Estimating production functions when prices are partially observed*

Xulia González[†] and Daniel Miles-Touya[‡]

Universidad de Vigo

Abstract

One limitation when estimating production functions using firm-level data is that quantities are typically not reported. When ouput is proxied by deflated sales the omited price bias arises (Klette and Griliches, 1996). The aim of this paper is to propose a strategy to estimate production functions that overcome this problem when the price rate of change is observed but not the price level. In such cases the base year price appear as an unobserved omited variable in the production function and we treat it as a fixed effect. Our approach is based on the Olley and Pakes (1996) framework and the partially linear semiparametric model with fixed effects proposed by Su and Ullah (2011). We apply our approach to a survey of Spanish manufacturing firms and compare the results with standard methods. We find relevant differences in point estimates of the parameters although productivity measures are highly correlated.

Keywords: production function productivity estimation
JEL Classification: L11,L60

*We are grateful for suggestions given by Jose Carlos Fariñas, Ariel Pakes, Consuelo Pazó, and Mark Roberts. We also thank the comments of participants at NAPW 2014, Universidad Complutense seminar 2013, SAEE 2012, Encuentro de Economía Aplicada 2012, EARIE 2010 and Jornadas de Economía Industrial 2010. Financial support from the Ministerio de Economía y Competitividad and FEDER through grant ECO2015-69334-R is gratefully acknowledged. All errors remain our own responsibility.

[†]Universidad de Vigo y GRiEE. E-mail: xgzlez@uvigo.es.

[‡]Universidad de Vigo y RGEA. E-mail: dmiles@uvigo.es.

INTRODUCTION

The estimation of production function using firm level data plays a key role in empirical studies from different fields, including industrial organization, labor or trade. Indeed, meaningful policy parameters (markups, returns to scale, productivity), policy assessments (effects of trade liberalization, entry regularization, R&D subsidies, etc.), and aggregate productivity growth can all be derived by estimating production functions. However, such estimation raises a set of empirical issues that have no straightforward solution (Ackerberg et al., 2007; Syverson, 2011).

One well-known issue arises when the firm's optimal choice of inputs depends on its productivity, thus giving rise to an endogeneity bias. Structural estimation methods are increasingly used to address this problem; see, for example, Olley and Pakes (1996; hereafter OP), Levinsohn and Petrin (2003), or Ackerberg, Caves, and Frazer (2006; hereafter ACF). These authors adopt a control function approach to estimating the production function and exploit lagged input decisions or adjustment frictions as a source of identification.

Another set of issues concerns input and output measures, since firm-level databases typically report revenues rather than output quantities, and prices are seldom available. Klette and Griliches (1996) show that the standard estimators of the production function parameters are biased when firm's revenues (deflated by an industry price index) are used instead of quantities —namely, within-industry price differences are embodied in these output measures.[1] Additional problems arise when materials expenditures are also deflated by an industry price index (Ornaghi, 2006; Katayama *et al.*, 2009).

There are a number of firm-level datasets from different countries that report the firm's output (input) price rate of change but not its price levels.[2] With this information it is

[1] They show that using deflated sales as a proxy for real output will create a downward bias in the scale estimates (omitted price bias). Furthermore, when a revenue function is estimated rather than a production function, residuals obtained from the estimated specifications are combinations of demand and supply side shocks and, thus, measured "productivity" should be interpreted as variations in the "revenue generating ability" or profitability rather than technology shock.

[2] For example the Spanish Encuesta Sobre Estrategias Empresariales (ESEE) survey, the Bank of Italy's

possible to recover firm level prices (and therefore quantities) only by fixing the unobserved base-year price level to a constant—for example, normalizing it to 1. This has been the approach of several papers that estimate production function with these data sets (see e.g. Eslava *et al.*, 2004; Mairesse and Jaumandreu, 2005; Dolado *et al.* 2016; Ortigueira and Stuchi, 2016). The implication of this normalization is then that all firms in all industries have the same observed output price level (i.e., there is no intra or inter-industry price dispersion) in the base year despite the existence of price heterogeneity in all other years. From an econometric standpoint, this implies assuming that the measurement error -the base-year price level- is statistically independent of the right hand side variables in a production function equation i.e., firms are positioned randomly in the price distribution at the base-year.

However, several authors, for example, Roberts and Supina (2000) suggest that a firm's position in the distribution of output price levels depends on its structural factors (e.g., productivity advantages), which are slow to change over time; Foster et al. (2008) show that the firm's optimal output price is a function of persistent structural demand and productivity factors, which suggests a negative correlation between price level and productivity. In other terms, the unobserved price level at the base-year seems not to be completely arbitrary and could be correlated with the right-hand side variables of the production function equation. In that case, the estimators of the production function parameters may be biased.

Our paper contributes an approach to address this problem by considering the unobserved base-year price level as a fixed effect in the production function. In order to estimate the parameters in the presence of this fixed effect, besides the unobserved productivity, we rely on a modification of the procedure originally proposed by OP. That procedure comprises two stages: in the first stage, a proxy function replaces the unobservable productivity in

Survey on Italian Manufacturing Firms (INVIND), the Colombian Encuesta Anual de Manufacturas (EAM), and the French manufacturing survey. Less surveys report directly physical quantities as the Prowess data from the Centre for Monitoring the Indian Economy (De Loecker *et al*, 2016) or the Longitudinal Business Database at the U.S. Census Bureau for a set of industries (Foster, *et al*, 2016).

the production function leading to a semiparametric partially linear model. In this stage, the linear coefficients and the nonparametric function are estimated. In the second stage, the nonparametric function is used to recover the remaining parameters of the production function.

In the presence of the base-year price level fixed effect, the first stage of the OP procedure turns out to be a semiparametric partially linear model with fixed effects (Su and Ullah, 2010). Therefore, it is necessary to remove the fixed-effect to consistently estimate the nonparametric function to be used in the second stage. Here we employ the maximum likelihood estimator of Su and Ullah (2006; hereafter SU) to recover the nonparametric function and the linear coefficients in the first stage. The second stage of the OP procedure is not affected by the fixed effect given that it relies on the nonparametric function estimated in the first stage. Furthermore, we assess whether the base-year price level is a random effect, i.e., does not impact any decision of the firm, by means of a nonparametric Hausman test (Henderson *et al.* 2008).

We apply the procedure just described to an unbalanced panel of small and medium-size Spanish manufacturing firms. This survey reports the rate of firm-level price changes as well as relevant information on production, inputs, and firm characteristics, and it has been widely used to estimate production functions.

The results can be summarized as follows. We start by rejecting the Hausman test's null hypothesis that the output base-year price level is a random effect. Next, we find relevant differences in point estimates of the parameters when considering the base-year price level as a fixed effect versus standard methods although productivity measures are highly correlated. Finally, the productivity measures obtained when the base-year price level is viewed as a fixed effect exhibit systematically higher dispersion than those obtained when that level is viewed as a random effect.

The rest of the paper is organized as follows. Section 2 discusses the problem of not observing quantities and addresses the initial condition problem attendant upon constructing a price index using rate-of-change data. Section 3 presents our estimate of the production function. Section 4 describes data and discuss the results, and Section 5 concludes.

UNAVAILABLE QUANTITIES AND PARTIALLY OBSERVED FIRM-LEVEL PRICES

A limitation when using firm microdata to estimate production functions is that quantities are typically not observed. In this section, we address the problem that stems from deriving a firm-specific price index from data on the firm's output price level rate of change.

Following Klette and Griliches (1996; hereafter KG), we assume that the logarithm of the production function for the i^{th} manufacturing firm at time t can be represented:

$$q_{it} = \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \omega_{it} + \eta_{it} \tag{1}$$

where q_{it} denotes the logarithm of the quantity produced by firm i at time t, l_{it} the log of the number of workers, k_{it} the log of physical capital, and m_{it} the log of quantities of materials. The parameters $\beta_l, \beta_k, \beta_m$ relate input choices to output; ω_{it} is the unobservable productivity shock (resulting from, e.g., managerial ability and entrepreneurship or expected downtime due to machine repairs); and η_{it} captures all those shocks that affect production but cannot be anticipated or predicted by the firm when making its input decisions–or are due to pure measurement error.

The main empirical shortcoming with regard to equation (1) is that quantities are seldom available in many firm-level data sets, which instead report total revenue: $R_{it} = P_{it}Q_{it}$, where P_{it} is the ith firm's output price, usually not reported in firm-level data sets.

The standard approach is to deflate R_{it} by an industry price index, P_{It}, and thereby recover quantities through $\overline{R}_{it} = (P_{it}/P_{It})\, Q_{it}$. However, Klette and Griliches (1996) show that the estimators of the production function parameters are biased as a consequence of correlation between the firm's relative price (P_{it}/P_{It}) and the firm's input decisions.

Some firm-level data sets—such as the Spanish ESEE, the Bank of Italy's Survey on Italian Manufacturing Firms (INVIND), the Colombian Encuesta Anual de Manufacturas (EAM), and the French manufacturing survey—report the rate of change in the firm's output price. This firm price rate of change information has been used to obtain firm-specific price indices.

Let the i^{th} firm price rate of change be $\triangle P_{it}/P_{it-1}$, where $\triangle P_{it} = P_{it} - P_{it-1}$. Following Eslava *et al.* (2004), among others, the firm's output price level at time t is

$$P_{it} = \prod_{t=b+1}^{t} (1 + \triangle P_{it}/P_{it-1}) P_{ib},$$

where P_{ib} is the price level set by the firm for the year $t = b$, which is the base-year chosen by the analyst.

Hence, the firm's revenue deflated by the firm's specific price index is $R_{it}^{*} = (P_{it}Q_{it}) / (P_{it}/P_{ib}) = P_{ib}Q_{it}$; in logs, we have

$$q_{it} = r_{it}^{*} - p_{ib}.$$

The limitation of this approach is that p_{ib} is not observed in surveys that report the price rate of change, i.e., the nominal anchor is not identifiable. The standard approach has been to normalize the base-year price level to a particular constant—for instance $P_{ib} = 1$ for all i at $t = b$, which implies that $p_{ib} = 0$. However, such normalization implies that the firm-level prices at the base-year are completely arbitrary and does not impact any decision of the firm. More precisely, if base-year price levels were arbitrary we could define the productivity as $\widetilde{\omega}_{it} = \omega_{it} + p_{ib}$ and ignore the problem introduced by not observing the price level at the base-year.

In this paper, we consider p_{ib} as fixed-effects and test the random effects hypothesis by means of a nonparametric Hausman test. The price level at the base-year, p_{ib}, need not be statistically independent of the production function's explanatory variables, the firm's position in the distribution of output price levels depends on its structural factors (e.g., productivity advantages), which are slow to change over time (Robert and Supina (2000).

Notice that an analogous problems appears with intermediate goods. Here we follow the standard literature and assume that firms are price takers of an intermediate homogeneous input within an industry and intermediate input prices can differ across firms with respect to the industrywide price index as consequence of exogenous shock, independent of the state variables and not serially correlated.[3]

Therefore, the firm's intermediate goods price index can be recovered under the above assumptions and the input quantity is measured as $m_{it} = e_{it} - p_{it}^{m}$, where p_{it}^{m} is the log of the observed firm's specific input price index and e_{it} is the log of the firm's intermediate goods expenditure.

In sum, the empirical production function, where output is measured by $q_{it} = r_{it}^{*} - p_{ib}$, introduces a fixed effect into the production function (1):

$$r_{it}^{*} = \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \omega_{it} + \tau_i + \eta_{it},$$

where $\tau_i = p_{ib}$.

In the next section we present the OP method, discuss how the unobserved base-year price level fixed effect affects this method and show how semiparametric fixed-effect partially linear panel data can be used to estimate parameters of the production function.

[3] That is, let the input relative price be P_{it}^{m}/P_{It}^{m}, where P_{It}^{m} is the industry-wide intermediate materials price index and in logs $\vartheta_{it} = p_{it}^{m} - p_{It}^{m}$ where ϑ_{it} is the relative input price shock and $E\left(\vartheta_{it}\right) = 0$ in t for each i. The firm's log intermediate goods price level can be written as $p_{it}^{m} = F_{it}^{m} + p_{ib}^{m}$ where $F_{it}^{m} = \sum_{t=1}^{t} \log\left(1 + \triangle P_{it}^{m}/P_{it-1}^{m}\right)$ is observed from the data and p_{ib}^{m} is the unobserved base-year price level. Then, if we substract the intermediate materials industry price index in both sides of the previous expression $p_{it}^{m} - p_{It}^{m} = F_{it}^{m} - p_{It}^{m} + p_{ib}^{m}$ and take the time mean for each firm, $\overline{p}_{i}^{m} - \overline{p}_{I}^{m} = \left(\overline{F}_{i}^{m} - \overline{p}_{I}^{m}\right) + p_{ib}^{m}$where $\overline{p}_{i}^{m} - \overline{p}_{I}^{m} = \sum_t \left(\frac{p_{it}^{m} - p_{It}^{m}}{T}\right)$ and $\overline{F}_{i}^{m} = \sum_t \frac{F_{it}^{m}}{T}$. Given that the deviations of the input prices to the industry price index are assumed zero, $E\left(\vartheta_{it}\right) = 0$ in t, then, $\overline{p}_{i}^{m} - \overline{p}_{I}^{m} \simeq 0$ so that $p_{ib}^{m} \simeq \left(\overline{p}_{I}^{m} - \overline{F}_{i}^{m}\right)$. Doraszelski and Jaumandreu (2013, appendix) follow a similar approach as the one described above but assuming that $p_{i0}^{m} = \overline{F}_{i}$.

ESTIMATING THE PRODUCTION FUNCTION

In this section we describe how to obtain estimates of the production function coefficients. We rely on the insight of Olley and Pakes (1996) as regards proxies for unobserved productivity and follow the timing assumptions suggested by Ackerberg *et al.* (2006) when establishing conditions under which the parameters of interest can be identified. We start by describing the original OP procedure (with the ACF assumptions); we then introduce the modifications needed to estimate the coefficients with the base-year price level fixed effect.

The original OP approach considers firms that make production choices to maximize the present discounted value of current and future profits. A firm's production function resembles equation (1), where the unobserved productivity ω_{it} is assumed to follow an exogenous first-order Markov process, i.e., $\Pr\left(\omega_{it} | \{\omega_{it-j}\}_{j=1}^{t}\right) = \Pr(\omega_{it}|\omega_{it-1})$, which is stochastically increasing in ω_{it-1}. Inputs to be used in period t can be classify in flexible inputs, which can be adjusted in every period, e.g. materials, or quasi-flexible, subject to adjustment frictions, which are chosen at $t-1$, e.g. capital or labor. Furthermore, inputs can either be dynamic, in which case the current period's input choices affect the firm's future profits (these are so-called state variables), or nondynamic. In this economic environment, the firm's profit maximization problem results in an investment policy rule that depends on the firm's unobserved productivity and the state variables—namely, $i_{it} = f_t(\omega_{it}, k_{it}, l_{it}, p_{it})$ (Ackerberg *et al.*, 2007). Finally, in the original OP procedure firms are assumed to operate in perfectly competitive output and input markets in order to subsume the unobserved price levels into the time dummies, i.e., $i_{it} = f_t(\omega_{it}, k_{it}, l_{it}, p_{it}) = f_t(\omega_{it}, k_{it}, l_{it})$.

The original OP model relies crucially on the notion that this investment policy rule can be inverted to proxy for unobserved productivity in the production function equation via a function of investment and the state variables: $\omega_{it} = f_t^{-1}(i_{it}, k_{it}, l_{it})$. Pakes (1994) gives conditions for invertibility when only one unobservable affects firm behavior (i.e.„ a scalar unobservability assumption); the implication is that the investment function is strictly increasing in ω_{it} in the region where i_{it} is positive.[4]

Under these assumptions, the original OP method consists of two stages. In the first stage, the unobservable productivity in the production function is replaced by the inverse of the investment function (i.e.„ the proxy function) to yield a partial linear model. That is, in light of (1) we have:

$$q_{it} = \beta_m m_{it} + \phi_t\left(i_{it}, k_{it}, l_{it}\right) + \eta_{it}$$

where $\phi_t\left(i_{it}, k_{it}, l_{it}\right) = \beta_l l_{it} + \beta_k k_{it} + f_t^{-1}\left(i_{it}, k_{it}, l_{it}\right)$. In principle, this first stage could be used to estimate coefficients for the nondynamic variable inputs—that is, the intermediate inputs coefficient β_m as well as ϕ_t. We shall discuss the former in what follows.

The OP method's second stage identifies the parameters of the dynamic inputs by relying on the moment conditions that result from (a) the first-order Markov assumption, $\omega_{it} = E\left(\omega_{it}|\omega_{it-1}\right) + \varepsilon_{it} = g\left(\omega_{it-1}\right) + \varepsilon_{it}$ and (b) the time to build assumption, $E\left(k_{it}\varepsilon_{it}\right) = E\left(l_{it-1}\varepsilon_{it}\right) = 0$. We can recover the residual ε_{it} by nonparametrically regressing $\omega_{it}\left(\beta_k, \beta_l\right)$ on $g\left(\omega_{it-1}\left(\beta_k, \beta_l\right)\right)$, given that

$$\omega_{it}\left(\beta_k, \beta_l\right) = \widehat{\phi}_t - \left(\beta_l l_{it} + \beta_k k_{it}\right)$$

where $\widehat{\phi}_t$ is an estimator of $\phi_t\left(i_{it}, k_{it}, l_{it}\right)$ and then estimate the coefficients by minimizing the sample analogue of the moment conditions,

$$\frac{1}{N}\frac{1}{T}\sum \varepsilon_{it}\left(\beta_k, \beta_l\right)\begin{bmatrix} k_{it} \\ l_{it-1} \end{bmatrix}$$

using standard GMM techniques.[5]

[4] Recent papers have relaxed this scalar unobservable assumption by admitting additional unobservables in the investment equation (DeLoecker, 2007; Ackerberg et al., 2007; Huang and Hu, 2011).

[5] For initial values $\left(\beta_k^0, \beta_l^0\right)$ we have $\omega_{it}^0\left(\beta_k^0, \beta_l^0\right) = \widehat{\phi}_t - \left(\beta_l^0 l_{it} + \beta_k^0 k_{it}\right)$ and can estimate $\varepsilon_{it}\left(\beta_k^0, \beta_l^0\right)$ from the residuals of the nonparametric regression $\omega_{it}^0\left(\beta_k^0, \beta_l^0\right) = g\left(\omega_{it-1}^0\left(\beta_k^0, \beta_l^0\right)\right) + \varepsilon\left(\beta_k^0, \beta_l^0\right)$. The moment conditions allow us to obtain new values for $\left(\beta_k, \beta_l\right)$, after which we iterate until convergence. As usual in this

Ackerberg *et al.* (2015) discuss the ability of the OP procedure to identify the variable input coefficients in the first stage.[6] In order to allow for identification in the first stage, ACF propose additional timing assumptions concerning when productivity shocks hit the firm and when inputs are chosen.[7] The idea is to introduce an independent source of cross-sectional variability by assuming that input decisions are based on an information set other than that of investment.[8] Hereafter, we incorporate these ACF assumptions into the OP procedure.

Furthermore, as firm level prices are not usually observed in firm level datasets, and given that the OP procedure is identified under any degree of imperfect competition in the output market, it is standard in the empirical literature to invoke different assumptions so as to capture the unobserved price level variability in the optimal investment equation through observable variables, such as productivity, capital, inputs or aggregate demand (see, among other, Aw *et al.* 2011; De Loecker, 2011 or De Loecker and Warzynski 2012).[9] For example, Aw *et al.* (2011) assume a constant markup setting where price variability is explicitly captured through the marginal cost heterogeneity that arises through productivity and capital differences across firms; De Loecker (2011) considers that the constant elasticity

literature, the objective function is based on $\min\{\beta_k, \beta_l\} \sum_j^2 \left(\sum_{it} \widehat{\varepsilon}_{it} (\beta_k, \beta_l) Z_{it}^j \right)^2$, where $Z_{it} = (k_{it}, l_{it-1})$ and j indexes the elements of Z.

[6] Under OP assumptions, the variable input demand could be a function of productivity, in which case—conditional on the first stage's proxy function—no variability would be left for intermediate materials. That is, $m_{it} = g_t(\omega_{it}, k_{it}, l_{it})$ so $m_{it} = g_t\left(f_t^{-1}(i_{it}, k_{it}, l_{it}), k_{it}, l_{it}\right) = z_t(i_{it}, k_{it}, l_{it})$, which implies that β_m is not identified in the first stage because there remains no variation conditional on the nonparametric function.

[7] ACF present an alternative approach to dealing with the problem of multicollinearity: estimate only the nonparametric function in the first stage and then estimate all the parameters in the second stage. Thus the first-stage estimation would be given by

$$q_{it} = \phi_{it}^*(i_{it}, k_{it}, l_{it}, m_{it}) + \eta_{it}$$

and the second stage would use moment conditions to estimate the coefficients.

[8] So, suppose that ω_{it} evolves between subperiods $t-1$, $t-b$, and t according to a first-order Markov process and the m_{it} is chosen at $t-b$. In this case the firm's optimal material input will be a function not of ω_{it} but rather of ω_{it-b}, which resolves the multicollinearity problem and allows the first stage to be identified. The intuition is that intermediate goods are chosen without perfect information about ω_{it}, and this incomplete information is what moves m_{it} independently of the nonparametric function.

[9] The monotonicity needed in Olley and Pakes (1996) to assure invertibility of the investment function does not depend on the degree of competition in the output market (e.g. De Loecker, 2007, appendix C).

demand assumption is sufficient to justify that the unobservable prices variability is capture through the variation in inputs and aggregate demand, though prices are not explicitly modeled nor the marginal cost in the constant markup framework clearly stated; De Loecker and Warzynski (2012, pg 2446 and appendix) assume that the investment function depends in a set of observable variables that implicitly would capture price variability (i.e., additional variables affecting optimal decision), but there is no specific model of price determination where these variables can be derived from. In other terms, the standard approach is to capture the unobserved output price variables through those observable variables that are usually find in the literature to be behind the models that explain price variability, such as productivity, dynamic inputs, input prices, demand shifters, etc, but without specifying an explicit price model.

In our framework, output price levels are not observed -we only observe the price rate of change. Hence, following De Loecker and Warzynski (2012, appendix), who introduce a set of additional variables that could affect the optimal decision, i.e. $i_{it} = f_t\left(\omega_{it}, k_{it}, l_{it}, p_{it}^m, \mathbf{z}_{it}\right)$, we include as these additional variables the firm specific demand shocks (De Loecker, 2011) and the hourly wage rate. Spain has some of the most stringent regulations for permanent contracts among developed countries (Aguirregarabiria and Alonso-Borrego, 2014). The hourly wage rate tries to capture possible output price movements as a consequence of an increase in the effective utilization of labour, i.e. transitory working hours increase[10]. In sum, the proxy function is expressed as $\omega_{it} = f_t^{-1}\left(i_{it}, k_{it}, l_{it}, p_{it}^m, D_{it}, w_{it}\right)$, where D_{it} captures firm specific demand shifters and w_{it} the hourly wage rate.

The foregoing discussion indicates that the first-stage equation can be written as

$$r_{it}^* = \beta_m m_{it} + \phi_t\left(i_{it}, k_{it}, l_{it}, D_{it}, p_{it}^m, w_{it}\right) + \tau_i + \eta_{it},$$

when we substitute $q_{it} = r_{it}^* - p_{ib} = r_{it}^* - \tau_i$ and where $\phi_t\left(i_{it}, k_{it}, l_{it}, D_{it}, p_{it}^m, w_{it}\right) = \beta_l l_{it} +$

[10] Movement in the effective hours worked when labour is fixed (high adjustment costs) can be seen as movements in capital utilization.

$\beta_k k_{it} + \omega_{it}$.

As is argued by Asker, Collard-Wexler, and De Loecker's (2014) in Theorem 1, while the OP/ACF moments would not be affected by a fixed-effect in the productivity process, the first-stage is not valid unless the fixed effects are accounted for.

We next show how to modify the first stage of the OP procedure by means of semiparametric, partially linear, panel data fixed-effect models that enable us to estimate not only the first-stage nonparametric function but also parameters of the nondynamic inputs. The nonparametric function is then used in the second stage to estimate the dynamic inputs, just as in the original OP/ACF procedure.

Partially linear panel data models

The first stage of the OP method can be described as a partially linear panel data fixed-effect model. We employ the notation most commonly used in semiparametric partial linear models:[11]

$$y_{it} = x_{it}'\beta + m(z_{it}) + \tau_i + \eta_{it}, i = 1, ..., n; t = 1, ..., T_i.$$

Here y_{it} represents r_{it}^*; x_{it} is a $p \times 1$ vector of explanatory variables, including materials inputs; $m(\cdot) = \phi(\cdot)$; and $z_{it} = (i_{it}, k_{it}, l_{it}, D_{it}, p_{it}^m, w_{it})$, which is a $q \times 1$ vector.[12]

Recent years have seen an increasing number of papers that propose alternative estimation methods for use with fixed-effect nonparametric or semiparametric panel data models (Su and Ullah, 2010). Each of these papers proposes an alternative method of estimating the linear or nonparametric component. For example, Li and Stengos (1996) use a kernel instrumental variable to estimate the linear—but not the nonparametric—component. Henderson *et al.*

[11] These methods were developed for data from balanced panels: we have extended these procedures to unbalanced panels.

[12] Although OP, ACF, and Levinsohn and Petrin (2003) all assume that the functional form of the investment proxy is time dependent, in their empirical applications this form is presumed not to be time dependent. In fact, it is possible for these functions to depend on t, but that would require an estimation methodology not yet available in the literature.

(2008) propose a similar iterative kernel estimator for the nonparametric function that is, however, computationally more demanding.

In this paper we apply the profile likelihood estimation for the nonparametric component (i.e., profile least squares; Su and Ullah, 2006) for estimating the unknown component of the fixed-effects partially linear mod el. We have chosen this method because they are computationally less costly and have good asymptotic properties (Su and Ullah, 2011; Gao, 2014).

Su and Ullah (2006) propose estimating the semiparametric fixed-effect model by means of profile maximum likelihood (i.e.,, profile least squares).[13] Let the local linear approximation of their equation can be written as

$$\begin{aligned} y_{it} &= x_{it}'\beta + m(z) + \frac{\partial m}{\partial z}(z_{it} - z) + \tau_i + \eta_{it} \\ &= x_{it}'\beta + Z_{it}(z)\,\delta(z) + \tau_i + \eta_{it} \end{aligned}$$

where $\frac{\partial m}{\partial z} = \left(\frac{\partial m}{\partial z_1}, ..., \frac{\partial m}{\partial z_q}\right)$, $Z_{it}(z) = \left[1, (z_{it} - z)'\right]_{1\times(q+1)}$ and $\delta(z) = \left(m(z), \frac{\partial m}{\partial z}\right)'$.

The idea behind this approach is to profile out the individual effect and the linear parameters, $\theta = (\tau', \beta')'$ and then consider the concentrated least squares for $\delta(z)$. Therefore, if θ is known the estimate of $\delta(z)$ is given by

$$\delta_\theta(z) = \arg\min_{\delta \in R^{q+1}} \sum_i \sum_t \left(y_{it} - \tau_i - x_{it}'\beta - Z_{it}(z)\,\delta(z)\right)^2 K_H(z_{it}, z),$$

where $K_H(z_{it}, z) = \prod_{k=1}^{q} h_k^{-1} k\left(\frac{z_{it,k} - z}{h_k}\right)$ for $k(\cdot)$ a univariate kernel function and subject to the

[13] In the end a profile estimator is a two stage procedure. In the first stage it assumes known the coefficients (or the unknown function, depending on the method), and derives an estimator of the unknown function as a function of these coefficients. In the second stage, the estimator of the unknown function is plugged in the least square objective function from where a consistent estimator of the parameters is derived. And substituting back in the first stage we obtain an estimator of the nonparametric function.

identification condition $\sum_i \tau_i = 0$. Thus,

$$\delta_\theta(z) = \left[\sum_i \sum_t Z_{it}(z)' K_H(z_{it}, z) Z_{it}(z)\right]^{-1} \sum_i \sum_t Z_{it}(z)' K_H(z_{it}, z)(y_{it} - \tau_i - x_{it}'\beta).$$

However, $\delta_\theta(z)$ is not an operational term because it depends on the unknown parameter θ. Hence this method profiles out $m(z)$ by estimating θ as

$$\widehat{\theta} = \arg\min_\theta \sum_i \sum_t (y_{it} - \tau_i - x_{it}'\beta - m_\theta(z_{it}))^2$$

where $m_\theta(z_{it})$ is the first component of $\delta_\theta(z)$. Given the profile estimates $\widehat{\theta} = \left(\widehat{\tau}', \widehat{\beta}'\right)'$, the profile likelihood estimator of $m(z)$ is

$$\begin{aligned} \widehat{m}(z) &= m_{\widehat{\theta}}(z) = e'\delta_{\widehat{\theta}}(z) \\ &= e'S(z)\left(\mathbf{y} - D\widehat{\tau} - \mathbf{X}\widehat{\beta}\right) \end{aligned}$$

in matrix form, where $e = (1, \mathbf{0}_{q\times 1})_{(q+1)\times 1}$,

$$\begin{aligned} S(z) &= \left[Z(z)' K_H(z) Z(z)\right]^{-1} Z(z)' K_H(z) \\ Z(z) &= \left(Z_{11}(z)' \ldots Z_{1T_1}(z)' \ldots Z_{nT_n}(z)'\right)'_{R\times(q+1)}, \\ K_H(z) &= diag\left[K(z_{11} - z), K(z_{12} - z) \ldots K(z_{1T_1} - z) \ldots K(z_{nT_n} - z)\right]_{R\times R} \\ D_{R\times n-1} &= \left[(-\iota_{n-1}, I_{n-1}) \otimes \iota_{T_i}\right]_{i=1,\ldots,n} \\ \mathbf{y} &= (y_{11,\ldots} y_{nT_n})'_{R\times 1} \end{aligned}$$

for $R = \sum_{i=1}^n T_i, , \tau = (\tau_2, \ldots \tau_n)'$ and $\mathbf{X}_{R\times p}$.

The parameter profile estimates are given by

$$\widehat{\beta} = \left[\widetilde{X}'\widetilde{X}\right]^{-1}\widetilde{X}'\widetilde{\mathbf{y}}$$
$$\widehat{\tau} = \left[D^{*\prime}D^{*}\right]^{-1}D^{*\prime}\left(\mathbf{y}^{*} - X^{*\prime}\widehat{\beta}\right).$$

Here the asterisk * indicates the residual after regressing the respective variable on the matrix defined by $S = [s(z_{it})] = [e'S(z_{it})]_{\ i=1,\ldots,n;t=1,\ldots,T_n}$ and the tilde $\widetilde{\cdot}$ denotes the residuals of regressing X^{*} and $\mathbf{y}^{*}$ on D^{*}. Su and Ullah (2006) demonstrate the consistency of the nonparametric estimator and the $\sqrt{n}-$consistency of the parameters estimator. In Su and Ullah (2007) these authors develop a similar estimation procedure for random-effects models.

Nonparametric Hausman test

Finally, we test whether the unobserved base-year price level should be considered as a random effect rather than as a fixed effect by means of a nonparametric Hausman test for semiparametric panel data models (Henderson *et al.,* 2008). The null hypothesis is that τ_i is a random effect. Hence, the null could be written as

$H_0 : E(\tau_i | x_{it}, z_{it}) = 0$

where x_{it} capture the non-dynamic inputs and z_{it} the proxy function inputs, and where alternative hypothesis is the negation of the null. Since η_{it} represents all the shocks that affect production yet cannot be anticipated or predicted by the firm when making its input decisions, it follows that $E(\eta_{it} | x_{it}, z_{it}) = 0$. Therefore, the null and alternative hypotheses can be equivalently stated as

$H_0 : E(u_{it} | x_{it}, z_{it}) = 0$

$H_A : E(u_{it} | x_{it}, z_{it}) \neq 0$

here $u_{it} = \tau_i + \eta_{it}$

The proposed test is based on the sample analogue of $J = E\{u_{it}E(u_{it} | x_{it}, z_{it}) f(x_{it}, z_{it})\} =$

$E\left\{[E(u_{it}|x_{it}, z_{it})]^2 f(x_{it}, z_{it})\right\}$. If the model is assumed to be correctly specified, then the statistic equals zero under random effects -H_0– but exceeds zero under fixed effects -H_A – . In the Appendix B we describe the procedure.

It is relevant to remark that rejecting the Hausman test may suggest that specifications of the functional forms are wrong. In other words, in the field of applied industrial organization it is conventional to assume that functional forms behind the empirical implementation are correctly specified. Under this assumption, rejection of the null hypothesis suggests that the unobserved base-year price level should be considered as a fixed effect. However, rejection of the null hypothesis could instead imply that the functional forms assumption is inaccurate or could be due to the presence of measurement errors (Bierens, 2009).[14] That is, functional form misspecification or measurement errors can severely affect any nonlinear estimator, such as the semiparametric OP estimator, or the semiparametric the first-difference or fixed-effect estimators, leading to reject the null hypothesis.

In the next section we apply the methods just described to data from a survey of Spanish manufacturing firms.

DATA AND RESULTS

Our data comes from the Encuesta Sobre Estrategias Empresariales (ESEE) 1991–2006 survey, an unbalanced firm-level panel of Spanish manufacturing firms that is sponsored by the Ministry of Industry. These data have been frequently used to estimate the production function. We selected a subsample consisting of small and medium-size firms (i.e., firms with 10–199 workers) that exhibited positive investment levels and at least two consecutive periods in the sample. This selection criteria yielded a total of 10,484 observations of 1,616

[14] As a referee kindly points out, a related issue could be to test the linearity of $\omega_{it} = h(i_{it}, k_{it}, l_{it})$ by way of some recently developed consistent nonparametric tests–such as the one proposed in Su and Lu (2013) or in Li and Sun (2014). Not rejecting the null hypothesis greatly simplifies the estimation process because in that case semiparametric estimation procedure is not required. We leave this issue for future research.

firms belonging to ten different industrial sectors.

The first issue we address in this empirical section is testing the relevance of the fixed effect through the Hausman nonparametric test: not rejecting the null hypothesis implies that the unobserved base-year price level can be treated as random, enabling us to follow standard implementation of the OP/ACF procedure. We will apply the test on two commonly used specifications: first, we omit input prices and hourly wages from investment equation under the assumption that differences in output prices are exclusively a consequence of differences in the firm's efficiency index, capital, labour and specific demand shifters in constant markup setting with firms facing competitive input markets (Erickson and Pakes, 1995; Aw et al., 2011). Second, we apply the test to the specification discussed in the previous sections where we additionally include input prices and hourly wages to the proxy equation.

In Table 1 we present Hausman test statistics and p-values when SU is used to estimate the fixed-effect residual term needed to derive the test statistic. Implementing the SU estimator requires that we use a Gaussian kernel for continuous variables and the Racine and Li (2004) kernel in the presence of continuous and discrete variables. When selecting the bandwidth we follow the empirical literature; however, test results are robust to different bandwidths and orders of expansion.

Insert Table 1

Overall, the results in Table 1 reject the null hypothesis by which the unobserved base-year price level is a random effect. These results suggest two alternative interpretations depending on the assumption of correct functional form specification.

One the one hand, under the conventional IO assumption of a correct specification of the functional forms, these results suggest that a firm's position in the base-year price level distribution may not be randomly determined but instead be correlated with the explanatory variables of the production function.

Hence, this result implies that the normalization of the base-year price level through the same constant for all firms in all industries may bias the production function estimators.

On the other hand, the results of Table 1 may be suggesting that the conventional functional forms adopted in the empirical implementation of the OP/ACF procedure might be incorrect (Bierens, 2009).

For example, the designation of the dynamic inputs or the monotonicity assumptions may be incorrect leading to the rejection of the null hypothesis despite the fact that the base-year price level might not be correlated with the production function explanatory variables.

Likewise, the presence of measurement errors correlated with the regressors may bias the coefficient estimates and induce the test to reject the null hypothesis.

In Table 2 we report estimates of the production function parameters when the proxy function depends on capital, labor, input prices, demand shifters and a trend variable based on Olley and Pakes's (1996) as well as Su and Ullah's (2006) fixed-effects estimator.

Insert Table 2

These results confirm that there are differences in most of the parameter point estimates obtained when considering the base-year price level as fixed-effects compared with the standard OP/ACF approach.

Overall, our results show that, for most cases, the point estimates under the SU-FE estimator of the capital and labor coefficients are larger and the materials systematically lower than under the OP/ACF approach. In particular, capital coefficient is larger in seven of the ten industries analyzed, and labor is higher in six of them, while materials coefficient is lower in 9 industries under SU-FE estimator.

These point estimate movements leads to a RS closer to 1. For example, in the case of Chemical products industry, the estimated RS drops from 1.11 to 1.00 in the OP/ACF and SU-FE respectively. Similarly, in Transport & equipment industry the RS falls from 1.16 (OP/ACF) to 1.05 (SU-FE). The opposite happens in industries as Timber & furniture where the RS rises from 0.81 to 0.95 and Paper & printing industry, where the RS increases from 0.81 to 1.04 when moving from the standard OP framework to the fixed effect approach.

Even so, the parameter estimates of the different methods are statistically similar in some cases; that is, the confidence intervals intersect. This result is consistent with the findings of Van Biesebroeck (2008) who suggest that different production estimation methods lead to statistically similar estimates. Yet from a statistical standpoint, the results reported in Tables 2 may suggest the need to search for more efficient estimators (Su and Ullah, 2011).[15]

Naturally, the construction of measures that depend on point estimates, such as markups, may be significantly affected.

In Table 3 we present the pooled correlation and standard deviation of the productivity measures obtained by the OP/ACF or SU-FE. This table also reports coefficients for the respective lagged productivity dependent variables in simple autoregressive regressions, which are needed to examine the persistence of these measures.

Insert Table 3

Table 3 reveals a high level of correlation and persistence, which is consistent with the findings of previous studies (Foster el al. 2008; Van Biesebroeck, 2008; Syverson 2011)[16]. Notice that the persistence is a function of whether the base-year price level is considered a fixed effect (in SU) or a random effect (as in OP), i.e., is generally larger in the fixed-effects estimation.

Also noteworthy is the substantial dispersion within sectors: standard deviations exceed 20 log points. The productivity measures obtained when the base-year price level is viewed

[15] Naturally, the relavance of differences in point estimates against statistical significant differences is relevant to guide economic decisions (McCloskey and Ziliak, 1996).

[16] Van Biesebroeck (2008) shows that alternative productivity estimation methods generate "*surprisingly similar results*". As Syverson (2011) suggests: "the inherent variation in establishment- or firm-level microdata is typically so large as to swamp any small measurement-induced differences in productivity metrics. Simply put, high-productivity producers will tend to look efficient regardless of the specific way that their productivity is measured."

as a fixed effect exhibit systematically higher dispersion than those obtained when that level is viewed as a random effect.

This finding is consistent with more efficient businesses charging lower prices, which in turn reflects the negative correlation between productivity and price level; when the base-year price level is embodied in the productivity measure, that negative correlation results in productivity that is less dispersed than when we control for the base-year price level.

CONCLUSION

Production functions are important primitive components of many economic models and their estimation play a key role in many empirical analysis. One of the limitations that arise in the estimation of production functions is that typically we observe revenue but not physical output, and we do not have prices at the firm level.

Our paper makes two primary contributions to the empirical literature on the production function. First, we show that deflating revenue by a firm price index constructed from data on the rate of change in prices introduces a fixed effect into the empirical production function. Second, we estimate a production function with two unobservable variables structurally obtained. One is the unobserved productivity that it is assumed to evolve as a first-order Markov process and the other one is the price in the base year which is a fixed effect. We use recently developed semiparametric methods to handle with the estimation.

We compare the results of the estimation of a production function using an unbalanced panel of small and medium-size Spanish manufacturing. Our results show that estimated point coefficients for capital and labor are larger and the materials systematically lower when the base-year price level fixed effect is considered in the empirical specification of the production function. These changes in point estimators leads to a RS closer to 1. We also find that productivity exhibit more dispersion within sectors when the fixed effect is take into account due to the expected negative correlation between productivity and prices (more efficient firms charge lower prices). Finally, despite the differences in point estimates productivity measures obtained with and without taking into account the unobservable fixed effect are highly correlated.

Appendix A: Survey and definition of variables.—

The Encuesta Sobre Estrategias Empresariales (ESSE) 1990–2006 is a panel of firms. The raw data set consists of 4,357 manufacturing firms and a total of 30,827 observations. At the beginning of this survey in 1990, 5% of firms with fewer than 200 workers were sampled randomly by industry and size strata. All such firms were asked to participate, which they did at a rate of about 70%. The initial sample properties have been maintained in subsequent years because exit attrition is balanced by replacing exiting firms with newly created firms that satisfy the initial sampling criteria as in the first year.

For the research reported in this paper, we selected a subsample consisting of small and medium-size firms that are present in the panel data for at least two consecutive years; this subsample includes 15,757 observations for 2,616 firms. Because the OP procedure considers only firms with positive investment, we eliminated from the sample all firms that did not exhibit positive investment. Hence our final sample contained 10,484 observations for 1,616 firms.

The variables are defined as follows.

- *Capital.* Capital at current replacement values K_{it} is computed recursively from an initial estimate and using data on current investments in equipment goods I_{it}. We update the value of the past stock of capital by means of the price index of investment in equipment goods p_{It} as $K_{it} = (1-\delta)(p_{It}/p_{It-1})K_{it-1} + I_{it-1}$, where δ is an industry-specific estimate of the rate of depreciation. Capital in real terms is obtained by deflating capital at current replacement values by the price index of investment in equipment goods.

- *Investment.* The value of current investments in operative capital includes equipment goods but excludes buildings, land, and financial assets. The magnitude is deflated by the price index of investment (the equipment goods component of the index of industry prices computed and published by the Spanish Statistic Institute, INE).

- *Market dynamism.* Firms are asked to assess the current and future situation (slump, stability, or expansion) of up to five separate markets in which they operate. The market dynamism index is computed as a weighted average of the responses.

- *Materials.* Value of intermediate consumption (including raw materials, components, energy, and services) deflated by a firm-specific price index of materials.

- *Output.* Value of produced goods and services, computed as sales plus the variation of inventories and deflated by a firm-specific price index of output.

- *Employment.* Number of full-time plus half of part-time workers as of December 31.

- *Demand shifters*: firms are asked to assess the current and future situation of up to 5 separate markets which they operate: contraction, stability or expansion.

- *Materials price index*: Firm-specific price index for intermediate consumption: firms are asked about the price changes that occurred during the year for raw materials, components, energy, and services. The price index is computed as a Paasche-type index.

- *Output price index*: Firm-specific price index for output. Firms are asked about the price changes they made during the year in up to 5 separate markets in which they operate. The price index is computed as a Paasche-type index.

Appendix B: Testing random versus fixed effects: nonparametric Hausman test.—

The test statistic is based on the sample analog of $J = E\left\{u_{it}E\left(u_{it}|x_{it}, z_{it}\right)f\left(x_{it}, z_{it}\right)\right\}.$

Let $\widehat{m}(z)$ be a consistent estimator of $m(z)$ under the fixed effects assumption and $\widehat{\beta}$ a consistent estimator of β. Then $\widehat{u}_{it} = y_{it} - x_{it}'\widehat{\beta} - \widehat{m}(z_{it})$.

A feasible test statistic is given by

$$\begin{aligned}\widehat{J} &= \frac{1}{R}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\widehat{u}_{it}E_{-it}\left(\widehat{u}_{it}|Z_{it}\right)\widehat{f}_{-it}\left(Z_{it}\right)\\ &= \frac{1}{R(R-1)}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\widehat{u}_{it}\left[\sum_{j=1}^{n}\left(\sum_{\substack{s=1\\ j,s\neq i,t}}^{T_i}\widehat{u}_{js}K_H\left(Z_{it}, Z_{js}\right)\right)\right]\end{aligned}$$

Let

$$K_H\left(Z_{it}, Z\right) = \left[K_H\left(Z_{it}, Z_{11}\right), K_H\left(Z_{it}, Z_{12}\right)..., K_H\left(Z_{it}, Z_{1T_1}\right), K_H\left(Z_{it}, Z_{it}\right)...K_H\left(Z_{it}, Z_{nT_n}\right)\right]_{1\times R}$$

so

$$\sum_{j=1}^{n}\left(\sum_{\substack{s=1\\ j,s\neq i,t}}^{T_i}\widehat{u}_{js}K_H\left(Z_{it}, Z_{js}\right)\right) = K_H\left(Z_{it}, Z\right)\times\left(\widehat{\mathbf{u}}.*e_{it}\right)$$

where

$$\begin{aligned}\widehat{\mathbf{u}} &= \left(\widehat{u}_{11}, ..., \widehat{u}_{nT_n}\right)_{R\times 1}'\\ e_{it} &= \left(1, ...1, 0_{it}, 1...1\right)_{R\times 1}'\end{aligned}$$

and $.*$ is the element by element product, so

$$\left(\widehat{\mathbf{u}}.*e_{it}\right) = \left(\widehat{u}_{11}, ...\widehat{u}_{it-1}, 0_{it}, \widehat{u}_{it+1}..., \widehat{u}_{nT_n}\right)_{R\times 1}'$$

Hence

$$\widehat{J} = \frac{1}{R(R-1)} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \{\widehat{u}_{it} * [K_H(Z_{it}, Z) \times (\widehat{\mathbf{u}}. * e_{it})]\}$$

Wild bootstrap procedure:

The null hypothesis is a random effect model, so the bootstrap is based on the random effects residuals.

Let $\widehat{\mathbf{u}}^{re} = \left(\widehat{u}_{11}^{re}, ..., \widehat{u}_{nT_n}^{re}\right)'_{R\times 1}$ be the random effect residuals, from where we define the bootstrap residuals based on a two point wild bootstrap

$$u_{it}^* = \begin{cases} \left(\left(1-\sqrt{5}\right)/2\right) u_{it}^{re} & p = \left(1+\sqrt{5}\right)/\left(2\sqrt{5}\right) \\ \left(\left(1+\sqrt{5}\right)/2\right) u_{it}^{re} & 1-p \end{cases}$$

We generate the bootstrap sample $\{y_{it}^*, x_{it}, z_{it}\}_{i=1,..,n;t=1,...T_i}$, as

$$y_{it}^* = x_{it}'\widehat{\beta}_{re} + \widehat{m}_{re}(z_{it}) + u_{it}^*$$

where $\widehat{\beta}_{re}, \widehat{m}_{re}(z_{it})$ where estimated in the original sample with a random effect method.

For each bootstrap sample we perform the Hausman test and repeat for B bootstraps. The bootstrap p-value is given as $B^{-1} \sum I\left(\widehat{J}^* > \widehat{J}\right)$.

REFERENCES

Ackerberg, D., K. Caves & G. Frazer (2015). "Identification Properties of Recent Production Function Estimators" *Econometrica,* 83(6), 2411–2451.

Ackerberg, D., C. L. Benkard, S. Berry, & A. Pakes (2007). "Econometric Tools for Analyzing Market Outcomes" in James Heckman and Edward Leamer, eds., Handbook of Econometrics, Vol. 6(1), Amsterdam: North-Holland, 4171–4276.

Aguirregabiria, V. & C. Alonso-Borrego (2014). "Labor contracts and flexibility: evidence from a labour market reform in Spain" *Economic Inquire*, 52, 930–957.

Asker, J., A. Collard-Wexler & J. De Loecker (2014). "Dynamic inputs and resource (mis) allocation", *Journal of Political Economy* 122(5), 1013–1063

Aw, Y., M. Roberts and D. Xu (2011). "R&D Investment,Exporting, and Productivity Dynamics" *American Economic Review,* 101, 1312–1344.

Bierens, H. (2009). "Consistent specification tests" *mimeo.*

De Loecker, J. (2007)., "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity" *NBER* WP 13155.

De Loecker, J. (2011). "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity" *Econometrica,* 79(5), 1407–1451.

De Loecker, J. & F. Warzynski (2012). "Markups and firm level export status" *American Economic Review*, 102(6), 2437–2471.

De Loecker, J., Goldberg, P., Khandelwal, A. & Pavcnik, N. (2016), 'Prices, markups and trade reform', *Econometrica* 84(2), 445–510.

Dolado, J. Ortigueira, S. & R. Stucchi (2016). "Does dual employment protection affect TFP? Evidence from Spanish manufacturing firms" *SERIEs* 7(4), 421–459.

Doraszelski, U. & J. Jaumandreu (2013). "R&D and Productivity: Estimating Endogenous Productivity" *The Review of Economic Studies,* 80, 1338–1383.

Eslava, M., J. Haltiwanger, A. Kugler, & M. Kugler (2004). "The Effects of Structural Re-forms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia" *Journal of Development Economics,* 75 (2), 333–372.

Foster, L., J. Haltiwanger & C. Syverson (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review,* 98(1), 394–425.

Foster, L., Haltiwanger, J. & Syverson, C. (2016), 'The slow growth of new plants: Learning about demand?', Economica 83(329), 91–129.

Gao, J. (2014). "Identification, Estimation and Specification in Semi-Linear Models". In Handbook on Non- and Semi-Parametric Econometrics and Statistics (edited by Jeffrey Racine, Liangjun Su & Aman Ullah). Oxford University Press.

Henderson, D., R. Carroll & Q. Li (2008). "Nonparametric Estimation and Testing of Fixed Effects Panel Data Models" *Journal of Econometrics,* 144, 257–275

Huang G. & Hu Y. (2011). "Estimating production functions with robustness against errors in the proxy variables" CeMMAP working papers CWP35/1

Klette, T. J. & Griliches, Z. (1996). "The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and Endogenous" J*ournal of Applied Econometrics,* 11, 343–361.

Levinsohn, J. & A. Petrin (2003). "Estimating Production Functions Using Inputs to Control for Unobservables" *Review of Economic Studies,* 70(2), 317–342.

Li, Q. & T. Stengos (1996). "Semiparametric Estimation of Partially Linear Panel Data Models" *Journal of Econometrics,* 71, 389–397.

Li, Z. Q.Li & Sun Y. (2014). "A consistent nonparametric test ofparametric regression functional form in fixed effects panel data models" *Journal of Econometrics,* 178, 167–179.

Mairesse, J. & J. Jaumandreu (2005). "Panel-data Estimates of the Production Function and the Revenue Function" *Scandinavian Journal of Economics,* 107(4), 651–672.

McCloskey, D. & S. T. Ziliak (1996). "The Standard Error of Regression" *Journal of Economic Literature,* 34(1), 97–114.

Olley, S.& Pakes, A.(1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry" *Econometrica,* 64, 1263–97.

Pakes, A. (1994). "The estimation of dynamic structural models" J. Laffont and C. Sims eds. Advances in Econonometrics, Ch. 5.

Racine, J., & Li, Q. (2004). "Nonparametric estimation of regression functions with both categorical and continuous data" *Journal of Econometrics,* 119, 99–130

Roberts, M. & D. Supina (2000). "Output Price and Markup Dispersion in Producer Micro Data: The Roles of Producer Heterogeneity and Noise" in Advances in Applied Microeconomics, Vol 9, Industrial Organization, Michael Baye (ed.), JAI Press.

Su, L. & Lu, X. (2013). "Nonparametric dynamic panel data models: kernel estimation and specificaton testing" *Journal of Econometrics,* 176, 112–133.

Su, L. & A. Ullah (2006). "Profile likelihood estimation of partially linear panel data models with fixed effects" *Economics Letters,* 92(1), 75–81.

Su, L. & A. Ullah (2007). "More efficient estimation of nonparametric panel data models with random effects." *Economics Letters,* 92(1), 75–81.

Su, L. & A. Ullah (2011). "Nonparametric and Semiparametric Panel Econometric Models: Specification, Estimation and Testing" in A. Ullah and D. E. A. Giles (eds), Handbook of Empirical Economics and Finance, pp. 455–497. Taylor & Francis Group, New York.

Syverson, Ch. (2011). "What Determines Productivity?" *Journal of Economic Literature,* 49(2), 326–65.

Van Biesebroeck, J. (2008). "The Sensitivity of Productivity Estimates: Revisiting Three Important Debates" *Journal of Business and Economic Statistics,* 26(3), 311–328.

Table 1. Nonparametric Hausman test under two specifications of the OP/ACF proxy function.

	A	
	I	II
1. Food, drink & tobacco	0.213 (0.000)	0.101 (0.000)
2. Textile, leather, & shoes	0.177 (0.000)	0.158 (0.000)
3. Timber & furniture	0.539 (0.000)	0.207 (0.000)
4. Paper & printing products	0.232 (0.000)	0.142 (0.000)
5. Chemical products	0.329 (0.000)	0.099 (0.000)
6. Non-metallic minerals	0.176 (0.024)	0.088 (0.000)
7. Metals & metal products	0.248 (0.000)	0.121 (0.000)
8. Agric. & ind. machinery	0.454 (0.000)	0.118 (0.000)
9. Office, comp. & electr.	0.193 (0.016)	0.157 (0.000)
10. Transport equipment	0.135 (0.020)	0.034 (0.088)

Notes: Columns I report original OP/ACF specification, where the proxy function includes capital, labor, demand shifters, and time dummies; in columns II the proxy function additionally includes the price of input and wages. We use the mixed continuous-discrete kernel of Racine and Li (2004) with bandwidths for the continuous kernel $h = a_0 R^{-1/4+q}$where $a_0$denotes a constant which was set equal to the mean of the sample standard deviations, $q$the dimension of the continuous variables and $R$the sample size; the mixed discrete kernel is $\lambda = b_0 R^{-2/4+q}$where $b_0$denotes a constant which was set to one and where the continuous variables where were standarized so as to use the same bandwidth. The same bandwidths were used for the Hausman test; p-value in parentheses are based on 500 bootstrap replications.

Table 2: Olley and Pakes method and Su and Ullah random-effects estimator

	OP method[a]				Su-Ullah FE[b]			
	Capital	Labor	Materials	RS	Capital	Labor	Materials	RS
1. Food, drink & tobacco	0.127 (0.026)	0.144 (0.043)	0.733 (0.022)	1.00	0.129 (0.045)	0.178 (0.074)	0.689 (0.019)	0.99
2. Textile, leather, & shoes	0.040 (0.024)	0.351 (0.044)	0.595 (0.028)	0.98	0.111 (0.035)	0.269 (0.064)	0.591 (0.016)	0.97
3. Timber & furniture	0.054 (0.035)	0.062 (0.079)	0.700 (0.044)	0.81	0.076 (0.031)	0.365 (0.075)	0.506 (0.028)	0.95
4. Paper & printing products	0.077 (0.033)	0.174 (0.076)	0.563 (0.045)	0.81	0.093 (0.045)	0.429 (0.105)	0.514 (0.034)	1.04
5. Chemical products	0.078 (0.039)	0.376 (0.056)	0.662 (0.038)	1.11	0.032 (0.034)	0.353 (0.063)	0.617 (0.026)	1.00
6. Non-metallic minerals	0.092 (0.039)	0.366 (0.056)	0.574 (0.044)	1.03	0.096 (0.048)	0.343 (0.059)	0.573 (0.030)	1.01
7. Metals & metal products	0.125 (0.027)	0.192 (0.055)	0.641 (0.023)	0.97	0.087 (0.032)	0.310 (0.051)	0.566 (0.019)	0.96
8. Agric. & ind. machinery	0.037 (0.036)	0.264 (0.066)	0.596 (0.043)	0.90	0.048 (0.039)	0.302 (0.067)	0.549 (0.028)	0.90
9. Office, comp. & electr.	0.031 (0.045)	0.448 (0.091)	0.632 (0.049)	1.10	0.056 (0.051)	0.559 (0.097)	0.535 (0.031)	1.15
10. Transport equipment	0.160 (0.059)	0.287 (0.099)	0.709 (0.032)	1.16	0.065 (0.064)	0.267 (0.115)	0.714 (0.036)	1.05

Notes: The proxy function includes capital, labor, demand shifters, input prices and time dummies. RS: return to scale; Standard errors (reported in parentheses) based on 500 bootstrap replications.[a]Original OP method. [b]Su and Ullah (2006) fixed effect profile least square applied to the OP method in order to correct for unobservability of the base-year price level; here materials is the only totally flexible nondynamic input.

Table 3 Pooled correlation and standard deviation of different productivity measures

Sector		SU FE	St. Dev.	Persist rates	Sector	SU FE	St. Dev.	Persist rates
1	SU FE		0.22	0.84 (0.03)	6		0.23	0.78 (0.03)
	OP	0.98	0.21	0.83 (0.06)		0.95	0.22	0.75 (0.05)
2	SU FE		0.23	0.83 (0.05)	7		0.24	0.84 (0.04)
	OP	0.93	0.23	0.84 (0.03)		0.92	0.22	0.76 (0.05)
3	SU FE		0.25	0.86 (0.04)	8		0.25	0.86 (0.05)
	OP	0.84	0.21	0.75 (0.06)		0.89	0.22	0.78 (0.06)
4	SU FE		0.27	0.93 (0.02)	9		0.27	0.85 (0.03)
	OP	0.91	0.25	0.90 (0.03)		0.95	0.25	0.82 (0.03)
5	SU FE		0.28	0.94 (0.02)	10		0.23	0.81 (0.06)
	OP	0.93	0.26	0.92 (0.02)		0.79	0.34	0.90 (0.02)

Notes: Values report correlations among the productivity measures obtained by different estimation methods for the pooled sample after removing the year effect. Standard errors are reported in parenthesis: OP: Olley and Pakes (1996); SU FE: Su and Ullah (2006) fixed effect; SU RE: Su and Ullah (2007) random effect; St. Dev.: standard deviation; Persist rates: coefficient of the lagged dependent variable in a weighted autoregressive regression.