var(r_i)} \tag{8}$$

This bias of the OLS estimator in this model is smaller that the bias in Example 1.[2] 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. In fact, the bias of the OLS estimator goes to zero as $Var(r_i)/Var(\omega_i)$ increases. Nevertheless, the bias can be very significant if the exogenous variability in input prices is not much larger than the variability in unobserved productivity.

The rest of this section discusses different estimators which try to deal with this endogeneity or simultaneity problem.

2.1 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 $cov(\omega_i, r_i) \neq 0$, i.e., 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.

[2] The model in Example 1 is a particular case of the model in Example2, i.e., the case when $Var(r_i) = 0$.

2.2 Panel Data: Within-Firms Estimator

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:

$$y_{it} = \alpha_L \; l_{it} + \alpha_K \; k_{it} + \omega_{it} + e_{it} \tag{9}$$

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 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 intepreted as weather, or other random and unpredictable shock.

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

Assumption PD-4: 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 $\bar{k}_i \equiv T^{-1} \sum_{t=1}^{T} k_{it}$.

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

$$(y_{it} - \bar{y}_i) = \alpha_L \left(l_{it} - \bar{l}_i\right) + \alpha_K \left(k_{it} - \bar{k}_i\right) + (\omega_{it} - \bar{\omega}_i) + (e_{it} - \bar{e}_i) \tag{10}$$

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 (i.e., 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):

$$cov\left(x_{it}, \omega_{is}^{*}\right) = cov\left(x_{it}, e_{is}\right) = 0 \tag{11}$$

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-2) and (PD-3) 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:

$$Bias(\hat{\alpha}_L^{OLS}) = \frac{Cov(k, \eta)}{Var(k) + Var(e^k)} - \frac{\alpha_L \; Var(e^k)}{Var(k) + Var(e^k)} \tag{12}$$

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:

$$Bias(\hat{\alpha}_L^{WGE}) = -\frac{\alpha_L \; Var(\Delta e^k)}{Var(\Delta k) + Var(\Delta e^k)} \tag{13}$$

Suppose that k_{it} is very persistent (i.e., $Var(k)$ is much larger than $Var(\Delta k)$) and that e_{it}^{k} is not serially correlated (i.e., $Var(\Delta e^k) = 2 * Var(e^k)$). 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.

2.3 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:

$$\Delta y_{it} = \alpha_L \ \Delta l_{it} + \alpha_K \ \Delta k_{it} + \Delta\delta_t + \Delta\omega^*_{it} + \Delta e_{it} \tag{14}$$

We maintain assumptions (PD-1), (PD-3), and (PD-4), but we remove assumption (PD-2). 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:

$$y_{it} = \alpha\ x_{it} + \eta_i + \omega^*_{it} + e_{it} \tag{15}$$

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

$$x_{it} = \delta\ x_{i,t-1} + \lambda_1\ \eta_i + \lambda_2\ \omega^*_{it} \tag{16}$$

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:

$$\Delta y_{it} = \alpha\ \Delta x_{it} + \Delta\omega^*_{it} + \Delta e_{it} \tag{17}$$

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:

$$\hat{\alpha}_N = \frac{\sum_{i=1}^{N} x_{i,t-2}\ \Delta y_{it}}{\sum_{i=1}^{N} x_{i,t-2}\ \Delta x_{it}} \tag{18}$$

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:

$$p\lim R^2 = \frac{Cov\left(\Delta x_{it}, x_{i,t-2}\right)^2}{Var\left(\Delta x_{it}\right)\ Var\left(x_{i,t-2}\right)} = \frac{\left(\gamma_2 - \gamma_1\right)^2}{2\left(\gamma_0 - \gamma_1\right)\gamma_0} \tag{19}$$

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:

$$
\begin{aligned}
\gamma_0 &= \frac{\lambda_1^2\ \sigma_\eta^2}{(1-\delta)^2} + \frac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2} \\
\gamma_1 &= \frac{\lambda_1^2\ \sigma_\eta^2}{(1-\delta)^2} + \delta\ \frac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2} \\
\gamma_2 &= \frac{\lambda_1^2\ \sigma_\eta^2}{(1-\delta)^2} + \delta^2\ \frac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2}
\end{aligned}
\tag{20}
$$

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:

$$
\begin{aligned}
R^2 &= \frac{\left(\delta\ \dfrac{\lambda_2^2\sigma_\omega^2}{1+\delta}\right)^2}{2\left(\dfrac{\lambda_2^2\sigma_\omega^2}{1+\delta}\right)\left(\dfrac{\lambda_1^2\ \sigma_\eta^2}{(1-\delta)^2} + \dfrac{\lambda_2^2\ \sigma_\omega^2}{1-\delta^2}\right)} \\
&= \frac{\delta^2\,(1-\delta)^2}{2\,(1-\delta+(1+\delta)\,\rho)}
\end{aligned}
\tag{21}
$$

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 (i.e., δ is close to zero) and when adjustment costs are large (i.e., 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 $\lambda_1^2\sigma_\eta^2 = \rho = 0$ (i.e., no fixed effects in the input demand). Then, $R^2 = \delta^2\,(1-\delta)\,/2$. 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, $R^2 = \delta^2\,(1-\delta)^2\,/4$. 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 $t_i^* \leq 0$ (i.e., 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:

$$
\begin{aligned}
x_{it_i^*} &= \frac{\lambda_1 \eta_i}{1-\delta} \\
y_{it_i^*} &= \alpha \, \frac{\lambda_1 \eta_i}{1-\delta} + \eta_i
\end{aligned} \tag{22}
$$

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

$$
\begin{aligned}
x_{it} &= x_{it_i^*} + \lambda_2 \left(\omega_{it}^* + \delta \omega_{it-1}^* + \delta^2 \omega_{it-2}^* + ... \right) \\
y_{it} &= y_{it_i^*} + \omega_{it}^* + \alpha \lambda_2 \left(\omega_{it}^* + \delta \omega_{it-1}^* + \delta^2 \omega_{it-2}^* + ... \right)
\end{aligned} \tag{23}
$$

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 $j > 0$:

$$
\begin{aligned}
E\left(\Delta x_{it-j}\left[\eta_i + \omega_{it}^* + e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(\Delta x_{it-j}\left[y_{it} - \alpha x_{it}\right]\right) = 0 \\
E\left(\Delta y_{it-j}\left[\eta_i + \omega_{it}^* + e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(\Delta y_{it-j}\left[y_{it} - \alpha x_{it}\right]\right) = 0
\end{aligned} \tag{24}
$$

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 $j > 1$:

$$
\begin{aligned}
E\left(x_{it-j}\left[\Delta \omega_{it}^* + \Delta e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(x_{it-j}\left[\Delta y_{it} - \alpha \Delta x_{it}\right]\right) = 0 \\
E\left(y_{it-j}\left[\Delta \omega_{it}^* + \Delta e_{it}\right]\right) = 0 \quad &\Rightarrow \quad E\left(y_{it-j}\left[\Delta y_{it} - \alpha \Delta x_{it}\right]\right) = 0
\end{aligned} \tag{25}
$$

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.

2.4 Control Function Approach

In a seminal paper, Olley and Pakes (1996) propose a control function approach to estimate PFs. Levinshon 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:

$$
\begin{array}{ll}
(PF) & y_{it} = \alpha_L \; l_{it} + \alpha_K \; k_{it} + \omega_{it} + e_{it} \\
(LD) & l_{it} = f_L \left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right) \\
(ID) & i_{it} = f_K \left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right)
\end{array}
\quad (26)
$$

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, i.e., 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.

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:

$$\begin{aligned} y_{it} &= \alpha_L \ l_{it} + \alpha_K \ k_{it} + f_L^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t) + e_{it} \\ &= \alpha_L \ l_{it} + \phi_t(l_{i,t-1}, k_{it}, i_{it}) + e_{it} \end{aligned} \tag{27}$$

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 (27) is a semiparametric partially linear model. The parameter α_L and the functions $\phi_1(.)$, $\phi_2(.)$, ..., $\phi_T(.)$ 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 (i.e., 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 $f_L\left(l_{i,t-1}, k_{it}, \omega_{it}, r_t\right)$ with respect to ω_{it}; (b) $r_{it} = r_t$, i.e., 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:

$$\omega_{it} = \ \gamma_1 \ i_{it} + \gamma_2 \ l_{i,t-1} + \gamma_3 \ k_{it} + r_{it} \tag{28}$$

Solving this equation into the PF, we have that:

$$y_{it} = \alpha_L \ l_{it} + (\alpha_K + \gamma_3) \ \ k_{it} + \gamma_1 \ i_{it} + \gamma_2 \ l_{i,t-1} + (r_{it} + e_{it}) \tag{29}$$

Note that current labor l_{it} is correlated with current input prices r_{it}. That is the reason why we need Assumption OP-2, i.e., $r_{it} = r_t$. 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)*, i.e., 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:

$$\omega_{it} = E[\omega_{it} \mid \omega_{i,t-1}] + \xi_{it} = h\left(\omega_{i,t-1}\right) + \xi_{it} \tag{30}$$

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:

$$\begin{aligned} \phi_{it} &= \alpha_K \ k_{it} + h\left(\omega_{i,t-1}\right) + \xi_{it} \\ &= \alpha_K \ k_{it} + h\left(\phi_{i,t-1} - \alpha_K \ k_{i,t-1}\right) + \xi_{it} \end{aligned} \tag{31}$$

Though we do not know the true value of ϕ_{it}, we have consistent estimates of these values from step 1: i.e., $\hat{\phi}_{it} = y_{it} - \hat{\alpha}_L \ l_{it}$.[3]

If function $h(.)$ is nonparametrically specified, equation (31) 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, i.e., 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 $h(.)$: i.e., $h(\omega) = \pi_1\omega + \pi_2\omega^2$. 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.

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

We 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:

$$\begin{aligned} \hat{\phi}_{it} &= \alpha_K k_{it} + \pi_1 \hat{\phi}_{i,t-1} + \pi_2 \hat{\phi}_{i,t-1}^2 + \left(-\pi_1 \alpha_K\right) k_{i,t-1} + (\pi_2 \alpha_K^2) k_{i,t-1}^2 \\ &+ \left(-2\pi_2 \alpha_K\right) \hat{\phi}_{i,t-1} k_{i,t-1} + \xi_{it} \end{aligned} \tag{32}$$

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 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:

$$\phi_{it} = \alpha_K\ k_{it} + \rho\ \phi_{i,t-1} + \left(-\rho\alpha_K\right)\ k_{i,t-1} + \xi_{it} \tag{33}$$

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}.

2.5 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:

$$l_{it} = f_L\left(l_{i,t-1},\ k_{it},\ f_K^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t),\ r_t\right) = G_t\left(l_{i,t-1}, k_{it}, i_{it}\right) \tag{34}$$

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 becuase 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:

$$y_{it} = \alpha_L \; l_{it} + (\alpha_K + \gamma_3) \;\; k_{it} + \gamma_1 \; i_{it} + \gamma_2 \; l_{i,t-1} + (\gamma_4 r_t + e_{it}) \tag{35}$$

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

$$l_{it} = (\delta_1 + \delta_3 \gamma_2) l_{i,t-1} + (\delta_2 + \delta_3 \gamma_3) k_{it} + \delta_3 \gamma_1 i_{it} + (\delta_4 + \delta_3 \gamma_4) r_t \tag{36}$$

Equation (36) 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 (35). 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 (36) 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 (35) 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:

$$\begin{array}{ll} (LD') & l_{it} = f_L\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}^L\right) \\ (ID') & i_{it} = f_K\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}^K\right) \end{array} \tag{37}$$

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, i.e., $var(r_{it}^L\ |t, i_{it}, l_{i,t-1}, k_{it}) > 0$; 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.

3 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.

3.1 Selection Bias Due to Endogenous Exit

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 (i.e., the scrapping value of the firm). Then, the optimal exit/stay decision is:

$$d_{it} = I\left\{ \; V^1(l_{it-1}, k_{it}, \omega_{it}) - V^0(l_{it-1}, k_{it}, \omega_{it}) \geq 0 \; \right\} \tag{38}$$

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, i.e., 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:

$$d_{it} = I\left\{ \; \omega_{it} \geq \omega^*\left(l_{it-1}, k_{it}\right) \; \right\} \tag{39}$$

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: i.e., $d_{it} = 1$. Therefore, the error term in the estimation of the PF is $\omega_{it}^{d=1} + e_{it}$, where:

$$\omega_{it}^{d=1} \;\; \equiv \;\; \{\omega_{it} \mid d_{it} = 1\} \;\; = \;\; \{\omega_{it} \mid \omega_{it} \geq \omega^*\left(l_{i,t-1}, k_{it}\right)\} \tag{40}$$

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,

$$\begin{aligned} E\left(\omega_{it}^{d=1} \mid l_{i,t-1}, k_{it}\right) &= E\left(\omega_{it} \mid l_{i,t-1}, k_{it}, d_{it} = 1\right) \\ &= E\left(\omega_{it} \mid l_{i,t-1}, k_{it}, \omega_{it} \geq \omega^*\left(l_{i,t-1}, k_{it}\right)\right) \\ &= \lambda\left(l_{i,t-1}, k_{it}\right) \end{aligned} \tag{41}$$

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

$$y_{it} = \;\; \alpha_L \; l_{it} + \alpha_K \; k_{it} + \lambda\left(l_{i,t-1}, k_{it}\right) + \tilde{\omega}_{it} + e_{it} \tag{42}$$

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 $\lambda(l_{i,t-1}, k_{it})$ introduces bias in our estimates of the PF parameters. The selection term is an increasing function of the threshold $\omega^*(l_{i,t-1}, k_{it})$, 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, i.e., 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:

$$\Delta s_{it} = \alpha + \beta \ s_{i,t-1} + \varepsilon_{it} \tag{43}$$

where s_{it} represents the logarithm of a measure of firm size, e.g., 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,

$$d_{it} = I\{\ \varepsilon_{it} \geq \varepsilon^*(s_{i,t-1})\ \} \tag{44}$$

where $\varepsilon^*(.)$ is a decreasing function, i.e., 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 $\varepsilon_{it}^{d=1} \equiv \{\varepsilon_{it} | d_{it} = 1\} = \{\varepsilon_{it} | \varepsilon_{it} \geq \varepsilon^*(s_{i,t-1})\}$. Thus,

$$\Delta s_{it} = \alpha + \beta s_{i,t-1} + \lambda(s_{i,t-1}) + \tilde{\varepsilon}_{it} \tag{45}$$

where $\lambda(s_{i,t-1}) \equiv E(\varepsilon_{it} | \varepsilon_{it} \geq \varepsilon^*(s_{i,t-1}))$, and $\tilde{\varepsilon}_{it} \equiv \{\varepsilon_{it}^{d=1} - \lambda(l_{i,t-1}, k_{it})\}$ that, by construction, is mean-independent of firm size at $t-1$. The selection term $\lambda(s_{i,t-1})$ is an increasing function of the threshold $\varepsilon^*(s_{i,t-1})$, 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 $\beta = 0$, i.e., 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, i.e., normality of the error term, and linear relationship between firm size and firm growth.

3.2 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,

$$y_i = \begin{cases} x_i\ \beta + \varepsilon_i & \text{if} \quad d_i = 1 \\ \text{unobserved} & \text{if} \quad d_i = 0 \end{cases} \tag{46}$$

and selection equation

$$d_i = \begin{cases} 1 \quad \text{if} \quad h(z_i) - u_i \geq 0 \\ 0 \quad \text{if} \quad h(z_i) - u_i < 0 \end{cases} \tag{47}$$

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,

$$y_i = x_i\beta + E(\varepsilon_i^{d=1} | x_i, z_i) + \tilde{\varepsilon}_i \tag{48}$$

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 $E\left(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i)\right)$. Given that u_i and ε_i are independent of (x_i, z_i), it is simple to show that the selection term depends on the regressors only through the function $h(z_i)$: i.e., $E\left(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i)\right) = g(h(z_i))$. The form of the function g depends on the distribution of the unobservables, and it is unknown if we adopt a nonparametric specification of that distribution. Therefore, we have the following partially linear model: $y_i = x_i\beta + g(h(z_i)) + \tilde{\varepsilon}_i$.

Define the *propensity score* P_i as:

$$P_i \equiv \Pr(d_i = 1 \mid z_i) = F_u(h(z_i)) \tag{49}$$

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:

$$y_i = x_i\beta + \lambda(P_i) + \tilde{\varepsilon}_i. \tag{50}$$

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 (50) 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 (i.e., 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:

$$\phi_{it} = \begin{cases} \alpha_K\ k_{it} + \rho\ \phi_{i,t-1} + (-\rho\alpha_K)\ k_{i,t-1} + \xi_{it} & \text{if} \quad d_{it} = 1 \\ \\ \text{unobserved} & \text{if} \quad d_i = 0 \end{cases} \tag{51}$$

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:

$$d_{it} = \begin{cases} 1 & \text{if} \quad \xi_{it} \leq h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \\ \\ 0 & \text{if} \quad \xi_{it} > h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \end{cases} \tag{52}$$

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:

$$\phi_{it} = \alpha_K k_{it} + \rho \phi_{i,t-1} + (-\rho \alpha_K)\, k_{i,t-1} + \lambda\,(P_{it}) + \tilde{\xi}_{it} \tag{53}$$

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 (53) using standard methods for partially linear models.

4 Conclusion

In this paper, I have discussed the simultaneity and sample selection problems in the identification and estimation of production functions, and I have reviewed the advantages and limitations of different estimation methods. The main emphasis of the paper has been to explain the role of different identifying assumptions used in alternative estimation methods.

References

[1] Ackerberg, D., K. Caves and G. Frazer (2006): "Structural Estimation of Production Functions," manuscript. Department of Economics, UCLA.

[2] Aguirregabiria, V. and Alonso-Borrego, C. (2008): "Labor Contracts and Flexibility: Evidence from a Labor Market Reform in Spain," manuscript. Department of Economics. University of Toronto.

[3] 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.

[4] Anderson, T.W., and C. Hsiao (1981): "Estimation of dynamic models with error components," Journal of the American Statistical Association, 589-606.

[5] Anderson, T.W., and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," Journal of Econometrics, 18, 47-82.

[6] 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.

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

[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] Evans, D. (1987): "The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," Journal of Industrial Economics, 35, 567-81.

[13] 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.

[14] Hall, B. (1987): "The Relationship Between Firm Size and Firm Growth in the US Manufacturing," Journal of Industrial Economics, 35, 583-606.

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

[16] Mansfield, E. (1962): "Entry, Gibrat's Law, innovation and the growth of firms," American Economic Review, 52, 1023-1051

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

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

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

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

[21] Robinson, P. (1988): "Root-N-Consistent Semiparametric Regression," Econometrica, 56, 931-54.

[22] Yatchew, A. (2003): "Semiparametric Regression for the Applied Econometrician," in Themes in Modern Econometrics, ed. P.C.B. Phillips, Cambridge University Press.