Estimating Production Functions with Heterogeneous Firms

Amit Gandhi, Salvador Navarro and David Rivers*
University of Wisconsin-Madison

February 15, 2008

Abstract

We present a new approach to the estimation of production functions that allows for richer patterns of firm heterogeneity than can be accommodated under the proxy variable methods of Olley and Pakes (1996) and Levinsohn and Petrin (2003). In particular, we show that the economics of the firms static input choice contains the necessary identifying information to control for the endogeneity problem in the production function. From an econometric point of view, our estimation proceeds in a single GMM step, and thus standard asymptotic standard errors are available. We consider the identification and estimation of models with heterogeneity in both input and output prices, as well as heterogeneity in factor specific productivity. Our empirical results show that we control for more of the endogeneity problem than the proxy variable approach, resulting in estimates of labor productivity nearly half as small.

1 Introduction

As first pointed out by Marschak and Andrews (1944), using inputs and outputs to estimate production functions gives rise to an endogeneity problem. The endogeneity problem is caused by the presence of productive factors that are unobservable to the econometrician but that are "transmitted" to the firm's optimal choice of inputs. These unobservable factors are traditionally captured by a scalar productivity index that varies across firms and potentially evolves over time. The two traditional and oldest methods of controlling for the endogeneity problem in the production are fixed effect estimation using panel data and instrumental variables. However these solutions have proven unsatisfactory on both theoretical and empirical grounds (for a review, see Ackerberg et al. (2007)). Instead, the modern literature on estimating production functions initiated by Olley and Pakes (1996) (OP for short) uses restrictions from economic theory to identify production function parameters. In particular, under certain condition's on the firm's profit maximization problem, it is possible to use an observed input decision of the firm, such as investment (Olley and Pakes, 1996)or intermediate materials (Levinsohn and Petrin, 2003), and invert this inputs in order to recover a firm's level of productivity, thereby controlling for it in the estimation. We shall refer to this method of inverting inputs as the "proxy variable approach" since the observed level of the input essentially proxies for the unobserved level of productivity.

*We thank John Kennan, Rasmus Lentz, and Amil Petrin for their helpful comments. We also thank Jim Levinsohn and Jagadeesh Sivadasan for lending us their data and their time. Funding for this research was provided by the Graduate School at the University of Wisconsin-Madison.
Address: University of Wisconsin-Madison, 1180 Observatory Drive. Madison, WI. 53706.
E-mail: agandhi@ssc.wisc.edu.

Along with a number of regularity conditions on the firm's problem, the theory underlying the proxy variable approach depends critically upon the assumption that the firm's current level of total factor productivity is the *only* dimension of unobserved heterogeneity across firms. In particular, unobserved differences in input/output prices and technology differences are assumed away. This is a strong assumption, and if violated would mean that the econometrician does not fully control for productivity when inverting the proxy, leaving the endogeneity problem to remain. As Ackerberg et al. (2007) themselves suggest, productivity heterogeneity is possibly reflective of heterogeneity in the quality of labor, and thus wages are endogenous. However in so far as wages are not observed, controlling for their endogeneity is an important issue for production function estimation that has yet to be addressed. As Fox and Smeets (2007) have recently shown, total factor productivity as recovered by the proxy method appears to explain only a small portion of the measured productivity differences across firms in their data, which suggests that proxy method is not recovering productivity estimates that capture . Moreover, the theory and empirics of the firm size distribution (see e.g., Lucas (1978)) along with the observation that larger firms pay higher wages (see e.g., Mortensen and Pissarides (1994)) suggest that wages and the marginal product of labor are highly heterogeneous, even within a homogoenous good industry. While such patterns of heterogeneity naturally arise through a variety of sorting/matching mechanism for clearing factor markets (e.g., the matching of productive managers with productive workers), they have remained outside the scope of the empirical literature to date. In addition, even though manufacturing firm's may behave competitively in their output markets, it is quite likely that their products are partially differentiated, at the very least by firm location when the population under consideration is all firms in an industry at the country level (as is the case with the now popular Chilean data used in Levinsohn and Petrin (2003)). Such unobserved heterogeneity in output prices is also excluded from the proxy variable approach.

The purpose of this paper is to present a new a new method for estimating production functions that allows for richer patterns of firm heterogeneity than the proxy variable approach, and as such has a number of potential applications. In addition to presenting the basics of our approach, we present results from the application of our method to the same Chilean manufacturing data used in Levinsohn and Petrin (2003) (LP for short). Compared to the LP estimates, our approach yields significantly smaller estimates of labor productivity (on average half as small), and significantly smaller estimates of returns to scale (from increasing returns to decreasing). These differences are what we expect when the proxy variable (intermediate materials in the LP case) fails to fully control for the endogeneity problem due to the multiple dimensions of unobserved heterogeneity that are likely playing a role in the data.

The key to our method lies in using the economics of the firm's static input decision as a source of identifying information. Since the classic endogeneity problem is caused by a "transmission" of the the firm's productivity to its static input choices, the first order condition of the static input choice provides an exact source to control for unobserved productivity. In particular, we use the first order condition as a second structural equation that allows us to separate out the endogenous part of the error term in the production function. Appealing to the first order conditions of the firm's static input decision stands in contrast to the modern literature's focus on the first order conditions of the firm's dynamic input decision. Nevertheless, we form moment conditions much in the manner of LP and OP, thus our framework is consistent with the dynamics of their underlying economic model. Since we avoid altogether proxy variables, we are able to allow for richer patterns of firm heterogeneity, such as differences in prices and factor specific productivity, which explains why our estimates diverge from theirs in the particular direction we obtain.

We further examine the sources heterogeneity that are playing a role in the data. In particular we consider whether firms in an industry exhibit heterogeneity in their labor productivity. We show that the distribution of labor productivity can be nonparametrically identified within our framework, and our estimate of the distribution across industries reveals considerable heterogeneity in labor productivity.

The plan of the paper is as follows. We first lay the groundwork by specifying the general economic setting in which both our method and the proxy variable methods take place. The proxy variable approach and its specialized assumptions are then reviewed. The main body of the paper focuses on how to use the structure of the firm's labor input problem to control for the endogeneity problem and allow for multiple dimensions of heterogeneity. We then conclude with the application of our framework to the Chilean manufacturing data and directions for future research.

2 The Standard Setup

Research on the estimation of production functions has largely assumed a Cobb-Douglas technology. We will follow in this tradition in order to explain our approach. However the method of using the first order condition of the static input choice as an additional source of identifying information can be carried out under a variety of assumptions about the functional form of technology. In the last section, we sketch our strategy for studying other production functions, such as CES. What follows below is a presentation of just the most essential elements from the model of Olley and Pakes (1996) needed to explain the proxy variable approach along with our method.

We consider the problem of a firm j that produces an output Y using capital K and labor L. The firm's technology at time t is described by the following Cobb-Douglas production function:

$$y_{jt} = \alpha l_{jt} + \beta_j k_{jt} + a_{jt} \tag{1}$$

where small letters denote logs. Thus the productivity of labor α_j, the productivity of capital β_j, and the productivity residual a_{jt} will generally vary across firms. Furthermore, a firm's productivity residual a_{jt} can evolve over time. More generally, we can allow for several inputs, but for the sake of exposition will study the traditional two input case.

While (1) represents a fairly general case of a population of Cobb-Douglas firms, the focus of the empirical literature has been on the case where $\forall j$, $\beta_j = \beta$ and $\alpha_j = \alpha$. That is, the only technology difference across firms in operation at a given time lies in the productivity residual. We will focus for now on this special case and return later in the paper to the more general formulation inherent in (1).

As Griliches and Mairesse (1998) explain, the the productivity residual a_{jt} can be decomposed into three components, i.e.,

$$a_{jt} = \omega_{jt} + \epsilon_{jt} + \eta_{jt}. \tag{2}$$

In (2), ω_{jt} is an anticipated productivity shock that firm j observes before it makes its period t input decisions, while ϵ_{jt} is an unanticipated productivity shock that firm j observes only after it makes its period t input decisions. On the other hand, η_{jt} equals measurement error (brought about, for example, by deflating revenue in order to measure output), which does not enter into the firm's problem but rather enters only into the econometric specification. We refer to the component ω_{jt} as the firm's total factor productivity as it captures a systematic technology difference across firms. Moreover from the econometric point of view, total factor productivity represents the endogenous part of error term since current and past realizations of a firm's productivity, i.e., the set $\{\omega_{j\tau}\}_{\tau=1}^{t}$, can be transmitted to the firm's inputs choices k_{jt} and l_{jt} for the current period, thus invalidating the most obvious estimation strategy, i.e., pooled OLS regression on (1). The exact mechanism through which productivity is transmitted to the inputs depends on the timing of the input decisions, which we now discuss.

A firm's productivity ω_{jt} evolves according to a Markov process. That is, the transition probability of productivity can be written as

$$\Pr\left(\omega_{jt}|\{\omega_{j\tau}\}_{\tau=1}^{t-1}, J_{t-1}\right) = \Pr\left(\omega_{jt}|\{\omega_{j\tau}\}_{\tau=1}^{t-1}\right)$$

where J_{t-1}represents all other information available to the firm at time $t-1$. It is traditionally assumed that this process is first order, an assumption we maintain moving forward. More generally, our method allows for R&D expenditure this period to impact productivity next period, but we ignore this consideration for the time being in order to explain the most basic model.

Firm j faces output prices P_{jt} and wages W_{jt} in period t and behaves competitively, i.e., chooses inputs to maximizes profits taking the input and output prices it faces as given. The inputs on the other hand are chosen over different time horizons. Capital on the other hand is accumulated dynamically. That is, K_{jt} is determined by the sequence of prior investment decisions $\{I_{j\tau}\}_{\tau=1}^{t}$ via the accumulation process $K_{jt} = \mathbf{K}(K_{jt-1}, I_{jt-1})$. The key point is that the current period's capital stock is determined through last period's investment decision, and is thus independent of the innovation $\xi_{jt} = \omega_{jt} - E[\omega_{jt}|\omega_{jt-1}]$ in the current period's productivity. Labor on the other hand is a static input and is thus chosen each period conditional on (k_{jt}, ω_{jt}) as state variables. Labor is chosen to maximize each period's profit, while firm j sequence of investments $\{I_{jt}\}$ is chosen to maximize the expected discounted sum of profits.

3 The Proxy Variable Approach

In order to solve the endogeneity problem, the literature initiated Olley and Pakes (1996) and further developed by Levinsohn and Petrin (2003) assumes the existence of a proxy input d_{jt} (demand for investment ($d_{jt} = I_{jt}$) in Olley and Pakes (1996), demand for materials ($d_{jt} = M_{jt}$) or electricity ($d_{j,t} = C_{jt}$) in Levinsohn and Petrin (2003)), such that

$$d_{jt} = \delta_t\left(k_{jt}, \omega_{jt}\right) \tag{3}$$

The demand function δ_t is assumed to be strictly monotone in productivity ω_{jt} when demand is positive, i.e., when $d_{jt} > 0$. The proxy demand function δ_t is indexed by time to capture the effect of all other time specific variables on the proxy decision. However an important assumption is that δ_t is not index by j, so any firm level heterogeneity firm outside of capital stock k_{jt} and total factor productivity ω_{jt} is assumed away. This, for example, restricts all firms to face homogeneous input and output prices in all periods, i.e., $\forall j, t$ it is assumed that $P_{jt} = P_t$, and $W_{jt} = W_t$. Furthermore, the assumption does not allow for there to be an idiosyncratic error term in the proxy demand function, which could reflect such things as firm level differences in interest rate expectations and tax treatment in the case of investment demand, or stochastic supply of electricity in the case of intermediate input demand, or more generally measurement error in the proxy variable itself (due for example to applying to a deflator to measure real proxy demand).

We leave asides for the moment our reservations about the assumptions underlying proxy demand, and show how the proxy variable and how is used to control for unobserved productivity. First, for all observations (j, t) for which $d_{jt} > 0$, we can invert (3) to recover ω_{jt}, which we then replace in (**??**) to yield

$$\begin{aligned} y_{jt} &= \alpha l_{jt} + \beta k_{jt} + \delta_t^{-1}\left(k_{jt}, d_{jt}\right) + \varepsilon_{jt} \\ &= \alpha l_{jt} + \Phi_t\left(k_{jt}, d_{jt}\right) + \varepsilon_{jt} \end{aligned} \tag{4}$$

where $\varepsilon_{jt} = \epsilon_{jt} + \eta_{jt}$ and $\Phi_{dt}(k_{jt}, d_{jt}) = \alpha k_{jt} + \omega_{jt}$. Thus one can consistently estimate the labor coefficient α and the value $\Phi_{jt} = \Phi(k_{jt}, d_{jt})$ in (4) by semiparametric techniques (e.g., Robinson (1988)). This constitutes the first stage of the estimation. For the second stage, given the first stage estimates $(\hat{\alpha}, \hat{\Phi})$, and any possible value of the capital coefficient β, we can form the residual

$$\theta_{jt}(\beta) = y - \hat{\alpha} l_{jt} - \beta k_{jt} = \omega_{jt} + \varepsilon_{jt} \tag{5}$$

for all observations (j, t). We can also form the residual

$$\omega_{jt}(\beta) = \hat{\Phi}_{jt} - \beta k_{jt} \tag{6}$$

for all observations (j, t) such that proxy demand is positive, i.e., $d_{jt} > 0$. Then the nonparametric regression of θ_{jt} on ω_{jt-1} yields

$$E[\omega_{jt} + \varepsilon_{jt} \,|\, \omega_{jt-1}, d_{jt-1} > 0] = E[\omega_{jt} \,|\, \omega_{jt-1}],$$

where the last equality follows from the fact that ε_{jt} is uncorrelated with ω_{jt} and the first order Markov assumption. Finally the parameter β is estimated using the moment condition that $\xi_{jt} = \omega_{jt} - E[\omega_{jt} \,|\, \omega_{jt-1}]$ is mean independent of k_{jt}.

3.1 The Colinearity Problem

As pointed out by Bond and Söderbom (2005) and Ackerberg et al. (2006), while (4) appears to identify the labor coefficient, the assumptions made thus far about the data generating process suggest that the labor coefficient α is in fact not identified from the first stage regression in the proxy variable approach. We now present a proof of their non-identification argument, which the authors do not formally establish.

The key to understanding the nonidentification claim starts by recalling that labor is assumed to be a fully flexible factor, and as such is chosen optimally by firm j at time t given the state variables (k_{jt}, ω_{jt}). Recalling also the scalar unobservable assumption, we thus have that labor demand follows $l_{jt} = f_t(k_{jt}, \omega_{jt})$. However the proxy demand also has the form $d_{jt} = \delta_{dt}(k_{jt}, \omega_{jt})$, which we invert to obtain $\omega_{jt} = h_t(k_{jt}, d_{jt})$. From this it follows that

$$l_{jt} = f_t(k_{jt}, h_t(k_{jt}, d_{jt})) = F_t(k_{jt}, d_{jt}).$$

Going back to the first stage regression in the proxy variable approach, if we wish to be fully nonparametric about Φ_t, then it follows that we cannot simultaneously identify the labor coefficient α. To see this, consider $\alpha' \neq \alpha$. Then

$$\begin{aligned} y_{it} &= \alpha l_{jt} + \Phi_t(k_{jt}, d_{jt}) + \varepsilon_{jt} \\ &= (\alpha' + (\alpha - \alpha')) l_{jt} + \Phi_t(k_{jt}, d_{jt}) + \varepsilon_{jt} \\ &= \alpha' l_{jt} + (\alpha - \alpha') F_t(k_{jt}, d_{jt}) + \Phi_t(k_{jt}, d_{jt}) + \varepsilon_{jt} \\ &= \alpha' l_{jt} + \Phi'_t(k_{jt}, d_{jt}) + \varepsilon_{jt}. \end{aligned}$$

Thus we cannot be both fully nonparametric about Φ_t and still identify α.

Later in the paper, we offer a new approach to estimating the production function parameters . In particular we show that when the full economics of the static input decision are used, the labor coefficient is identified simultaneously with the other parameters. Thus we do not need to invoke timing restrictions or adjustment cost stories to break the above colinearity as proposed by Ackerberg et al. (2006) and Bond and Söderbom (2005). For now we will assume the labor coefficient β is known, and proceed to examine potential difficulties that arise with the second stage of the proxy variable approach.

3.1.1 Selection on the Proxy

Recall once again that the proxy variable approach requires that $\delta_t(k_{jt}, \omega_{jt})$ be monotonic in ω_{jt} conditional on the event that the proxy demand $d_{jt} > 0$. For the case of investment, the proxy demand is a lumpy and equals zero for sufficiently small values of productivity ω_{jt}, in which case it cannot inverted out. In the case of manufacturing data from Chile, Levinsohn and Petrin (2003) report that such zeroes occur for nearly 50% of their observations when using investment as a proxy variable. Although dropping such observations from the data in order to estimate the model surely results in a serious efficiency loss, so long as the moments are constructed in the fashion explained in Section 3, there is no selection bias introduced. However the economic relationships in the model give rise to a number of alternative and seemingly more efficient ways of constructing moments, and these alternative constructions could very well give rise to a selection problem.

For example Ackerberg, Caves and Frazer (2006); Ackerberg, Benkard, Berry and Pakes (2007) advocate an "equivalent" method of constructing the moments. To quote Ackerberg, Benkard, Berry, and Pakes (p. 53, adjusting their notation to square with ours)

> An equivalent way to construct a moment condition ... is as follows ... Given β, construct $\hat{\omega}_{jt} = \hat{\Phi}_{jt} - \alpha k_{jt}$. Nonparametrically regress $\hat{\omega}_{jt}$ on $\hat{\omega}_{jt-1}$ to construct the estimated residual $\hat{\xi}_{jt}$. Construct a moment condition interacting $\hat{\xi}_{jt}$ with k_{jt}.

However this manner of constructing the moment does in fact suffer from selection bias. To see this, recall (6) and notice that we can only construct the residual $\hat{\omega}_{jt}$ for those observations (j, t) such that $d_{j,t} > 0$. However

$$E[\omega_{jt} \,|\, \omega_{jt-1}, d_{jt} > 0, d_{jt-1} > 0] = E[\omega_{jt} \,|\, \omega_{jt-1}, d_{jt} > 0] \neq E[\omega_{jt} \,|\, \omega_{jt-1}].$$

Thus this procedure using the investment proxy will not lead to the right moment condition.

Another important difficulty can arise from the fact that $\hat{\omega}_{jt}$ can only be constructed for those observations for which $d_{jt} > 0$. For example Olley and Pakes (1996) wish to use their estimated model to study the distribution of productivity among active firms at any given time. Thus they do not wish to select on those firms for which $d_{jt} > 0$. As a result they can only form the "noisy" productivity residual using (5), i.e., $\widehat{\omega_{jt} + \varepsilon_{jt}}$. However the distribution of ω_{jt} is not the same as $\omega_{jt} + \varepsilon_{jt}$, and unless distributional restrictions are placed on ε_{jt}, it is not possible to know apriori the impact of this bias other than that the degree of heterogeneity will be overstated As Fox and Smeets (2007) discuss, manufacturing data show considerably more heterogeneity in their underlying technology (as measured by the residual from the standard pooled OLS estimator of 1) than can be explained by the productivity residuals ω_{jt} recovered from the Olley and Pakes procedure, suggesting that much of the variance in $\omega_{jt} + \varepsilon_{jt}$ is coming from ε_{jt}. Their findings also suggest that the proxy variable does not sufficiently recover underlying heterogeneity in productivity. This is naturally a result of the scalar unobservability assumption, which recall if violated implies that the proxy variable does not in fact control for unobserved productivity, leading to inconsistent production function estimates.

4 A "New" Approach to Estimating Production Functions

A more serious misgiving we have is the idea that, by inverting either investment, materials, electricity, etc. under the proxy variable approach, we are recovering the same ω_{jt}. This follows from the scalar unobservability assumption. That is, in the proxy demand function $\delta_t(k_{jt}, \omega_{jt})$, ω_{jt} is the only firm level unobserved factor that generates differences in proxy demand behavior across firms. In particular, the proxy variable machinery does not allow firms to face heterogeneity in input and output prices and does not allow heterogeneity in factor specific productivity (versus total factor productivity) differences across firms. Moreover the proxy variable approach, since it requires a nonlinear inversion, is sensitive to measurement error in the proxy variable, which will typically be introduced when we apply deflator to measuring real proxy demand. All of these factors would cause the proxy approach to fail to recover productivity exactly, and hence control for the endogeneity problem in the production function.

In this section, we develop an alternative to the proxy variable approach that solves the endogeneity problem under the standard setup of Section (2), but allows for the above mentioned elements of heterogeneity to be included in the model. We refer to our method as a "new" approach, because the basic idea behind it is actually quite in the literature on labor demand. In particular, the point of departure of our approach is that we use the first order condition for the static input decision in each period, i.e., the labor input decision, as our second structural equation instead of the proxy demand equation. Since the endogeneity problem is generated by the transmission of ω_{jt} to the l_{jt}, the first order condition for labor provides an exact source to control for the endogeneity problem that is robust to multiple dimensions of heterogeneity. The idea that the firm's first order condition for profit maximization has identifying power for the marginal product of labor has been much discussed in labor demand literature (for a review, see for example Hamermesh (1993)). The "newness" of our idea comes from combining the first order condition with the production function itself, and recognizing that we can use both equations *jointly* to invert out productivity. Thus the first order condition for labor plays a role analogous to the proxy demand equation. However as we show below, it contains much more information such that we are able to control for heterogeneity unobserved input and output prices.

Another key difference is that we estimate this second equation simultaneously with the production function itself, which leads to an straightforward one step estimator.

4.1 The Labor Input Problem

Recall the empirical specification (1) and the decomposition of the error term (2), and consider the usual constrained case where each firm has the same labor and capital coefficients, i.e., $\beta_j = \beta$ and $\alpha_j = \alpha$ for all firms j. These together imply a production function

$$Q_{jt} = A_{jt} U_{jt} L_{jt}^{\alpha} K_{jt}^{\beta} \tag{7}$$

where $\log(A_{jt}) = \omega_{jt}$ and $\log(U_{jt}) = \epsilon_{jt}$, where the expectation is taken in the cross section over firms, i.e, U_{jt} is drawn from the distribution G_t. Thus A_{jt} represents the firm's total factor productivity or "anticipated" productivity shock, whereas U_{jt} is the firm's unexpected productivity shock or expectational error in anticipating productivity, and the terms ω_{jt} and ϵ_{jt} exactly match their meaning in the decomposition of the econometric error term from equation (2). For the purposes of presentation. Thus we have merely backed out the underlying production function implied by the standard empirical specification.

For simplicity, let $G_t = G$ for all t (more generally we can allow U_{jt} to be drawn from a time varying G_t but this generalization complicates the exposition). Notice then that the term $E[log(U_{jt})]$ gets subsumed in the constant term β_0 in (1) along with the mean of measurement error η_{jt}. Recall $U_{jt} > 0$ is a productivity shock that the firm can only observe *after* it has made its period t input decisions. Thus when planning for these period t input decisions, the firm faces a *stochastic* production function in (7). We assume the unanticipated productivity shock U_{jt} is independent of the productive inputs (A_{jt}, L_{jt}, K_{jt}), i.e., it is an exogenous shock. Furthermore we assume and that the firm has rational expectations and hence knows the distribution G from which the shock is drawn. This implies that WLOG, $E[U_{jt}] = 1$, since any systematic expectation for the unexpected shock other than one simply gets subsumed in TFP component A_{jt} from the point of view of the firm. Thus the firm can form the conditional expectation

$$Q^*_{jt} = E[Y_{jt} \,|\, A_{jt}, L_{jt}, K_{jt}] = A_{jt} L^{\alpha}_{jt} K^{\beta}_{jt}.$$

Finally observe that by definition that the observable output Y_{jt} is related to the firm's expected output Q^*_{jt} by

$$\log Y_{jt} = \log Q^*_{jt} + \epsilon_{jt} + \eta_{jt}. \tag{8}$$

We now consider the firm's labor input problem. Given potentially heterogeneous output prices P_{jt} and wages W_{jt}, and assuming risk neutrality on the part of the firm, the first order condition for labor becomes.

$$\alpha \frac{P_{jt} Q^*_{jt}}{L_{jt}} = W_{jt}. \tag{9}$$

Now taking logs of both sides, adding $\epsilon_{jt} + \eta_{jt}$ to both sides, rearranging terms and recalling the definition $\varepsilon_{jt} = \epsilon_{jt} + \eta_{jt}$, we get

$$\ln \left(\frac{P_{jt} Y_{jt}}{W_{jt} L_{jt}} \right) = -\log(\alpha) + \varepsilon_{jt}. \tag{10}$$

Equation (10) provides us the essential additional information to solve the endogeneity problem inherent in the production function (1). To see this, combine (1) and (10) to yield the system

$$\begin{aligned} \ln \left(\frac{P_{jt} Y_{jt}}{W_{jt} L_{jt}} \right) &= -\log(\alpha) + \varepsilon_{jt} \qquad (11) \\ y_{jt} &= \alpha l_{jt} + \beta k_{jt} + \omega_{jt} + \varepsilon_{jt}. \end{aligned}$$

Letting $s_{jt} = \ln \left(\frac{P_{jt} Y_{jt}}{W_{jt} L_{jt}} \right)$ and $x_{jt} = (l_{jt}, k_{jt})$, we can express the above system more generally as

$$\begin{pmatrix} s_{jt} \\ y_{jt} \end{pmatrix} = \Upsilon(x_{jt}, \omega_{jt}, \varepsilon_{jt}). \tag{12}$$

For production functions more general than Cobb-Douglas, the system of combining the first order condition for labor with the production function will yield the form (12). Moreover, for a variety of production functions, for any realization of the data $D_{jt} = (s_{jt}, y_{jt}, l_{jt}, k_{jt})$ and value of the parameter vector (β, α), we can uniquely solve the two equation system (12) for the two econometric unobservables $(\omega_{jt}, \varepsilon_{jt})$. This especially clear in the case of the Cobb-Douglas case in (11), which forms a simple triangular system in $(\omega_{jt}, \varepsilon_{jt})$. The appendix to this paper present the corresponding development for a CES production function.

To estimate the parameter vector, we simply introduce the variety of possible moment conditions offered by the underlying economic model, which are similar to those used by OP and LP. For example, under the identifying assumption that[1] $E[\log U_{jt}] = E[\epsilon_{jt}] = 0$, and the assumption of mean zero measurement error $E[\eta_{jt}] = 0$, we have the two natural moments, $E[\xi_{jt} k_{jt}] = 0$ for $\xi_{jt} = \omega_{jt} - E[\omega_{jt} \,|\, \omega_{jt-1}]$, and $E[\varepsilon_{jt}] = 0$, which we can use to estimate the parameter vector (α, β). Moreover we have the natural overidentifying moments $E[\varepsilon_{jt} l_{jt}] = 0$ and $E[\varepsilon_{jt} k_{jt}] = 0$. All of these moments are available without appealing to lagged values for instruments.

[1]The identifying assumption $E[\epsilon_{jt}] = 0$is satisfied, for example, if there is no unanticipated productivity shock, i.e, $U_{jt} = 1$, or if U_{jt} is log-normally distributed. We currently have a number of approaches for relaxing this assumption (that include allowing firms to have arbitrary risk preferences), but present our method under this assumption for clarity.

Several points are to be noted concerning our approach. First, the role of introducing the first order condition for labor, as the system (12) makes clear, is to separate out the error terms ω_{jt} and ε_{jt}, which cannot be separated out from the production function alone. The intuition for why this separation is possible in a general production function setting is straightforward. The firm observes ω_{jt} and thus it gets absorbed in the first order condition, which is nothing other than usual endogeneity problem. However the expectational error/unanticipated shock ε_{jt} does not get absorbed by the firm but rather gets introduced when we move from the first order condition from the point of view of the firm to the first order condition from the point of view of the econometrician, i.e., the transition from (9) to (10). Thus the two error terms enter asymmetrically into the second equation, which is the key to the ability to invert the system (11).

Another key point is that we allow for input and output price heterogeneity. However it is not required required that we be able to observe these prices, because they enter the second equation through the term s_{jt}, and hence that we only need to observe the firm's expenditure on labor $W_{jt}L_{jt}$ and its revenue $P_{jt}Y_{jt}$, both variables being commonly available. But these are standard observables in the production setting. In fact, the measure of output Y_{jt} is typically derived by deflating revenue $P_{jt}Y_{jt}$, where revenue is what the firm actually reports in census data. The firm's expenditure on inputs is also typically reported. Thus $W_{jt}L_{jt}$ and $P_{jt}Y_{jt}$ are primitives in the data that do not suffer from measurement error generated by applying deflators (such as is required when trying to measure real intermediate input demand or real investment when applying the proxy, which require deflating expenditures on investment and intermediate inputs).

The above empirical strategy has been explained with reference to a two input production function in labor and capital, where labor is the static input and capital is the dynamic input. Such a two input production function is typical when output Y is measured in a value added fashion (for a discussion of value added versus gross output production functions, see Levinsohn and Petrin (2003) and the references cited therein). We point this out because in some settings it may be felt that a firm's labor decision does not adjust to fully adjust to TFP shock in a given period, either because labor is also a dynamic input or because adjustment costs prevent labor from being a fully flexible factor. In these cases, we can still proceed with the above empirical strategy, but instead of taking the first order condition for labor, we take the first order condition for *any* static input that was subtracted out of the value added production function. This will yield an analogous condition to (10), so long as we add this input back into the production function. This will introduce another parameter to be estimated. The key point is that our approach is not married to the assumption that labor is a static input, so long as there is some static input that enters the production function whose expenditure we can measure.

As a final remark about the above empirical strategy, since our estimation of the labor and capital coefficients amounts to a single step GMM procedure, we can appeal to the standard formulas for asymptotic standard errors. This stands in contrast to the more complicated derivation of standard errors for the two step estimators used in Olley and Pakes (1996) and Levinsohn and Petrin (2003).

5 Empirical Example

We now apply our method to the same Chilean manufacturing data used in Levinsohn and Petrin (2003). We use industry 311 (food products) for illustrative purposes. The data for each industry contain a panel of firms extending 18 years (1979-1996). See LP and the references cited therein for a further description of the data and the construction of the variables. In the following tables, we present estimates of the labor and capital coefficient derived from OLS, the LP estimator with electricity as the intermediate input (the preferred input in LP), and our method (which we label GNR). We use a third order polynomial approximation for our estimation of the first order Markov process. In addition, for both the LP and GNR estimates, we use the bootstrap to generate the confidence intervals.

Table 1: Industry 311

Method	Labor	95% CI	Capital	95% CI
OLS	.953	.932,.947	.400	.389,.411
LP	.647	.595,.700	.399	.292,.505
GNR	.414	.402,.425	.362	.274,.391

The most apparent feature that emerges from the tables is the extent to which the GNR estimates reduce the labor coefficient as compared to both OLS and LP. The elasticity of labor productivity is on average half as small under GNR as compared to LP. This reduction is exactly what we expect under the hypothesis that the proxy variable approach does not fully recover productivity due to ignored dimensions of heterogeneity, which leaves endogeneity in the error term during the first stage of the LP/OP estimation. While the capital coefficients achieved under both GNR and LP are roughly comparable, the reduction of the labor coefficient under GNR has an important implication for the understanding of these industries. In particular, all of the industries significantly exhibit decreasing returns to scale under GNR while they significantly exhibit increasing returns to scale under LP. Increasing returns to scale is a problematic finding since the underlying economic framework described in section 2 assumes price taking behavior, which cannot be easily reconciled with a market structure of firms having increasing returns to scale (mere existence of price equilibrium in a competitive setting is not straightforward if firm technologies are not convex). We view the GNR finding of decreasing returns as a further merit of the approach as competitive behavior is the intuitive expectations for the industries under question (food products, apparel, etc).

Another feature of the tables to notice are the much tighter confidence intervals found under GNR as compared to LP. This is a natural result of the fact that GNR estimation proceeds in a single step, with only one nonparametric term to estimate (namely the first order Markov process on productivity), as opposed to the two step estimator of LP with two nonparametric terms to estimate (both in the first and second stage).

It is straightforward to generalize our procedure to allow for a more general Markov process on productivity, such as letting the process on TFP be higher order Markov or dependent upon the firm's R&D expenditure. In addition, we can incorporate a selection correction into our estimator in order to account for the endogenous exit of firms from an industry due to an unfavorable productivity shock. Such exits lead to an unbalanced panel, and treating such observations as missing at random could bias the estimate of the capital coefficient as discussed by Olley and Pakes (1996). We can analogously incorporate a selection correction into our approach, but with the advantage once again of not requiring inversion on the proxy (the existence of a monotonic relationship between the proxy and productivity, and hence invertibility of the proxy, is central to their development of the correction procedure). In particular, the selection correction under our method amounts to controlling for capital k_{jt} when regressing ω_{jt} on ω_{jt-1}. All of these extensions are sketched in the working paper version of the proposal (Gandhi et al., 2008).

6 Heterogeneous Factor Productivity

The existing literature on production function estimation has focused mainly on heterogeneity through the TFP term, ω_{jt}. However, as pointed out earlier, there is evidence that ω_{jt} fails to account for a significant portion of heterogeneity. One possible explanation for this is the existence of heterogeneity in factor productivity. Since a model with this added level of heterogeneity nests a model without it, it is possible to test for the existence of such heterogeneity. In this section we show how to extend our method to account for different forms of factor productivity heterogeneity.

Depending on the data and form of heterogeneity one wants to allow for, the extension can be trivial. For example, if one wants to allow for time-specific coefficients (α_t, β_t) one can simply run our procedure separately for each time period. In the same manner, if one has access to a long enough panel allowing for firm specific coefficients (α_j, β_j) is also straightforward. In this case, we can simply run our procedure separately for each firm. Even though the ability to allow for these extensions is not new to our method, as this can also be done using the methods of OP and LP, it has not been explicitly addressed in the literature.

The more interesting case arises when we want to allow for firm specific coefficients without access to a long panel. This is the case for most data sets and it is not a trivial extension. However, as we show below, we can still identify the distributions of $\alpha_j, \beta_j, \omega_{jt}$ and ε_{jt} *nonparametrically.*

Heterogeneous Labor Productivity

Heterogeneity in labor productivity, which corresponds to allowing the coefficient on labor α_j vary across firms j, has theoretically appealing properties. In particular, it has the potential of explaining certain labor market "puzzles" such as the the simultaneous correlation between wages, firm size, and productivity observed in a variety of micro data sets (see e.g., Lucas (1978); Mortensen and Pissarides (1994); Davis and Haltiwanger (1995)). We begin by showing how to identify and estimate the distribution of labor productivity within our framework.

Consider a version of equation (10), where α is replaced with α_j which varies across firms j. In particular assume each α_j is drawn from a distribution G, which represents the population distribution of labor productivity. Assume each firm is in operation for at least two time periods which we generically call $t = 1, 2$. Consider now the system (10) for periods $t = 1$ and $t = 2$, which yields

$$\begin{aligned} \ln(\frac{P_{j1}Y_{j1}}{W_{j1}L_{j1}}) &= -\ln(\alpha_j) + \varepsilon_{j1} \\ \ln(\frac{P_{j2}Y_{j,2}}{W_{j2}L_{j,2}}) &= -\ln(\alpha_j) + \varepsilon_{j2}. \end{aligned} \tag{13}$$

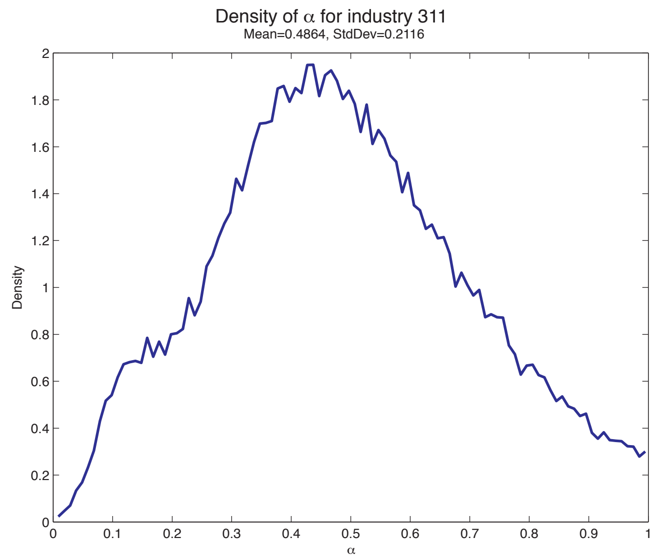
Let $s_{j1} = \ln(\frac{P_{j1}Y_{j1}}{W_{j1}L_{j1}})$ and $s_{j2} = (\frac{P_{j2}Y_{j2}}{W_{j2}L_{j2}})$. From the data, we can directly recover the joint distribution of the random vector (s_1, s_2). As we show below it follows that, from a theorem of Kotlarski (1967), we can separately identify *nonparametrically* the distribution of the random variable $\alpha \sim G$ and the distributions of the random variables $\varepsilon_i \sim F_i$ for $i = 1, 2$.

Theorem 1. *If the joint distribution of* (s_1, s_2) *admits a non-vanishing characteristic function and the random variables* $(\alpha, \varepsilon_1, \varepsilon_2)$ *are mutually independent, then the joint distribution of* (s_1, s_2) *identifies the the joint distribution of* $(\alpha, \varepsilon_1, \varepsilon_2)$.

Proof. Recall the assumption that $E[\varepsilon_i] = 0$ for $i = 1, 2$. With this restriction in place, refer to the system (13). The theorem of Kotlarski (1967) implies that that the distribution of $a = -\ln(\alpha)$ and the distributions of μ_i for $i = 1, 2$ are nonparametrically identified from the joint distribution of (s_1, s_2). Thus we can identify the distribution G of the one-to-one transformation $1/\exp(a) = \alpha$ and the distribution F_i of ε_i for $i = 1, 2$. □

The basic idea behind the identification theorem is that in a single cross section (for say $t = 1$), we cannot separately identify whether the heterogeneity in labor's share of revenue across firms is due to heterogeneity in labor productivity α, or heterogeneity in the shocks ε_1. However, if we bring a second cross section to bear on the problem through $t = 2$, we can use the *persistence* in labor's share of revenue within a firm across time to separate out the effects. Thus the higher the correlation in labor's share of revenue between the two time periods, the more of the observed cross sectional variance in labor's share of revenue we attribute to productivity differences in α.

To implement the estimation of the the distribution of $(\alpha, \{\varepsilon_t\}_{t=1}^{T})$ in the system (13), we can apply semi-parametric factor methods such as those described by e.g., Carneiro et al. (2003). We now present estimates of the distribution of α for industry 311 from the Chilean data (which we used in Section 5) assuming that each period t's exogenous shock ε_t comes from a common distribution F (an assumption that our identification theorem shows that we can relax). While we can recover the distributions nonparametrically, we instead estimate the distributions F and G using flexible parametric forms. In particular, we let the distribution of $a = -\ln(\alpha)$ follow a truncated (between $(0, \infty)$) mixture of normals with 5 components and F follow a mixture of normals with 3 components and obtain estimates by maximum likelihood. In both cases we treat missing observations for a firm as missing at random. In the graph below, we present the estimated densities of labor productivity for the four Chilean industries we previously examined. As is apparent, these industries reveal considerable heterogeneity, with average labor productivity roughly corresponding to the labor coefficient estimated under the GNR method of the previous section. The standard deviations are roughly .2, and thus there is a positive mass of firms where the labor productivity of some firms can be over twice as high as the labor productivity of other firms. Examining the full consequences of these findings for the relationship between wages, firm size, and productivity, is a subject of ongoing research.



Given our estimates for the distribution of α_j, we next turn to estimation of the capital coefficient β. The difficulty here is that we only know the marginal distribution of α_j. We do not know each individual firm's value of α_j, nor do we know the correlation between α_j and other variables such as labor and capital. This lack of information prevents us from solving directly for ω_{jt} in the production function and from using the procedure described in Section 5 to estimate β. It also causes problems for other potential methods for estimating β, as we discuss below. In spite of these limitations, we are able to exploit the first-order condition to provide a solution to this problem. Since we are unable to obtain an estimate of ω_{jt} directly, we cannot be non-parametric about the process on productivity as we were in section 5, because we cannot regress ω_{jt} on ω_{jt-1}. However, if we impose parametric assumptions the evolution of ω_{jt}, we can recover β given our estimates for the distribution of α_j.

For example, if we assume that ω_{jt} follows an AR(1) process: $\omega_{jt} = \rho\omega_{jt-1} + \eta_{jt}$, then using the production function in periods t and $t-1$, η_{jt} can be written as the following:

$$\begin{aligned}
\eta_{jt} &= \omega_{jt} - \rho\omega_{jt-1} \\
&= y_{jt} - \alpha_j l_{jt} - \beta k_{jt} - \varepsilon_{jt} - \rho\left(y_{jt-1} - \alpha_j l_{jt-1} - \beta k_{jt} - \varepsilon_{jt-1}\right) \\
&= \left(y_{jt} - \rho y_{jt-1}\right) - \alpha_j\left(l_{jt} - \rho l_{jt-1}\right) - \beta\left(k_{jt} - \rho k_{jt-1}\right) - \left(\varepsilon_{jt} - \rho\varepsilon_{jt-1}\right)
\end{aligned}$$

Notice that, since we do not know the value of α_j for any individual firm, we cannot form η_{jt} even given a value for β. Further, since α_j is correlated with l_{jt} (and l_{jt-1}), we cannot simply integrate it out using the marginal distribution of α_j we estimated above. However, if we let ψ_{jt} denote an instrument that is uncorrelated with η_{jt}, then

$$\begin{aligned} E[\eta_{jt}\psi_{jt}] &= E\{[(y_{jt}-\rho y_{jt-1})-\alpha_j(l_{jt}-\rho l_{jt-1})-\beta(k_{jt}-\rho k_{jt-1})-(\varepsilon_{jt}-\rho\varepsilon_{jt-1})]\psi_{jt}\} \\ &= E(y_{jt}\psi_{jt})-\rho E(y_{jt-1}\psi_{jt})-E(\alpha_j l_{jt}\psi_{jt})+\rho E(\alpha_j l_{jt-1}\psi_{jt}) \\ &\quad -\beta E(k_{jt}\psi_{jt})+\beta\rho E(k_{jt-1}\psi_{jt})-E(\varepsilon_{jt}\psi_{jt})+\rho E(\varepsilon_{jt-1}\psi_{jt}) \end{aligned} \tag{14}$$

Terms involving y and k can be directly calculated from the data. The last two terms are equal to zero as long as our instruments are uncorrelated with ε_{jt}. The two terms involving α_j are problematic since α_j is correlated with l_{jt} and l_{jt-1} and potentially the instrument as well. Notice, however, that we only need the expectation of $\alpha_j l_{jt}\psi_{jt}$ and $\alpha_j l_{jt-1}\psi_{jt}$. We show below that by taking advantage of the first-order condition, we can estimate this expectation for all instruments ψ_{jt} that are independent of ε_{jt}. [2] That is, although we cannot form η_{jt} for any firm, we can form the moment condition $E[\eta_{jt}\psi_{jt}]=0$ and use it to estimate β.

To see how we can form $E(\alpha_j l_{jt}\psi_{jt})$ and $E(\alpha_j l_{jt-1}\psi_{jt})$, let $s^{\psi}_{jt}\equiv s_{jt}-\ln(l_{jt}\psi_{jt})$ with s_{jt} the log inverse of the labor share, l_{jt} the log of labor, and ψ_{it} an instrument as defined previously. By rearranging the first order condition for labor and multiplying by $\frac{1}{l_{jt}\psi_{jt}}$ we have:

$$s^{\psi}_{jt}=-\ln(\alpha_j l_{jt}\psi_{jt})+\varepsilon_{jt} \tag{15}$$

Since the distribution of $\ln\left(s^{\psi}_{jt}\right)$ is given in the data and we have estimates for the distribution of ε_{jt} from above, we can deconvolve the distribution of $-\ln(\alpha_j l_{jt}\psi_{jt})$ and from that obtain an estimate of the distribution of $\alpha_j l_{jt}\psi_{jt}$. This yields an estimate of $E[\alpha_j l_{jt}\psi_{jt}]$. A similar argument holds for $E[\alpha_j l_{jt-1}\psi_{jt}]$. Thus we can form the moment conditions (14) and use them to estimate β and ρ. This can easily be extended to allow for more general processes on the Markovian evolution of ω_{jt}, for example an AR(2), although it requires additional instruments to estimate the additional parameters.

Heterogeneous Capital Productivity

So far we have shown how to estimate the distribution of α_j and to use that in estimating a homogeneous capital coefficient β. However, we can go further. If we assume that the random coefficient on capital is of a particular form, then we can estimate a random coefficient for capital, β_j, as well. Let $u_j\equiv\alpha_j-\overline{\alpha}$, where $\overline{\alpha}$ is the mean of α_j. Suppose that $\beta_j=\beta+u_j\lambda$, where λ denotes the sign and degree of correlation between the labor and capital coefficients. The moment equations for this case are very similar to the case with a homogeneous capital coefficient.

$$\begin{aligned} 0 &= E[\eta_{jt}*\psi_{jt}] \\ &= E(y_{jt}\psi_{jt})-\rho E(y_{jt-1}\psi_{jt})-E(\alpha_j l_{jt}\psi_{jt})+\rho E(\alpha_j l_{jt-1}\psi_{jt}) \\ &\quad -\lambda E[\alpha_j k_{jt}\psi_{jt}]+\lambda\rho E[\alpha_j k_{jt-1}\psi_{jt}]-\overline{\alpha}\lambda E[k_{jt}\psi_{jt}]+\overline{\alpha}\lambda\rho E[k_{jt-1}\psi_{jt}] \\ &\quad -\beta E(k_{jt}\psi_{jt})+\beta\rho E(k_{jt-1}\psi_{jt})-E(\varepsilon_{jt}\psi_{jt})+\rho E(\varepsilon_{jt-1}\psi_{jt}) \end{aligned} \tag{16}$$

We can use a variation of (15) to estimate $E[\alpha_j k_{jt}\psi_{jt}]$ and $E[\alpha_j k_{jt-1}\psi_{jt}]$. In this case there is one more parameter to estimate here, λ, so one additional instrument is required as compared to the homogeneous capital coefficient case. Note that $\overline{\alpha}$ does not need to be estimated since this is already known from the estimation of the distribution of alpha.

[2] There are many possible instruments here, since ε is measurement error and is assumed to be independent of all random variables except Y and functions of Y, such as s.

7 Directions for Further Work

7.1 Extensions to Other Production functions

We have thus far illustrated our method with respect to a Cobb-Douglas production function. As already mentioned, this functional form is the standard in the literature. However since we use the first order condition derived from this parametric form, whereas the proxy inversion relies upon monotonicity assumptions not tied to a particular production function, the question arises as to whether our approach is robust to deviations from Cobb-Douglas. Of course, even the proxy variable approach must start out by writing down a parametric form of the production function, which is used explicitly in the first stage of their estimation. However, the monotonic form of the proxy demand equation $\delta_t(k_{jt}, \omega_{jt})$ does not come from a particular parametric form of the production function, but does come at the cost of imposing the scalar unobservability assumption. In contrast, our approach makes full use of the parametric form inherent when the researcher writes down the production function, and takes the first order condition for static input demand with respect to this form. The advantage of course is that this gave us identifying power to control for unobserved input and output price heterogeneity. The question is thus whether using the parametric form of the production to derive input demand is a worse assumption than imposing the scalar unobservability assumption and treating input demand nonparametrically. Our initial results on the Chilean data suggest that we are in fact controlling for more unobserved productivity than the proxy approach, but a full analysis of robustness will require Monte Carlo simulations.

However from production theory, it is well known that the Cobb-Douglas and the CES functional form exhaust the space of production functions that are homogeneous and additively separable. Thus it is natural to ask whether, at the least, our approach can be analogously extended to a CES production function to yield a solvable system like (12). This is in fact possible, and we provide this development in the working paper version of our proposal (Gandhi et al., 2008). Further developing our approach to handle random coefficients in the CES production function is an avenue for future research.

7.2 Risk Aversion and Quantile Maximization

Our first order condition for labor input demand is based on the assumption that the firm is a risk neutral expected profit maximizer. Of course this is an assumption, and the question is whether we can add risk tastes as a parameter of the model. Under the assumption that the firm is quantile maximization (i.e., the firm chooses a lotteries over profits that maximize a certain quantile of the distribution, such as the median, min, max, etc), we can in fact easily allow for simultaneous estimation of risk tastes and the production function parameters under weaker identifying assumptions than expected profit maximization. Quantile maximization has the advantage of being robust to tail behavior in distributions, and has recently been established as a formal decision theory in and of itself. The working paper of the proposal (Gandhi et al., 2008)contains a sketch as to this development, whose empirical implementation is another avenue of further work. In particular, the quantile maximization approach allows us to relax the strict behavioral assumption required of expected profit maximization.

7.3 Application to the Analysis of Productivity

We ultimately hope that our approach to estimating production functions allows for a richer study of productivity itself, which was the original intent of Olley and Pakes (1996). Since our estimation method explicitly accounts for heterogeneity in wages, and labor productivity differences under random coefficients, our method produces different estimates of productivity than OP or LP. A particular application of interest is the extent to which opening up an economy enhances firm productivity. This question was previously studied by Pavcnik (2002) using the OP estimator and the Chilean data. Since our estimates of productivity control for wage differences (and hence labor quality), it allows us to more carefully study whether productivity enhancements due to the opening of an economy are due to genuine positive productivity shocks brought about by this event, or allocation of output to potentially larger firms that hire higher quality workers. Since we have the Chilean data in hand, this question is ripe for study.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Structural Identification of Production Functions," 2006. Unpublished Manuscript, UCLA Economics Department.

Ackerberg, Daniel, Lanier Benkard, Steven Berry, and Ariel Pakes, "Econometric Tools for Analyzing Market Outcomes," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics*, Vol. 6, Elsevier Science, 2007. Forthcoming.

Bond, Stephen and Måns Söderbom, "Adjustment Costs and the Identification of Cobb Douglas Production Functions," 2005. Unpublished Manuscript, The Institute for Fiscal Studies, Working Paper Series No. 05/4.

Carneiro, Pedro, Karsten Hansen, and James J. Heckman, "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, May 2003, *44* (2), 361–422. 2001 Lawrence R. Klein Lecture.

Davis, Steven K and John Haltiwanger, "Employer Size and The Wage Structure in U.S. Manufacturing," NBER Working Papers 5393, National Bureau of Economic Research, Inc December 1995.

Fox, Jeremy and Valerie Smeets, "Do Input Quality and Structural Productivity Estimates Drive Measured Differences in Firm Productivity?," 2007.

Gandhi, Amit, Salvador Navarro, and David Rivers, "On the Estimation of Production Functions," 2008. Unpublished manuscript, University of Wisconsin-Madison.

Griliches, Zvi and Jacques Mairesse, "Production Functions: The Search for Identification," in "Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium," New York: Cambridge University Press, 1998, pp. 169–203.

Hamermesh, Daniel S., *Labor Demand*, Princeton, NJ: Princeton University Press, 1993.

Kotlarski, Ignacy I., "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, 1967, *20*, 69–76.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, April 2003, *70* (243), 317–342.

Lucas, Robert E., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, Autumn 1978, *9* (2), 508–523.

Marschak, Jacob and W.H. Andrews, "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 1944, *12*, 143–205.

Mortensen, Dale T. and Christopher A. Pissarides, "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, July 1994, *61* (3), 397–415.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, November 1996, *64* (6), 1263–1297.

Pavcnik, Nina, "Trade Liberalization Exit and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies*, 2002, *69*, 245–276.

Robinson, Peter M., "Root-N-Consistent Semiparametric Regression," *Econometrica*, July 1988, *56* (4), 931–954.