Estimating Production Functions with Robustness Against Errors in the Proxy Variables*

Guofang Huang and Yingyao Hu
Yale University and Johns Hopkins University

October 29, 2011

Abstract

This paper proposes a new semi-nonparametric maximum likelihood estimation method for estimating production functions. The method extends the literature on structural estimation of production functions, started by the seminal work of Olley and Pakes (1996), by relaxing the scalar-unobservable assumption about the proxy variables. The key additional assumption needed in the identification argument is the existence of two conditionally independent proxy variables. The assumption seems reasonable in many important cases. The new method is straightforward to apply, and a consistent estimate of the asymptotic covariance matrix of the structural parameters can be easily computed.

1 Introduction

The literature on estimating production functions on panel data using control functions has focused mainly on two major issues. One is the simultaneity bias, and the other is the sample-selection problem caused by firm entry and exit. Both problems exist because of the unobserved productivity in the production function. The seminal paper by Olley and Pakes (1996) (hereafter OP) proposes the idea of using investment as a proxy variable to

*We thank James Levinsohn for providing us with the Chilean manufacturing industry survey data. We also thank Xiaohong Chen, Joe Harrington, Przemek Jeziorski, Ahmed Khwaja, Zhipeng Liao, Arvind Magesan, Zhengtao Shi, Matt Shum and K. Sudhir for helpful comments. Any remaining errors are our own.

correct for the simultaneity bias.[1] Their key observation is that if productivity is the only unobserved factor affecting investment and if investment is a strictly increasing function of productivity, then a nonparametric function of investment and other covariates can be used to control for the latent productivity when estimating the production function. Important discussions and extensions of the method have since been made, for example, by Levinsohn and Petrin (2003) (hereafter LP), Ackerberg, Caves, and Frazer (2006) (hereafter ACF), etc. LP suggest that intermediate inputs may be better proxy variables for productivity. They argue that when using intermediate inputs as proxy variables, the primitive conditions that ensure the monotonicity condition would be easier to come by, and that the intermediate inputs could be much less lumpy and have fewer values of zero. ACF suggest an improvement to avoid a potential identification issue in the first stage of LP's procedure. The problem they point out is very clear. Suppose that labor demand, like investment, is also a function of capital and productivity but no other unobserved factors. Then, after controlling for capital and productivity perfectly by the nonparametric function, there would be no independent variation in labor left to identify the coefficient of labor in LP's first stage.

The two fundamental assumptions maintained by these methods are: a) the demand functions of the proxy variables are strictly increasing in the unobserved productivity; and b) the productivity is the only unobserved determinant of the demands of these proxy variables. Under these assumptions, the methods have provided very intuitive and convenient ways to control for the latent productivity and consistently estimate the production function and firm-level productivity. These methods have been used in a large number of applications. The estimates of production functions and firms' productivity are often used as inputs in the analysis of issues such as the impact of deregulations, trade, etc, on firm productivity (e.g., Pavcnik (2002), Bernard, Eaton, Jensen, and Kortum (2003), Javorcik (2004), Aw, Roberts, and Xu (2008), etc).

However, the second assumption of productivity being the single unobserved factor affecting investment/intermediate inputs has raised concerns as far back as OP's original paper. LP also point out that a major criterion in selecting their proxy variable is to avoid inputs that could be subject to the influence of other unobserved factors (see p.326 in LP). The importance of the issue can be seen from the following two problems associated with the assumption. First, if there were additional unobserved factors affecting investment and input decisions, the OP/LP/ACF procedures cannot fully control for the latent productivity.[2] In

[1] They suggest using the propensity of exiting to correct for the sample-selection problem.

[2] Closely related to the literature, Imbens and Newey (2009) use the conditional CDF of the input given some instrumental variables, such as cost shocks, as the control variable for the latent productivity. But as

general, some other unobserved factors could be affecting the actual investment and inputs. For example, measurement errors may be ubiquitous in the data. As LP point out, disruptions in the supply of intermediate inputs and unobserved changes in their inventories could all make the actual inputs of materials and fuels differ from those observed in their data (see p.326 in LP). One may get some sense of the seriousness of the simultaneity bias problem by looking at the estimated labor coefficients and returns to scale. Without any correction for the simultaneity bias, the OLS tends to overestimate the labor coefficient and the returns to scale. For example, for the Food industry in Table 1 in ACF, with OLS, the estimated labor coefficient and returns to scale are, respectively, 1.080 and 1.416. After correction using the procedure of ACF or LP, the estimates make much more sense. For example, for ACF's method, the estimated labor coefficient is around 0.84, and the estimated returns to scale is about 1.24. However, the persistent existence, from 1979 to 1986, of significant increasing returns to scale in the Food industry may prompt one to ask whether the scalar unobservable assumption is seriously violated.

Second, the assumption forces us to give up some important sources of identification. This problem manifests itself most clearly in the first stage of LP's procedure, where they estimate the labor coefficient. Suppose, like what is assumed for the proxy variables, the demand of labor is also a function of only capital and the latent productivity (but not of any other unobserved factors). The identification issue in LP's first stage, as ACF point out, then is that the labor input is left with no independent variation after controlling for a nonparametric function of capital and the latent productivity. To maintain logical consistency, one does not want to both use the additional sources of variation in labor input—due to cost shocks, for example—to identify the labor coefficient in LP's first stage *and* use an intermediate input as a perfect proxy variable for the latent productivity. Relatedly, Bond and Söderbom (2005) point out the difficulty of identifying fully flexible inputs when there is no variation in input prices across firms; they suggest that one may use stochastic input adjustment costs to help identify the input coefficients. The authors argue that with stochastic adjustment costs it is better to use the instrumental variable methods, as in Blundell and Bond (2000), to estimate production functions since the model of OP and LP would be misspecified if the stochastic input adjustment costs were present.

We propose a new method in this paper to estimate production functions, *allowing the demand of the proxy variables to be affected by other unobserved factors in addition to*

Imbens (2007) points out, such an approach cannot correct all the simultaneity bias if the input demand is also affected by other unobservables in addition to the latent productivity.

the latent productivity.[3] The insight of our method is that, because researchers normally have multiple proxy variables, such as some intermediate inputs and investment, available for productivity, we may be able to find two such proxy variables that, conditional on productivity, are independent of each other in some reasonable cases. We may intuitively view these two proxy variables as two contaminated measures of productivity. Then, loosely speaking, we can use one proxy variable as the instrument for the other contaminated measure of productivity to fully control for the latent productivity in the estimation of production functions. Hu and Schennach (2008) establish the corresponding identification results for a general class of nonclassical measurement-error models. In this paper, we apply their results to show that production functions can be identified and estimated in many important cases even when the scalar-unobservable assumption is not satisfied by the proxy variables.

Two key conditions are needed for our identification of production functions. The first one is the conditional independence condition alluded to above. As we will discuss in detail later, this condition seems reasonable in many important cases. The second is that the conditional density of each proxy variable given productivity and all the control variables satisfy an injectivity condition. This condition is satisfied, for example, if the conditional expectations of the proxy variables given productivity and all the control variables are strictly increasing in productivity. We can view this condition as a generalization of the original monotonicity condition of OP in the more general specification for the demand of the proxy variables.

Our identification strategy provides the foundation for an alternative estimation method without the above two problems associated with the scalar-unobservable assumption about the proxy variables. As our identification explicitly allows for additional unobservable factors to affect investment/intermediate-input demand, it is robust against such a possibility and frees up many important sources of variation for identifying the input coefficients in the production function.

We propose a semi-nonparametric maximum likelihood estimation method to estimate the structural parameters in the production function. The latent densities included in the likelihood functions are approximated by using Hermite series (Gallant and Nychka (1987)). We compare our method to those of LP and ACF in a Monte Carlo exercise. The Monte Carlo evidence shows the robustness of our method, but not of the other methods, against the additional unobserved factors. The methods of LP and ACF tend to overestimate the

[3]We focus on dealing with the simultaneity bias in this paper.

labor coefficient when there are additional unobserved factors affecting the demand of the intermediate inputs. We also apply all these methods to the same data of the Chilean manufacturing industry (1979-1986), as used by LP and ACF, and our comparison yields similar results. The estimates generated using the methods of LP and ACF show relatively large labor coefficients and significant increasing returns to scale. In comparison, our method produces smaller labor coefficients and largely constant returns to scale.

The rest of the paper proceeds as follows. Section 2 briefly reviews the methods of estimating production functions proposed by OP, LP and ACF. Section 3 describes our model, shows the identification of the model, and discusses the conditions required in our identification. Section 4 proposes a new estimation method based on our identification result. Section 5 discusses the calculation of asymptotic covariance matrix of the estimates. Section 6 contains the comparison of our method with those of LP and ACF, using simulated data and the Chilean manufacturing-industry census data. Section 7 concludes.

2 Literature Review

We start by putting down the model used by OP/LP/ACF. Throughout the paper, we follow the tradition of using uppercase letters to denote levels and lowercase letters to denote the log of levels. And to simplify notation, we omit the subscript for firms. The goal is to estimate the following form of industry production function

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t$$

by using firm-level panel data, where y_t, l_t, and k_t are, respectively, the output (value added), labor and capital inputs; ω_t is the latent productivity that is serially correlated; and η_t is the residual term with $E(\eta_t | l_t, k_t) = 0$. The productivity ω_t follows an exogenous first-order Markov process

$$\omega_t = E(\omega_t | \omega_{t-1}) + \xi_t$$

where ξ_t is mean-independent of ω_{t-1}. The capital is depreciated and accumulated according to following equation:

$$K_t = (1 - \delta) K_{t-1} + I_{t-1}$$

where $\delta \in (0, 1)$ is the depreciation rate and I_{t-1} is the investment made in period $t - 1$. OP note that, under certain conditions, the firm investment is determined as

$$i_t = \iota_t(\omega_t, k_t)$$

where $\iota_t(\omega_t, k_t)$ is the investment demand function that is strictly increasing in ω_t for any given k_t. LP make use of the following intermediate input demand function:

$$m_t = \mu_t(\omega_t, k_t)$$

that is similarly assumed to be strictly increasing in ω_t for any given k_t in their estimation procedure. The difficulty in estimating the production function is that normally l_t and k_t are correlated with ω_t, and we do not observe ω_t.

2.1 Olley and Pakes (1996)

OP propose a structural approach to estimate the production function. The key observation of OP is that we can use investment as a proxy for ω_t. More specifically, under some conditions, the investment demand function $\iota_t(\omega_t, k_t)$ is strictly increasing in ω_t , so we can invert $\iota_t(\omega_t, k_t)$ to get $\omega_t(i_t, k_t)$. Based on this insight, OP propose the following procedure to estimate the production function:

Step 1: semiparametrically estimate

$$y_t = \beta_l l_t + \phi_t(i_t, k_t) + \eta_t$$

where $\phi_t(i_t, k_t) = \beta_0 + \beta_k k_t + \omega_t(i_t, k_t)$ is estimated nonparametrically. We get an estimate of β_l, ϕ_t and ϕ_{t-1} in this stage.

Step 2: semiparametrically estimate

$$y_t - \hat{\beta}_l l_t = \beta_k k_t + \rho(\hat{\omega}_{t-1}) + \xi_t + \eta_t$$

where $\rho(\omega_{t-1}) \equiv E(\omega_t|\omega_{t-1})$ is estimated nonparametrically based on $\hat{\omega}_{t-1} = \hat{\phi}_{t-1} - \beta_k k_{t-1}$ and $\hat{\omega}_t = \hat{\phi}_t - \beta_k k_t$. Here, one gets a consistent estimate of β_k using the condition that k_t is mean-independent of ξ_t.

2.2 Levinsohn and Petrin (2003)

The insight of LP is that we can actually use intermediate inputs, such as materials and energy inputs, as the proxy for ω_t if similarly the demand functions for such inputs are also strictly monotonic in ω_t for any given k_t. For example, suppose that we have the following demand function for the material input m_t:

$$m_t = \mu_t\left(k_t, \omega_t\right)$$

where $\mu_t\left(\omega_t, k_t\right)$ is strictly increasing in ω_t for any given k_t. Then, following OP's idea, we can use a nonparametric function, $\phi\left(m_t, k_t\right)$, of k_t and m_t to control for ω_t when estimating the production function.

The LP method has two advantages over the original OP method. First, one does not have to get rid of the observations with zero investment. Second, primitive conditions that ensure monotonic intermediate input demand functions are easier to derive and test since intermediate inputs have no dynamic implications.

2.3 Ackerberg, Caves and Frazer (2006)

The critique of ACF is that the first stages in OP and LP's procedures are actually not identified because l_t would have no independent variation when ϕ_t is nonparametrically estimated. To see this, suppose that, similar to the demand of m_t and i_t we have the following labor demand function:

$$l_t = l_t\left(\omega_t, k_t\right)$$

And for LP's method, by assumption, one has $\omega_t = \omega_t\left(m_t, k_t\right)$. Then, $l_t = l_t\left(\omega_t\left(m_t, k_t\right), k_t\right)$ is thus also a function of $\left(m_t, k_t\right)$, and would be collinear with the terms used to approximate the unknown function of $\phi\left(m_t, k_t\right)$. ACF assume that the decision on l_t is made before that of m_t, and the intermediate input demand function would be $m_t = \mu_t\left(\omega_t, k_t, l_t\right)$, where μ_t is assumed to be strictly increasing in ω_t for any given $\left(k_t, l_t\right)$. So, after substituting in the expression for ω_t, the production function can be written as follows:

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t\left(m_t, k_t, l_t\right) + \eta_t$$

To get around the identification problem of l_t in the first stage of LP's procedure, they suggest estimating the coefficients of both l_t and k_t in the second stage. They propose estimating the production function through the following two steps:

Step 1. To net out the effect of η_t, nonparametrically estimate the unknown function of $\varphi\left(m_t, l_t, k_t\right) = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t\left(m_t, k_t, l_t\right)$.

Step 2. Estimate $\left(\beta_l, \beta_k\right)$ using the following set of two moment conditions,

$$E\left(\xi_t \cdot \begin{pmatrix} k_t \\ l_{t-1} \end{pmatrix}\right) = 0$$

where $\xi_t(\beta_l, \beta_k) = \omega_t - E(\omega_t|\omega_{t-1})$, $\hat{\omega}_t = \hat{\varphi} - \beta_l l_t - \beta_k k_t$, and $E(\omega_t|\omega_{t-1}) = \rho(\omega_{t-1})$ is also estimated nonparametrically.

2.4 Discussion

All the above methods rely critically on the key assumption that the latent productivity is the only unobservable affecting the intermediate inputs and investment. So, in the cases where the observed intermediate inputs and investment are also affected by measurement errors, optimization errors, cost shocks, etc., these methods would not be able to eliminate the simultaneity bias. To illustrate the problem, suppose that the material demand function is a linear function as the following:

$$\begin{aligned} m_t &= \mu_t + \epsilon_t \\ \mu_t &= \tilde{\gamma}_0 + \tilde{\gamma}_1 \omega_t + \tilde{\gamma}_2 k_t + \tilde{\gamma}_3 l_t \end{aligned}$$

In this case, the latent productivity can be written as a linear function of (k_t, m_t, l_t) and ϵ_t

$$\omega_t = \gamma_0 + \gamma_k k_t + \gamma_l l_t + \gamma_m (m_t - \epsilon_t)$$

where $(\gamma_0, \gamma_k, \gamma_l, \gamma_m)$ are functions $(\tilde{\gamma}_0, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3)$. Substituting the expression for ω_t into the production function, we have

$$y_t = (\beta_0 + \gamma_0) + (\beta_l + \gamma_l) l_t + (\beta_k + \gamma_k) k_t + \gamma_m m_t - \gamma_m \epsilon_t + \eta_t$$

However, the equation now cannot be consistently estimated since $Cov(m_t, \epsilon_t) \neq 0$. Thus, when one tries to use a nonparametric function of (l_t, k_t, m_t) to control for ω_t, part of ω_t that is a linear combination of m_t and ϵ_t would always be missed as long as $var(\epsilon_t) > 0$. Therefore, in this case, the first stage estimates of LP and ACF's procedure would be inconsistent, and the estimates of their second stage would also be problematic.

In the next section, we show that with commonly available data, we can still identify the structural parameters in the production function even if the observed intermediate inputs and investment are also affected by other unobservables.

3 Model and Identification

Starting in this section, we study the identification and estimation of production functions assuming that each observed intermediate input (and investment) is affected by another unobservable factor in addition to productivity. In the following, we first outline the main idea of our identification strategy and some issues involved in implementing the idea; then we set up our generalized model and show the identification of the model.

Our key identification idea is to use two intermediate inputs (or one intermediate input plus the investment) simultaneously as two proxy variables for productivity. The two proxy variables can be thought as two contaminated measures of the latent productivity. Although, now, one cannot directly invert the demand function of an input to fully control for the latent productivity due to the additional unobserved factor in the demand of the input, we can use the other input as an instrument for the first crude input proxy variable. Given this perspective of the model, we can employ Hu and Schennach (2008)'s identification result for nonclassical measurement-error models to show the identification of parameters in the production function. More specifically, suppose that we are interested in estimating the following equation of a dependent variable y_t,[4]

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t$$

And suppose that we have two contaminated measures of the latent variable ω_t: x_t and z_t, such that 1) the three dependent variables (y_t, z_t, x_t) are independent of each other conditional on ω_t and the control variables—i.e., $f\left(y_t|\omega_t, z_t, x_t, k_t, l_t\right) = f\left(y_t|\omega_t, k_t, l_t\right)$ and $g\left(z_t|\omega_t, x_t, k_t, l_t\right) = g\left(z_t|\omega_t, k_t, l_t\right)$; 2) $g\left(z_t|\omega_t, k_t, l_t\right)$ and $g\left(\omega_t|x_t, k_t, l_t\right)$ are injective for any given (k_t, l_t).[5] Then, it can be shown that the conditional density of $f\left(y_t|k_t, l_t, \omega_t\right)$, as well as $g\left(z_t|\omega_t, k_t, l_t\right)$ and $g\left(\omega_t|x_t, k_t, l_t\right)$, are identified through the following equation based on the observed conditional density of $f\left(y_t, z_t|x_t, k_t, l_t\right)$ [6]

$$\begin{aligned} & f\left(y_t, z_t|x_t, k_t, l_t\right) \\ = & \int_{-\infty}^{\infty} f\left(y_t|k_t, l_t, \omega_t\right) g\left(z_t|\omega_t, k_t, l_t\right) h\left(\omega_t|x_t, k_t, l_t\right) d\omega_t \end{aligned}$$

Furthermore, the structural parameters (β_l, β_k) are identified given that $f\left(y_t|k_t, l_t, \omega_t\right)$ is identified.

We note that the above two assumptions put much less restriction on y_t than on z_t and x_t. The second condition normally requires z_t and x_t to be continuous and have no point

[4] y_t may be simply the output, but as we will see it can also be some other dependent variable.

[5] As we will discuss below, some additional technical assumptions are needed to prove the identification results, but these two conditions are the most substantive assumptions we need. A conditional density function $g\left(z|\omega\right)$ is injective if the integral operator defined by it, $L_g\left(h\left(.\right)\right) = \int g\left(z|\omega\right) h\left(z\right) dz$, is invertible.

[6] The equation is a result of the total law of probability and the first condition.

mass, as ω_t is normally modeled as continuous and having no point mass. Yet, y_t can even be binary as long as the independence conditions are satisfied. Meanwhile, all three latent conditional densities are identified. Thus to get the full potential of the identification idea, the production output does not have to be put as the "main" dependent variable y_t. We can either put the output as y_t, *or* we can treat it as a "proxy variable" (i.e. x_t or z_t) since it satisfies the second (injectivity) condition automatically by its specification. This observation is useful when we have to use investment i_t in our identification, which often has large point mass at zero and is not easily verified for the second condition even if we are willing to throw out all the observations with zero investment. As will become more clear later, this feature makes our identification idea more generally applicable.

There are a few complications in applying the above identification idea to the estimation of production functions. First, we need think about how to carefully choose (z_t, x_t) given industry background knowledge and the underlying structural framework as described in Olley and Pakes (1996), such that the above two identification conditions can be satisfied. Second, normally, there are no direct crude measures of the productivity term ω_t. The relation between ω_t and proxy variables, such as intermediate inputs and investment, depends on (k_t, l_t). This creates a collinearity problem in the identification of the structural parameters. We need to find a way to avoid the collinearity problem. In the rest of this section, we first specify our econometric model and then discuss in greater detail the identification of our model and our approach to dealing with the above complexities.

3.1 Model

We assume that the underlying structural model is the same as described by Olley and Pakes (1996). The notation is the same as in our literature review. We assume the following Cobb-Douglas value-added production function:

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t$$

where $\eta_t \perp (l_t, k_t, \omega_t)$, and $E(\eta_t | l_t, k_t, \omega_t) = 0$. The functional form assumption is made here for the ease of demonstration. The identification result applies equally well to other forms of production functions. Our interest here is to identify and estimate (β_l, β_k) given that ω_t is correlated with (l_t, k_t), but is not observed by the econometrician. For the productivity ω_t, let $E(\omega_t | \mathcal{I}_{t-1})$ be the prediction of ω_t based on the information, $\mathcal{I}_{t-1}$, available in period $t-1$, and $\xi_t = \omega_t - E(\omega_t | \mathcal{I}_{t-1})$ is the prediction error. In the following, we will assume that ω_t follows an exogenous first-order Markov process, such that $E(\omega_t | \mathcal{I}_{t-1}) = E(\omega_t | \omega_{t-1})$.

We define $\phi(\omega_{t-1}) \equiv E(\omega_t|\omega_{t-1})$. Later, we will demonstrate that the more general case of ω_t following a controlled Markov process can be similarly treated as long as the control variable is observed.

The timing assumptions about the input decisions determine the appropriate arguments to be included in the input demand functions. In applications, these assumptions should be made to match the specific industries under analysis. To compare our estimates to those obtained under previous methods, we assume that decisions about intermediate inputs are made after observing the contemporaneous capital and labor input and productivity. And we assume that the capital and labor input of period t are determined in period $t-1$, without observing the period-t innovation, ξ_t, of productivity.[7]

As an important extension of the literature, we model the intermediate input demand more generally as follows:

$$m_t = \mu_t(k_t, l_t, \omega_t) + \epsilon_t$$

where $\mu_t(\omega_t, l_t, k_t)$ is the "theoretical" input demand function that is strictly increasing in ω_t for any given (k_t, l_t), and ϵ_t is the residual error. We assume $\epsilon_t \perp (l_t, k_t, \omega_t)$. In the following, we will simply call $\mu_t(\omega_t, l_t, k_t)$ an input demand function, keeping in our mind that there is a deviation in the actual demand from what is predicted by its demand function.

Similarly, we specify the demand of electricity (or any other intermediate input) in the following way:

$$u_t = \psi(k_t, l_t, \omega_t) + v_t$$

where, similarly, $\psi(\omega_t, l_t, k_t)$ is the "theoretical" electricity demand function that is strictly increasing in ω_t for any given (k_t, l_t), and we assume $v_t \perp (l_t, k_t, \omega_t)$.

Given the Cobb-Douglas production function and the timing assumption, labor demand has the following linear form:

$$l_t = \alpha_0 + \alpha_1 k_t + \alpha_2 \omega_{t-1} + \varepsilon_t$$

We assume $\varepsilon_t \perp (k_t, \omega_{t-1})$.

The data-generating process for the investment I_t is somewhat different from that of the above static inputs. We often observe a significant portion of the firms in the data making

[7]The case of l_t being determined after observing ξ_t will be discussed later.

no investment in physical capital in some periods. To account for the fact that there are a lot of zero observations for investment, we model investment as a censored variable as follows:

$$\begin{aligned} I_t^* &= \iota_t(\omega_t, k_t) + \zeta_t \\ I_t &= I_t^* \times 1(I_t^* \geq 0) \end{aligned}$$

where I_t is the observed investment, I_t^* is a latent index variable, and $\iota_t(\omega_t, k_t)$ is the "theoretical" investment function that is strictly increasing in ω_t for any given k_t. We assume $\zeta_t \perp (k_t, \omega_t)$. The observed investment data are censored at zero.

To complete the model, we assume that capital is accumulated according to the following equation:

$$K_t = (1 - \delta) K_{t-1} + I_{t-1}$$

where δ is the depreciation rate. And we assume that $(\eta_t, \xi_t, \epsilon_t, v_t, \zeta_t)$ are mutually independent.

3.2 Identification

We note that the equations of the dependent variables—i.e. (y_t, m_t, u_t, I_t)—are very similar in that they all depend on the unobserved productivity ω_t, similar control variables (k_t, l_t) or just k_t, and some error terms $\eta_t, \epsilon_t, v_t, \zeta_t$. Our argument about the identification of the model below includes the production function equation of y_t and the equations of two other dependent variables. In the following, we base our identification discussion on the current value-added output y_t and two intermediate inputs, m_{t-1} and u_{t-1}, from the previous period. However, similar arguments can be made about y_t, i_{t-1} and an intermediate input variable. We use (m_{t-1}, u_{t-1}) instead of (m_t, u_t) in the identification to avoid the collinearity problem we alluded to before. This point will become clear after we give the identification equation. Now we begin our identification argument by listing the conditions we need to prove identification.

Condition 1 *(Conditional Independence)* $f(y_t|m_{t-1}, u_{t-1}, \omega_{t-1}, l_t, l_{t-1}, k_t, k_{t-1}) =$ $f(y_t|\omega_{t-1}, l_t, l_{t-1}, k_t, k_{t-1})$, *and* $g(m_{t-1}|u_{t-1}, \omega_{t-1}, l_t, l_{t-1}, k_t, k_{t-1}) = g(m_{t-1}|\omega_{t-1}, l_t, l_{t-1}, k_t, k_{t-1})$, *for all* $(l_t, l_{t-1}, k_t, k_{t-1})$, *where* f *and* g *are conditional density functions.*

Condition 2 *(injectivity) i) characteristic functions of* g *and* h *do not vanish on the real line; ii)* $\mu_t(k_t, l_t, \omega_t)$ *and* $\psi_t(k_t, l_t, \omega_t)$ *are monotonic in* ω_t *for any given* (k_t, l_t).

The first equality in condition 1 states that m_{t-1} and u_{t-1} do not provide information about y_t beyond what is already contained in ω_{t-1}. Condition 2 guarantees that the integral operators defined by $g\left(m_{t-1}|\omega_{t-1},k_{t-1},l_{t-1}\right)$ and $h\left(u_t|\omega_{t-1},k_{t-1},l_{t-1}\right)$ are invertible. These assumptions make m_{t-1} and u_{t-1} two potential proxy variables for the unobserved ω_{t-1}. The second equality in condition 1 says that the two proxy variables are independent of each other, conditional on ω_{t-1} and other control variables. This condition and the injectivity condition make one proxy variable a valid instrument for the error in the other proxy variable. We also need the following two technical conditions for our identification.

Condition 3 *(distinctive eigenvalues) for any given (k_t,l_t) and any $\overline{\omega}_t \neq \widetilde{\omega}_t$, there exists a set A such that $g\left(y_t|\overline{\omega}_{t-1},l_t,k_t\right) \neq g\left(y_t|\widetilde{\omega}_{t-1},l_t,k_t\right)$ for all $y_t \in A$ and* $\Pr(A) > 0$.

Condition 4 *(normalization) $E\left(m_t|\mu_t\right) = \mu_t$, that is, $E\left(\epsilon_t|\mu_t\right) = 0$.*

Condition 3 guarantees that we can always find distinctive eigenvalues and, consequently, different eigenfunctions, in the spectral decomposition that we employ in the proof of our identification. And condition 4 will be used to pin down the eigenfunctions for each given ω_{t-1}.[8] We will discuss the substantive implications of these assumptions later.

Given condition 1 and by the law of total probability, we have

$$
\begin{aligned}
& f\left(y_t,u_{t-1},m_{t-1}|l_t,l_{t-1},k_t,k_{t-1}\right) \\
= & \int f\left(y_t|u_{t-1},m_{t-1},\omega_{t-1},l_t,l_{t-1},k_t,k_{t-1}\right) g\left(m_{t-1}|u_{t-1},\omega_{t-1},l_t,l_{t-1},k_t,k_{t-1}\right) \\
& h\left(\omega_{t-1},u_{t-1}|l_t,l_{t-1},k_t,k_{t-1}\right) d\omega_{t-1} \\
= & \int f\left(y_t|\omega_{t-1},l_t,k_t\right) g\left(m_{t-1}|\omega_{t-1},l_{t-1},k_{t-1}\right) h\left(\omega_{t-1},u_{t-1}|l_t,l_{t-1},k_t,k_{t-1}\right) d\omega_{t-1}
\end{aligned}
$$

The second equality follows from the conditional independence condition and the model specification. Rewriting the above equation, we have

$$
\begin{aligned}
& f\left(y_t,u_{t-1},m_{t-1}|l_t,l_{t-1},k_t,k_{t-1}\right) \qquad (1)\\
= & \int f\left(y_t|\omega_{t-1},l_t,k_t\right) g\left(m_{t-1}|\omega_{t-1},l_{t-1},k_{t-1}\right) h\left(\omega_{t-1},u_{t-1}|l_t,l_{t-1},k_t,k_{t-1}\right) d\omega_{t-1}
\end{aligned}
$$

Intuitively speaking, m_{t-1} works as a crude proxy for ω_{t-1}, and u_{t-1} works as an instrument for the error in m_{t-1}. Note that we use (m_{t-1},u_{t-1}), instead of (m_t,u_t), as the proxy

[8] In fact, any known functional of $f\left(m_t|\mu_t\left(k_t,l_t,\omega_t\right)\right)$, such as median and known quantiles, works here. The mean function is a natural choice here, as a constant term can always be included in the function of $\mu_t\left(k_t,l_t,\omega_t\right)$.

variables. If we use (m_t, u_t) as the proxy variables, the control variables in the conditioning set would be same—i.e. $((k_t,\ l_t)$—for y_t and m_t. This creates an identification problem, because we then cannot both nonparametrically identify the conditional density of m_t and the structural parameters in the production function. Using (m_{t-1}, u_{t-1}) as the proxy variables gets us around this problem. Now, given equation (1), the identification question is whether we can identify the latent densities, especially $f(y_t|\omega_{t-1}, l_t, k_t)$, given the observed density of $f(y_t, u_{t-1}, m_{t-1}|l_t, l_{t-1}, k_t, k_{t-1})$.

Given the conditions above, Theorem 1 in Hu and Schennach (2008) (p. 202) can be applied to show that the latent densities $f(y_t|\omega_{t-1}, l_t, k_t)$, $g(m_{t-1}|\omega_{t-1}, k_{t-1}, l_{t-1})$, and $h(\omega_{t-1}, u_{t-1}|l_t, l_{t-1}, k_t, k_{t-1})$ are identified.[9] In the following, we sketch the main idea of the proof of the identification to help make the key identification sources more transparent. We will omit the control variables of k and l for notational simplicity. First, we define an integral operator based on a conditional density.

Definition 1 *Let $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Z})$ be spaces of functions defined on the domains of $\mathcal{X}$ and $\mathcal{Z}$ respectively. Then, the integral operator $L_{x|z}$ is defined as*

$$\left[L_{x|z}g\right](x) = \int_{\mathcal{Z}} f(x|z)\, g(z)\, dz$$

where the operator $L_{x|z}$ maps a function $g(z)$ in $\mathcal{F}(\mathcal{Z})$ into a function in $\mathcal{F}(\mathcal{X})$.

Then equation (1) now can be equivalently written in corresponding integral operators as

$$L_{y;m|u} = L_{m|\omega}\Delta_{y;\omega}L_{\omega|u} \tag{2}$$

where $L_{y;m|u}$ is defined similarly to $L_{m|u}$ with $f(m|u)$ replaced by $f(y, m|u)$ for a given y, and where $\Delta_{y;\omega}$ is a "diagonal operator" mapping a function $h(\omega)$ to $f(y|\omega)\, h(\omega)$. Meanwhile, by integrating both sides of equation (2) over y, we get $L_{m|u} = L_{m|\omega}L_{\omega|u}$, which is equivalent to

$$L_{\omega|u} = L_{m|\omega}^{-1}L_{m|u}$$

Now, we substitute the above expression of $L_{\omega|u}$ into (2) and rearrange the operators based on observable densities to the left-hand side, and we get

$$L_{y;m|u}L_{m|u}^{-1} = L_{m|\omega}\Delta_{y;\omega}L_{m|\omega}^{-1} \tag{3}$$

[9] Hu and Schennach's theorem is stated without control variables. We can define, for example, $\tilde{y}_t \equiv y_t - \beta_0 + \beta_l l_t + \beta_k k_t$, such that their identification results can be applied directly given that (β_l, β_k) are identified from the variation in the data.

The inverse of $L_{m|u}$ used in the above equation can be shown to exist by using the second condition in condition 2. Equation (3) means that $L_{y;m|u}L_{m|u}^{-1}$ admits an eigenvalue-eigenfunction decomposition. The left-hand-side operator based on observed conditional densities is decomposed to obtain $g(m|\omega,.)$, and $f(y|\omega,.)$, the latent conditional densities of interest. Theorem XV.4.5 in Dunford and Schwartz (1971) can be used to show that the decomposition is unique given that the operators are defined with density functions. Lastly, conditions 3 and 4 are employed to ensure the uniqueness of the ordering and indexing of the eigenvalue and eigenfunctions.

The independence and injectivity conditions are implicitly applied in the above identification argument. The independence assumptions play two roles in the identification. First, it helps reduce the dimensionality of the latent conditional densities to make the spectral decomposition possible. Second, it makes one proxy variable a potential instrument for the measurement error of the other proxy variable. The injectivity assumptions make sure that the integration operators are invertible. This role played by the injectivity condition bears some similarity to that of the rank conditions for the instrumental variable method in the classical linear regression models.

We summarize the above identification results in the following Lemma.

Lemma 1 *Under conditions 1, 2, 3, 4, the observed density of* $f(y_t, u_{t-1}, m_{t-1}|l_t, k_t, k_{t-1}, l_{t-1})$ *uniquely determines the latent conditional densities of* $f(y_t|\omega_{t-1}, l_t, k_t)$, $g(m_{t-1}|\omega_{t-1}, k_{t-1}, l_{t-1})$ *and* $h(\omega_{t-1}, u_{t-1}|l_t, l_{t-1}, k_t, k_{t-1})$.

Given the identification of the conditional densities and the assumptions of $\xi_t + \eta_t \perp (k_t, l_t, \omega_{t-1})$ and $E(\xi_t + \eta_t) = 0$, the density of $f_{\xi_t+\eta_t}$ and the production function are identified nonparametrically given enough variation in (k_t, l_t). Thus, for the Cobb-Douglas production function, the structural parameters of interest, (β_l, β_k), are identified. We summarize the identification results in the following Theorem.

Theorem 5 *Under conditions 1, 2, 3 and 4, the observed density* $f(y_t, u_{t-1}, m_{t-1}|l_t, k_t, k_{t-1}, l_{t-1})$ *uniquely determines* (β_l, β_k)*, together with* $f_{\xi+\eta}$*,* g_ϵ *and* h *from the following equation:*

$$\begin{aligned} & f(y_t, u_{t-1}, m_{t-1}|l_t, l_{t-1}, k_t, k_{t-1}) \qquad (4) \\ = & \int_{-\infty}^{\infty} f_{\xi+\eta}(y_t - (\beta_0 + \beta_l l_t + \beta_k k_t + E(\omega_t|\omega_{t-1}))|l_t, k_t, \omega_{t-1}) \\ & \times g_\epsilon(m_{t-1} - \mu_t(\omega_{t-1}, k_{t-1}, l_{t-1})|\omega_{t-1}, k_{t-1}, l_{t-1}) \times h(\omega_{t-1}, m_{t-1}|k_{t-1}, l_{t-1}, l_t, k_t)\, d\omega_{t-1} \end{aligned}$$

We note that similar identification arguments can also be made with (y_t, m_{t-1}, u_{t-1}) replaced by (y_t, m_{t-1}, i_{t-1}).[10] As investment can also be used as one of the proxy variables if we keep only the part of sample with positive investment, we can derive the same identification result by using investment and one of the intermediate inputs as the two proxy variables. So, we can similarly identify the production function by using the following equation:

$$\begin{aligned} & f(y_t, m_{t-1}, i_{t-1}|l_t, l_{t-1}, k_t, k_{t-1}, (I_{t-1}>0)) \qquad (5) \\ = & \int_{-\infty}^{\infty} f_{\xi+\eta}(y_t-(\beta_0+\beta_l l_t+\beta_k k_t+E(\omega_t|\omega_{t-1}))|l_t, k_t, \omega_{t-1})\times \\ & g_\epsilon(m_{t-1}-\mu_t(\omega_{t-1}, k_{t-1}, l_{t-1})|\omega_{t-1}, k_{t-1}, l_{t-1})\times \\ & h\left(\omega_{t-1}, i_{t-1}|k_{t-1}, l_{t-1}, l_t, k_t, (I_{t-1}>0)\right) d\omega_{t-1} \end{aligned}$$

We lose the part of sample with zero investment by using the above equation. Yet, as we have noted, the output y_t satisfies the injectivity assumption automatically without having to throw out any observations. So we can switch the roles played by the three dependent variables to avoid throwing out a lot of observations. Specifically, we can use I_{t-1} as the variable that satisfies only the conditional independence assumption and use y_t and one intermediate input as the proxy variables that satisfy both the independence and injectivity assumptions. So now, we can base our identification on the following equation:

$$\begin{aligned} & f\left(y_t, m_{t-1}, I_{t-1}|l_t, l_{t-1}, k_t, k_{t-1}\right) \qquad (6) \\ = & \int_{-\infty}^{\infty} f_{\xi+\eta}(y_t-(\beta_0+\beta_l l_t+\beta_k k_t+E(\omega_t|\omega_{t-1}))|l_t, k_t, \omega_{t-1})\times \\ & g_\epsilon\left(m_{t-1}-\mu_t(\omega_{t-1}, k_{t-1}, l_{t-1})|\omega_{t-1}, k_{t-1}, l_{t-1}\right)\times \\ & h\left(\omega_{t-1}, I_{t-1}|k_{t-1}, l_{t-1}, l_t, k_t\right) d\omega_{t-1} \end{aligned}$$

As we discuss below, the above two alternative equations for identification become useful when the identification assumptions are more likely to hold for (y_t, m_{t-1}, i_{t-1}) than for (y_t, m_{t-1}, u_{t-1}) in some cases. The trade-off here is that the computational burden in the estimation may be heavier when we use equation (6), whereas only the observations with positive investment can be used if we use equation (5).

[10] The same observation is also true with (y_t, m_{t-1}, u_{t-1}) replace with (y_t, u_{t-1}, i_{t-1}).

3.3 Extension

As an important extension of the above model, we can allow ω_t to be endogenously determined. This extension is important for applications in which it is essential that firms are assumed to actively spend resources to improve productivity. Our model can very conveniently accommodate the case of productivity following a controlled first-order Markov process. Specifically, suppose that the control variable affecting the process of ω_t is determined in the following way:

$$r_t = R(k_t, \omega_t) + \varrho_t$$

where r_t is the R&D spending in period t (or some other control variable affecting the evolution of productivity), and ϱ_t captures other unobserved factors affecting r_t. Under the alternative assumption, we have $E(\omega_t|\mathcal{I}_{t-1}) = E(\omega_t|\omega_{t-1}, r_{t-1})$. Given that R&D spending is observed, our identification arguments above can be largely replicated as long as we replace the term of $E(\omega_t|\omega_{t-1})$ in the output equation with $E(\omega_t|\omega_{t-1}, r_{t-1})$.

Doraszelski and Jaumandreu (2008) extend LP's method by modeling productivity as a controlled Markov Process. Their interest is in estimating the impact of R&D on productivity in the Spanish manufacturing sector. Obviously, the assumption of ω_t following an exogenous Markov Process is conceptually inconsistent with their goal. To estimate such a model, they assume a parametric form for the labor demand function to avoid the identification issues associated with the nonparametric input demand functions. Similar to LP, their method also relies on the assumption that the only unobserved determinant of labor input is productivity. The identification issue they avoid by assuming a parametric labor demand function also has its root in the implicit assumption that other unobserved factors, such as cost shocks, do not affect investment and input decisions. We do not need parametric assumptions for input demand functions to allow for endogenous productivity, because we explicitly allow other unobserved shocks to affect investment and input decisions, and these random shocks break up the collinearity problem when the demand functions of inputs used as proxy variables are estimated nonparametrically.

3.4 Discussion

Although we have shown that the identification of production function can be achieved even if we allow additional unobservables to show up in the proxy variable equations, the validity of the underlying conditions still needs to be assessed carefully in order to verify the applicability of the above identification results for estimation. In the following, we discuss the key conditions in turn, assuming that the underlying the structural framework is the same as described in Olley and Pakes (1996).

3.4.1 The conditional independence assumption

The conditional independence assumption can be equivalently stated through the residuals in the corresponding equations. For example, the assumption of mutual independence among y_t, m_{t-1}, u_{t-1} is equivalent to the assumption of mutual independence among the corresponding residuals—i.e. $\eta_t, \epsilon_{t-1}, v_{t-1}$. Whether it is reasonable to assume that the three residual terms are mutually independent depends on what are captured by them. Conceptually, the residual of the output equation, η_t, often reflect measurement errors of the output and/or unanticipated technology shocks, such as the number of defective products and machine breakdowns. The errors associated with intermediate inputs and investment could be results of measurement errors, optimization errors, or idiosyncratic cost shocks. To assess the validity of the independence assumption, we need to carefully assess the following two questions: 1) What unobserved factors are being captured in the residuals? 2) Whether the unobserved factors can be reasonably assumed independent of each other? The residuals under consideration here are $\eta_t, \epsilon_{t-1}, v_{t-1}$ (and ζ_{t-1}), which are, respectively, the residuals in the equations of y_t, m_{t-1}, u_{t-1} (and I_{t-1}). The purpose of our discussion below is to find some guidelines for picking out the two proxy variables that, together with y_t, satisfy the independence assumption. In the following, we discuss the independence assumption for each of three common types of unobserved factors.

Measurement error Measurement errors in the proxy variables are often the top concern for researchers applying OP,LP and ACF's methods. If only measurement errors are involved, we expect that the residual in the equation of a proxy variable contains only its own measurement error. Suppose that the measurement errors are simply recording errors. Then, it seems reasonable to assume that the measurement error of one proxy variable is conditionally independent of that of another proxy variable. However, measurement errors due to other reasons may need to be analyzed more carefully. For example, as LP point out, some intermediate inputs —such as materials and fuels—may be storable, and it could become a cause of measurement errors if the econometrician can only observe the new purchase of such inputs instead of the actual usage of them. The unobserved inventory changes make the actual inputs differ from what are observed in the data. If, furthermore, the inventory decisions of the two proxy variables are correlated, then the measurement errors of the two proxy variables would also be correlated. In this case, using material and fuel, for example, as the two proxy variables would be problematic. However, as electricity normally cannot be stored, we may use the material and electricity inputs as the two proxy variables in the estimation. Hence, to assess the independence assumption for the measurement error case in application, we need to carefully consider the main causes of such errors. The independence assumption could be satisfied if we carefully select the two proxy variables.

Optimization error In the model, we assumed that the decisions on intermediate inputs, m_{t-1} and u_{t-1}, are made simultaneously after observing $(k_{t-1}, l_{t-1}, \omega_{t-1})$. With optimization errors, the timing assumption implies, for example, that if more than the optimal material input were used, the extra material would not be complemented by additional energy input. Thus, under this timing assumption, it seems reasonable to assume that the residual of each intermediate input contains only its own optimization error and that the residuals are independent. Now suppose that the intermediate inputs are determined sequentially. If the decision sequence is known to the econometrician, the independence assumption seems reasonable as we can modify the model by including the input determined first as a control variable in the equation of the input determined later. However, if the decision timings are unknown, the independence assumption would be violated in a misspecified model, because now the optimization error of one input can be captured by the residual of the other input. Lastly, as we expect no interaction between the static input decision and the dynamic investment decision, it should be reasonable to assume that the optimization error in investment is independent of the two input optimization errors. So, if the residuals capture only optimization errors, we expect the independence assumption to be satisfied for both (m_{t-1}, u_{t-1}) and (m_{t-1}, i_{t-1}), with the independence assumption for (m_{t-1}, i_{t-1}) more robust to alternative timing assumptions on input decisions.

Idiosyncratic cost shocks The literature has been assuming a competitive market for intermediate inputs and capital such that the costs in each period are the same for all the firms. Now suppose that there are idiosyncratic cost shocks. Given that firms observe the cost shocks while making simultaneous decisions on the two inputs, we expect the residuals of the two inputs to capture the cost shocks of both inputs. Thus, the independence assumption would be violated for (m_{t-1}, u_{t-1}). However, suppose that the cost shocks of the intermediate inputs are independent across time. In this case, the investment decision should not be affected by the cost shocks to static inputs. So, the independence assumption would still hold for (m_{t-1}, i_{t-1}) and (u_{t-1}, i_{t-1}). Finally, if cost shocks of the inputs are serially correlated, the current cost shocks would contain information about future input cost shocks and, consequently, future returns to the current investment. Then, the investment would respond to the cost shocks to the inputs, and the residual in the investment equation would capture the cost shocks of both the inputs and the investment. Hence, if the input cost shocks are serially correlated, the independence assumption would most likely be violated.[11]

To finish our assessment of the independence assumption, we still need to evaluate the relation between the residual of the output equation, η_t, and that of the two proxy variables. Given that the output errors are normally results of measurement errors and technological shocks, it seems reasonable to assume that the error of the output is independent of the errors of the intermediate inputs and investment for all the three types of errors above. However, it is worth noting that the errors of the *contemporary* intermediate inputs could be captured in the residual of the output equation if the output is measured by value-added. This creates no problem if the errors of the intermediate inputs are independent across time. Otherwise, one can get around the problem by estimating the production function for the gross output instead of value-added.

In summary, the independence assumption seems reasonable in many important cases. In some cases, carefully chosen proxy variables for estimation would make the independence assumption more likely to be satisfied. For each specific application, special attention should be given to the interpretations of the residual terms when assessing the validity of the assumption. We summarize the above discussion in Table 1.

Table 1: Assessment of the validity of the independence assumption

	1: Likely valid; 0: Unlikely valid		
Error types	(y_t, m_{t-1}, u_{t-1})	(y_t, m_{t-1}, i_{t-1})	(y_t, u_{t-1}, i_{t-1})
Measurement errors	1	1	1
Optimization errors, simultaneous	1	1	1
Optimization errors, sequential	0	1	1
Cost shocks, independent over time	0	1	1
Cost shocks, serially correlated	0	0	0

[11] In the extreme case of firm-specific cost shocks being constant over time, the methods of OP, LP and ACF would work as the cost shocks can be combined into the firm-specific productivity term.

3.4.2 The injectivity assumption

We also require that the demand function (the certainty part) of the inputs and investment to be strictly increasing in productivity in order to satisfy the injectivity condition. In comparison to the method of OP/LP/ACF, we do not need the inputs/investment to be a deterministic function of productivity. This generalization should also make the monotonicity assumption easier to satisfy. As LP point out, the primitive conditions ensuring the monotonicity condition are easier to obtain for the demand of intermediate inputs than for investment. This observation suggest that intermediate inputs can be preferable candidates for the two proxy variables in our estimation.

3.4.3 The distinctive eigenvalues and the normalization

The distinctive eigenvalue condition is a very weak assumption, which should be easily satisfied in practice. The normalization assumption of $E(m_t|\mu_t) = \mu_t$ does not add additional restrictions given the assumption of additive errors in the intermediate input demand functions.

In the next section, we propose an estimation method based on our identification strategy.

4 Estimation

In the estimation, we treat each firm as an observation, and the data as i.i.d across firms. A complete specification of the likelihood for each firm would be very complicated, especially for longer panels. The likelihood of the observation of a firm would involve, for example, the conditional density of the firm's last period data given its data in all previous periods. Specifying such complete models requires many additional assumptions, which are undesirable and are unnecessary for estimating the structural parameters of interest here. In our case, the structural parameters in the production functions are identified with the partial conditional likelihood, which involves only two periods' data . Thus we adopt the partial likelihood framework for our estimation (c.f. Wooldridge (2002)).

We follow Gallant and Nychka (1987) by approximating the conditional densities with Hermite series. More specifically, to compute the likelihood function, we approximate a unknown density function $h(u)$, for example, by using Hermite series:

$$h(u) \approx P_k^2(u)\,\phi(u)$$

where $P_k(u)$ is a kth-order polynomial of u, and $\phi(u)$ is the density of the standard normal distribution. The advantage of using this approximation method for estimating our model is threefold. First, the Hermite series automatically impose some smoothness requirements for the approximated density. This avoids having to add a penalty function that punishes nonsmoothness in the density functions, which is necessary, for example, when one uses a polynomial to approximate the densities. Second, the approximation method, in principle, automatically guarantees that the approximated density is nonnegative at all values of the parameters. This avoids having to impose restrictions in estimation that the approximated density has to be nonnegative. Lastly, the likelihood function in our method involves integration of a product of multiple conditional densities over the support of the latent variable. With the approximation by Hermite series, the integration is very convenient to compute and has an analytical result.

For estimation, we first spell out the observed density, $f(y_t, m_{t-1}, u_{t-1}, l_t, k_t|l_{t-1}, k_{t-1})$, as a mixture of the product of several latent conditional densities as follows:

$$\begin{aligned}
& f(y_t, m_{t-1}, u_{t-1}, l_t, k_t|l_{t-1}, k_{t-1}) \\
= & \int g_y(y_t|l_t, k_t, \omega_{t-1}, m_{t-1}, u_{t-1}, l_{t-1}, k_{t-1})\, g_l(l_t|k_t, \omega_{t-1}, m_{t-1}, u_{t-1}, k_{t-1}, l_{t-1}) \\
& g_k(k_t|\omega_{t-1}, m_{t-1}, u_{t-1}, l_{t-1}, k_{t-1})\, g_m(m_{t-1}|\omega_{t-1}, u_{t-1}, l_{t-1}, k_{t-1})\, g_u(u_{t-1}|\omega_{t-1}, l_{t-1}, k_{t-1}) \\
& g_\omega(\omega_{t-1}|l_{t-1}, k_{t-1})\, d\omega_{t-1} \\
= & \int g_y(y_t|l_t, k_t, \omega_{t-1})\, g_l(l_t|k_t, \omega_{t-1})\, g_k(k_t|\omega_{t-1}, k_{t-1}) \\
& g_m(m_{t-1}|\omega_{t-1}, l_{t-1}, k_{t-1})\, g_u(u_{t-1}|\omega_{t-1}, l_{t-1}, k_{t-1})\, g_\omega(\omega_{t-1}|l_{t-1}, k_{t-1})\, d\omega_{t-1}
\end{aligned}$$

The first equality above follows by the total law of probability; the second equality follows from the conditional independence assumption and the fact that the variables in period $t-1$ are independent of the period-t innovation in the latent productivity. Thus we can estimate the model using Semi-Nonparametric Maximum Likelihood estimation (SNPMLE) method as follows

$$
\begin{aligned}
& (\hat{\beta}, \hat{\alpha}, \hat{\phi}, \hat{\mu}, \hat{\psi}, \hat{g}_{\xi+\eta}, \hat{g}_{\varepsilon}, \hat{g}_k, \hat{g}_{\epsilon}, \hat{g}_v, \hat{g}_{\omega}) \\
= \; & \arg \max_{\beta,\alpha,\phi,\mu,\psi,\left(g_{\xi+\eta}, g_{\varepsilon}, g_k, g_{\epsilon}, g_v, g_{\omega}\right)\in\mathcal{A}_n} \frac{1}{J}\sum_{j=1}^{J}\sum_{t=1}^{T}\ln\int_{-\infty}^{\infty} g_{\xi+\eta}\left(y_{jt} - \left(\beta_l l_{jt} + \beta_k k_{jt} + \phi\left(\omega_{j,t-1}\right)\right)\right) \\
& g_{\varepsilon}\left(l_{jt} - \alpha_0 - \alpha_1 k_{jt} - \alpha_2 \omega_{j,t-1}\right) g_k\left(k_{jt}|k_{j,t-1}, \omega_{j,t-1}\right) \qquad (7) \\
& g_{\epsilon}\left(m_{j,t-1} - \mu\left(k_{j,t-1}, l_{j,t-1}, \omega_{j,t-1}\right)\right) g_v\left(u_{j,t-1} - \psi\left(k_{j,t-1}, l_{j,t-1}, \omega_{j,t-1}\right)\right) \\
& g_{\omega}\left(\omega_{j,t-1}|l_{j,t-1}, k_{j,t-1}\right) d\omega_{j,t-1}.
\end{aligned}
$$

where $\mathcal{A}_n$ is the set of all nth-order Hermite series that integrate to one.[12] Note that the sum of per-period likelihoods over t for each firm j is not the likelihood of the observation of firm j.

A few things worth pointing out about the above estimation method. First, the above likelihood is analytical in the parameters as long as we use polynomials to approximate the unknown functions of μ, ψ and ϕ. If these unknown functions are linear in $\omega_{j,t-1}$, the likelihood can be easily computed using the values of the moments of normal distributions. When the unknown functions involves higher order terms of $\omega_{j,t-1}$, the likelihood is still analytical in the parameters, though the coefficients of which involve the gamma functions evaluated at some constants. This is a very convenient property that greatly simplifies the computation of the likelihood.

Second, if one assumes Cobb-Douglas production function, the functions of μ and ψ would be linear in their arguments. Furthermore, if ω_t follows a $AR(1)$ process, then ϕ would also be a linear function. These assumptions would make the likelihood very easy to compute.

Third, we use partial likelihood method with constraints, which needs to be carefully accounted for when we compute the covariance matrix of the estimates. And, as we detail below, the recent numerical equivalence results by Ackerberg, Chen, and Hahn (2011) show that the asymptotic variances of our semi-nonparametric estimates of the structural parameters are no more difficult to compute than those of a parametric model.

Lastly, the above likelihood function is a mixture of latent conditional densities. It is well-

[12] For example, an nth-order approximation of g_{ε} would be given by

$$\widehat{g}_{\varepsilon}(x) = \left(\sum_{s=1}^{n} a_s x^s\right)^2 \exp\left(-\frac{x^2}{\sigma_{\varepsilon}^2}\right)$$

where $\left(\{a_s\}_{s=1}^{n}, \sigma_{\varepsilon}\right)$ are the parameters to be estimated.

known that such likelihood functions often have some local maximums. When computing the estimates, we look for the global maximum of the likelihood function by repeating the optimization routine many times, starting from different random initial values centered around a reasonable guess. As most of the structural parameters in the model can be interpreted as elasticities, they are normally relatively small numbers, and many of them should actually be roughly in $(0, 1)$. This makes finding the global maximum easier.

4.1 Using Alternative Proxy Variables

As we discussed above, in some cases, the residuals in the equations of m_{t-1} and u_{t-1} may capture the optimization errors/cost shocks of both m_{t-1} and u_{t-1}, but the errors may be independent across time. In this case, to satisfy the conditional independence assumption, we can replace one of the proxy variables (m_{t-1} and u_{t-1}) by investment i_{t-1} and use only the sample with positive investment for estimation. So we can use the decomposition of the following observed density for estimation:

$$
\begin{aligned}
& f\left(y_t, m_{t-1}, i_{t-1}, l_t, k_t | l_{t-1}, k_{t-1}\right) \\
= & \int g_y\left(y_t | l_t, k_t, \omega_{t-1}\right) g_l\left(l_t | k_t, l_{t-1}, \omega_{t-1}\right) g_k\left(k_t | i_{t-1}, k_{t-1}\right) \\
& g_i\left(i_{t-1} | k_{t-1}, \omega_{t-1}\right) g_m\left(m_{t-1} | k_{t-1}, l_{t-1}, \omega_{t-1}\right) g_\omega\left(\omega_{t-1} | l_{t-1}, k_{t-1}\right) d\omega_{t-1}
\end{aligned}
$$

In the above expression, the density of k_t does not involve the latent variable ω_{t-1}, and thus can be dropped in the estimation.[13]So, we can estimate the model using the method of SNPMLE as follows

$$
\begin{aligned}
& (\hat{\beta}, \hat{\alpha}, \hat{\phi}, \hat{\mu}, \hat{\iota}, \hat{g}_{\xi+\eta}, \hat{g}_\varepsilon, \hat{g}_\zeta, \hat{g}_\epsilon, \hat{g}_\omega) \\
= & \arg \max_{\beta,\alpha,\phi,\mu,\iota,\left(g_{\xi+\eta}, g_\varepsilon, g_\zeta, g_\epsilon, g_\omega\right) \in \mathcal{A}_n} \frac{1}{J} \sum_{j=1}^{J} \sum_{t=1}^{T} \ln \int_{-\infty}^{\infty} g_{\xi+\eta}\left(y_{jt} - \left(\beta_l l_{jt} + \beta_k k_{jt} + \phi\left(\omega_{j,t-1}\right)\right)\right) \\
& g_\varepsilon\left(l_{jt} - \alpha_0 - \alpha_1 k_{jt} - \alpha_2 \omega_{j,t-1}\right) g_\zeta\left(i_{jt-1} - \iota\left(\omega_{jt-1}, k_{jt-1}\right)\right) \\
& g_\epsilon\left(m_{j,t-1} - \mu\left(k_{j,t-1}, l_{j,t-1}, \omega_{j,t-1}\right)\right) g_\omega\left(\omega_{j,t-1} | l_{j,t-1}, k_{j,t-1}\right) d\omega_{j,t-1}.
\end{aligned}
$$

Finally, one might want to use the investment variable but have only a relatively small sample. To use the entire sample, as we have pointed out before, we may switch the roles of the three key dependent variables. We can let y_t and one of the intermediate inputs (m_{t-1} or u_{t-1}) be the two variables that satisfy the assumptions of both conditional independence

[13]So the likelihood is constructed based on $f\left(y_t, m_{t-1}, i_{t-1}, l_t, k_t | l_{t-1}, k_{t-1}\right) / g_k\left(k_t | k_{t-1}, i_{t-1}\right)$

and monotonicity, and let i_{t-1} be the variable that satisfy only the conditional independence assumption. So, for the entire sample, we can use the following decomposition:

$$\begin{aligned}
& f\left(y_t, m_{t-1}, i_{t-1}, l_t, k_t | l_{t-1}, k_{t-1}\right) \\
= & \int g_y\left(y_t | l_t, k_t, \omega_{t-1}\right) g_l\left(l_t | k_t, \omega_{t-1}, l_{t-1}\right) g_k\left(k_t | i_{t-1}, k_{t-1}\right) \\
& \left(\Pr\left(I_{t-1}=0 | k_{t-1}, \omega_{t-1}\right)\right)^{1\{I_{t-1}=0\}}\left(g_i\left(i_{t-1} | k_{t-1}, \omega_{t-1}\right)\right)^{1-1\{I_{t-1}=0\}} \\
& g_m\left(m_{t-1} | k_{t-1}, l_{t-1}, \omega_{t-1}\right) g_\omega\left(\omega_{t-1} | l_{t-1}, k_{t-1}\right) d\omega_{t-1}
\end{aligned}$$

Now the integration does not have an analytical result. We have to simulate the integration, for example, via importance sampling, and estimate the model using simulated partial likelihood method. Suppose that we draw ns_1 and ns_2 simulation draws from two independent standard normal distribution for ω_{t-1} and ζ_{t-1}; then we can estimate the parameters as follows:

$$\begin{aligned}
& (\hat{\beta}, \hat{\alpha}, \hat{\phi}, \hat{\mu}, \hat{\iota}, \hat{g}_{\xi+\eta}, \hat{g}_{\varepsilon}, \hat{g}_{\zeta}, \hat{g}_{\epsilon}, \hat{g}_{\omega}) \\
= & \arg \max_{\beta,\alpha,\phi,\mu,\iota,\left(g_{\xi+\eta},g_{\varepsilon},g_{\zeta},g_{\epsilon},g_{\omega}\right)\in\mathcal{A}_n} \frac{1}{J}\sum_{j=1}^{J}\sum_{t=1}^{T}\ln\{\frac{1}{ns_1}\Sigma_{s=1}^{ns_1} g_{\xi+\eta}\left(y_{jt}-\left(\beta_l l_{jt}+\beta_k k_{jt}+\phi\left(\omega_{s,t-1}\right)\right)\right) \\
& g_{\varepsilon}\left(l_{jt}-\alpha_0-\alpha_1 k_{jt}-\alpha_2\omega_{s,t-1}\right) \\
& \left[\frac{1}{ns_2}\Sigma_{s'=1}^{ns_2}\left[1\left\{\zeta_{s',t-1}\leq-\iota\left(\omega_{s,t-1},k_{j,t-1}\right)\right\} g_{\zeta}\left(\zeta_{s',t-1}\right)/h\left(\zeta_{s',t-1}\right)\right]\right]^{1\{I_{j,t-1}=0\}} \\
& g_{\zeta}\left(i_{j,t-1}-\iota\left(\omega_{s,t-1},k_{j,t-1}\right)\right)^{1-1\{I_{j,t-1}=0\}} \\
& g_{\epsilon}\left(m_{j,t-1}-\mu\left(\omega_{s,t-1},l_{j,t-1},k_{j,t-1}\right)\right) g_{\omega}\left(\omega_{s,t-1}|l_{j,t-1},k_{j,t-1}\right)/h\left(\omega_{s,t-1}\right)\}.
\end{aligned}$$

where $h\left(\omega_{s,t-1}\right)$ and $h\left(\zeta_{s',t-1}\right)$ are the density of the standard normal distribution.

4.2 Alternative Timing Assumption for the Labor-Input Decision

We assumed in the model that l_t is determined in period $t-1$ before the realization of ξ_t, and thus $E\left(\xi_t|l_t\right)=0$. It is a reasonable assumption, given that it often takes time to adjust the labor input. If l_t is determined in period t instead, then we have $Cov\left(l_t,\xi_t|k_t,\omega_{t-1}\right)\neq 0$. In such a case, we can use the assumption of $Cov\left(l_{t-1},\xi_t|k_t,\omega_{t-1}\right)=0$ to help identify the labor elasticity and achieve similar identification results. This has been the standard alternative identification condition used in the literature.

In the case of l_t being determined in period t after observing ω_t, the appropriate specification for the labor demand equation is:

$$l_t = \alpha_0 + \alpha_1 k_t + \alpha_2 \omega_t + \varepsilon_t$$

Then, the identification of $f\left(y_t | l_t, k_t, \omega_{t-1}\right)$ above does not lead to the identification of the coefficient of l_t, because l_t is now correlated with ξ_t. Using l_{t-1} as an instrument for l_t in the production function, a Full Information Maximum Likelihood estimator can be employed to consistently estimate the structural parameters.

To illustrate, the estimation now can be based on the following decomposition of the observed conditional density:

$$\begin{aligned}
& f\left(y_t, m_{t-1}, u_{t-1}, l_t, k_t | l_{t-1}, k_{t-1}\right) \\
= & \int f\left(y_t, l_t | k_t, \omega_{t-1}, l_{t-1}\right) g_k\left(k_t | \omega_{t-1}, k_{t-1}\right) \\
& g_m\left(m_{t-1} | \omega_{t-1}, l_{t-1}, k_{t-1}\right) g_u\left(u_{t-1} | \omega_{t-1}, l_{t-1}, k_{t-1}\right) g_\omega\left(\omega_{t-1} | k_{t-1}, l_{t-1}\right) d\omega_{t-1}
\end{aligned}$$

Then, the we can estimate the structural parameters using the SNP FIML method as follows:

$$\begin{aligned}
& (\hat{\beta}, \hat{\alpha}, \hat{\phi}, \hat{\mu}, \hat{\psi}, \hat{g}_{\xi+\eta,\xi+\varepsilon}, \hat{g}_k, \hat{g}_\epsilon, \hat{g}_v, \hat{g}_\omega) \\
= & \arg \max_{\beta,\alpha,\phi,\mu,\psi,\left(g_{\xi+\eta,\xi+\varepsilon}, g_k, g_\epsilon, g_v, g_\omega\right) \in \mathcal{A}_n} \frac{1}{J} \sum_{j=1}^{J} \sum_{t=1}^{T} \\
& \ln \int_{-\infty}^{\infty} g_{\xi+\eta,\xi+\varepsilon}\left(y_{jt} - \left(\beta_l l_{jt} + \beta_k k_{jt} + \phi\left(\omega_{j,t-1}\right)\right), l_{jt} - \alpha_0 - \alpha_1 k_{jt} - \alpha_2 \phi\left(\omega_{j,t-1}\right) - \alpha_3 l_{j,t-1}\right) \\
& g_k\left(k_{jt} | k_{j,t-1}, \omega_{j,t-1}\right) g_\epsilon\left(m_{j,t-1} - \mu\left(k_{j,t-1}, l_{j,t-1}, \omega_{j,t-1}\right)\right) \\
& g_v\left(u_{j,t-1} - \psi\left(k_{j,t-1}, l_{j,t-1}, \omega_{j,t-1}\right)\right) g_\omega\left(\omega_{j,t-1} | l_{j,t-1}, k_{j,t-1}\right) d\omega_{j,t-1}
\end{aligned}$$

where, similarly, the $g_{\xi+\eta,\xi+\varepsilon}, g_k, g_\epsilon, g_v, g_\omega$ are nth-order approximations of the densities using Hermite series. Here, as l_t is endogenous, we employ the joint distribution of $g_{\xi+\eta,\xi+\varepsilon}$ in the likelihood function, and use l_{t-1} as the variable that is affecting l_t, but excluded from the y_t equation.

5 Asymptotic Inference

In this section, we discuss how to estimate the covariance matrix of our estimator. As the densities in the likelihood function of (7) are nuisance parameters, we focus on how to estimate the covariance matrix of the structural parameters in the production function. The covariance matrix of the estimators of OP/LP/ACF are normally estimated through bootstrapping. We can similarly estimate the covariance matrix of the parameters via bootstrapping. Alternatively, we can employ the recent numerical equivalence result proved by Ackerberg, Chen, and Hahn (2011), with which computing the asymptotic variances of our SNPMLE estimator turns out to be as simple as computing the asymptotic variances of a parametric MLE estimator.

Suppose that we use second-order Hermite series to approximate the densities in the likelihood function of (7). Shen (1997)'s results imply that our estimator of the structural parameters are $\sqrt{n}$ consistent and asymptotically normal. These asymptotic properties are based on the assumption that both the sample size and the order of the approximating series go to infinity. Then, a consistent estimator can be derived for the asymptotic covariance matrix for structural parameters (see Appendix D in Ackerberg, Chen, and Hahn (2011) for details). Meanwhile, if one assumes that the functional forms of the true conditional densities are exactly second-order Hermite series, the model becomes a parametric model, and the asymptotic covariance matrix can be easily estimated using standard results for parametric MLE. Obviously, here, the asymptotic properties of parametric MLE are based on the assumption that only the sample size goes to infinity. A bit surprisingly, the results of Ackerberg, Chen, and Hahn (2011) show that the consistent estimator of the asymptotic covariance matrix for the semi-parametric model is numerically exactly the same as that for the fictitious parametric model, even though the estimated structural parameters in the two models have different limiting distributions. In short, for the given Hermite series used to approximate the densities, we can compute the asymptotic covariance matrix of the structural parameters as if the Hermite series are the known functional form for the densities.

In the following section, we first compare our method to the previous methods through a simple Monte Carlo exercise. Then, we apply all these methods to the same Chilean manufacturing census data that LP and ACF used for illustration.

6 Empirical Example

6.1 Monte Carlo Evidence

We simulate two random samples by using models with a Cobb-Douglas production function. The general set-up is the same as in the model we described above. Productivity is the single unobservable affecting the intermediate inputs and investment in the first sample, whereas additional additive errors are also affecting these variables in the second sample. The exact simulation set-up is described in the Appendix. We then apply the method of OLS, ACF and ours (HH) to estimate the structural parameters of the labor and capital elasticities. The estimation results are presented in Table 2 and Table 3.

Table 2: The Estimation Results with Perfect Proxy Variables, with Simulated Data

	Parameters (True Value)					
	Labor Elasticity (0.625)			Capital Elasticity (0.375)		
	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE
OLS	0.957	0.004	0.332	0.078	0.004	0.298
ACF	0.636	0.022	0.025	0.365	0.021	0.023
HH_Parametric	0.621	0.021	0.021	0.380	0.023	0.024
HH_SNP	0.620	0.021	0.021	0.381	0.023	0.024

Table 3: The Estimation Results with Mismeasured Proxy Variables, with Simulated Data

	Parameters (True Value)					
	Labor Elasticity (0.625)			Capital Elasticity (0.375)		
	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE
OLS	0.957	0.005	0.332	0.078	0.005	0.297
ACF	0.832	0.010	0.207	0.205	0.011	0.171
HH_Parametric	0.621	0.034	0.035	0.374	0.031	0.031
HH_SNP	0.622	0.035	0.035	0.373	0.031	0.031

Table 2 shows the estimation results for the sample with the latent productivity being the single unobservable in the demand of the proxy variables. We do not compute the LP or OP estimates because of the identification issue pointed out by ACF. The coefficients estimated using the methods, except for OLS, are all very similar. Thus, without additional errors in the input/investment demand functions, both ACF and our method (HH_parametric and HH_SNP) can control for endogeneity caused by the unobserved productivity. Table 3 shows the comparison of the estimates when the demands of the intermediate inputs/investment are also affected by other unobserved factors. In this case, as we have argued, the ACF method cannot fully control for the unobserved productivity. The ACF and OLS estimates are biased in the same direction, with ACF having a smaller bias relative to OLS. Our Parametric and SNP estimates are much closer to true values than OLS and ACF estimates, and achieve significant reductions in the root mean squared error relative to the OLS and ACF estimates.

Given the above Monte Carlo evidence, one can build a test of measurement error, for example, by comparing our estimates to the ACF estimates. Under the null hypothesis that the inputs and investment are not affected by unobserved factors other than productivity, the two methods should produce close estimates.

6.2 Chilean Manufacturing Industry Data

We use an eight-year plant-level panel, from 1979 to 1986, for the Chilean manufacturing industries to illustrate our method. The data are the same as those used by Liu (1991), LP and ACF. We only briefly describe the data here. Readers are referred to Liu (1991) and LP for the details about the data.

The panel is from an annual census of the Chilean manufacturing industry, covering all plants with at least ten employees. The data we use include revenue (the measure of output, net of intermediate inputs), labor, capital, and the intermediate inputs of electricity and material. All the variables are measured in 1980 Chilean Pesos, deflated using their own annual price deflators. The capital input is measured by the total value of machineries, buildings and vehicles at each plant. The labor input is measured by the total wage for both blue collar and white collar workers. Liu (1991) gives the details of how the capital input is constructed. The intermediate inputs are measured by the net purchases of the inputs in each year. These intermediate input variables, especially the material, are likely measured with errors since the inventories are not observed. The material input is positive for over 99 percent observations. The electricity input is also positive for more than 90 percent observations in most industries. However, only less than half of the observations of investment are nonzero.

We follow LP and ACF by focusing on the Food Products industry (311), which has much larger number of plants and observations than the other three industries that LP and ACF also looked at. Table 4 gives the number of plants and observations for each industry in the data.

Table 4: The Plants and Observations in the Four Industries

Industry	Num. of plants	Num. of observations
Food (311)	926	4699
Textile (321)	193	803
Wood Products (331)	174	686
Metal (381)	258	1101

In our estimation, we use polynomials to approximate the unknown input demand functions in the model. Particularly, we specify the material and electricity demand functions as follows

$$\begin{aligned}\mu_t(k_t, l_t, \omega_t) &= \gamma_0 + \gamma_1 l_t + \gamma_2 k_t + \gamma_3 l_t^2 + \gamma_4 k_t^2 + \gamma_5 k_t l_t + \\ &\quad \left(\gamma_6 + \gamma_7 l_t + \gamma_8 k_t + \gamma_9 l_t^2 + \gamma_{10} k_t^2 + \gamma_{11} k_t l_t\right)\omega_t \\ \psi_t(k_t, l_t, \omega_t) &= \lambda_0 + \lambda_1 l_t + \lambda_2 k_t + \lambda_3 l_t^2 + \lambda_4 k_t^2 + \lambda_5 k_t l_t + \\ &\quad \left(\lambda_6 + \lambda_7 l_t + \lambda_8 k_t + \lambda_9 l_t^2 + \lambda_{10} k_t^2 + \lambda_{11} k_t l_t\right)\omega_t\end{aligned}$$

The approximation is linear in the latent productivity, which simplifies the computation of the likelihood function and is appropriate given the assumption of Hicks-neutral productivity. Furthermore, we assume that the productivity process to be AR(1) also to facilitate computation. All the density functions are approximated by second-order Hermite series. Following LP, we also allow the input demand functions to be different across the three periods in the eight years in the data (1979-1981, 1982-1983, and 1984-1986). Our estimates are computed as described in (7), using material and electricity as the two proxy variables. Those of LP and ACF are computed using material and electricity separately.

All estimates are presented in Table 5. First, we compare our SNPMLE estimates to those obtained using previous methods. The estimates of the labor elasticity using LP and ACF's methods range from 0.68 (LP(M)) to 0.87 (ACF(E)). Our estimate of the labor elasticity, 0.54, is significantly smaller than all these estimates. In comparison, all the estimates of the capital elasticity are similar, which are around 0.40. For the implied returns to scale, our estimate is quite close to one, while the previous estimates all suggest increasing returns to scale to some extent. As the labor input is normally more variable than the capital input, the labor elasticity tend to be overestimated when the latent productivity is not well controlled for. The comparison of the estimates suggests that our method can be a useful alternative when the single unobservable assumption is not appropriate for the proxy variables.

In the parametric version of our model, we assume that the conditional distributions in the likelihood functions are normal distributions. The parametric estimates are close to our SNP estimates. As the parametric model is very easy to code and compute, we use it to find a reasonable guess for the parameters in the SNP model. It could also be a useful substitute in the preliminary analysis in practice.

It is also interesting to check how the inputs demand varies with the latent productivity. Given the Hicks-neutral productivity, we should expect inputs demand to increase with productivity, controlling for the other factors, that is, the slopes of the intermediate inputs demand with respect to productivity should be positive at all values of labor and capital. Figure 1-Figure 3 (see Appendix) present our estimates of the slopes of material and electricity demand with respect to productivity at all values of labor and capital in the data. From the graphs, we see that the slopes of the input demand functions are positive at almost all values of labor and capital across all three periods. The only exception is that, in Figure 3, at a few small values of capital and labor the slope of the material demand is slightly negative in the third period (1984-1986). The input demand slopes also vary widely with capital and labor. The input demand increases with productivity more quickly at plants with large capital input and/or large labor input. The relationship between the input demand slope and the capital and labor input are relatively stable across all three periods. In general, the estimated relation between the intermediate inputs and productivity seem quite reasonable.

Table 5: Estimation Results, with the Chilean Manufacturing Data

	Parameters					
	Labor		Capital		Return to Scale	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev
Food Industry (311)						
OLS	1.080	0.042	0.336	0.025	1.416	0.026
LP (M)	0.676	0.037	0.455	0.038	1.131	0.035
LP (E)	0.765	0.040	0.446	0.032	1.210	0.034
ACF (M)	0.842	0.048	0.371	0.037	1.212	0.034
ACF (E)	0.865	0.047	0.379	0.031	1.244	0.032
HH_parametric (M&E)	0.610	0.032	0.372	0.019	0.982	0.039
HH_SNP (M&E)	0.539	0.071	0.415	0.030	0.954	0.047

Note: (M) and (E): material input and electricity input as the proxy variables. The standard deviations are bootstrapped.

7 Conclusions

In this paper, we proposed a new method to estimate production functions using panel data. Our method provides the robustness that is very important when the proxy variables used by OP, LP and ACF are affected not only by the latent productivity, but also by some other unobserved factors. In addition, the method frees up some important identification sources that were not applicable in the methods of OP, LP and ACF. In view of the large number of applications based on the previous methods, we believe that our contribution to this literature will be of value to future studies of various issues centered around firm productivity and production functions.

References

ACKERBERG, D., K. CAVES, AND G. FRAZER (2006): "Structural identification of production functions," *Unpublished manuscript*, 12, 19–20.

ACKERBERG, D., X. CHEN, AND J. HAHN (2011): "Asymptotic Variance Estimator for Two-Step Semiparametric Estimators," *Cowles Foundation Discussion Papers.*

AW, B., M. ROBERTS, AND D. XU (2008): "R&D investments, exporting, and the evolution of firm productivity," *The American Economic Review*, 98(2), 451–456.

BERNARD, A., J. EATON, J. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *American Economic Review*, pp. 1268–1290.

BLUNDELL, R., AND S. BOND (2000): "GMM estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19(3), 321–340.

BOND, S., AND M. SÖDERBOM (2005): "Adjustment costs and the identification of Cobb Douglas production functions.," .

DORASZELSKI, U., AND J. JAUMANDREU (2008): "R&D and productivity: Estimating production functions when productivity is endogenous," *Harvard Institute of Economic Research Discussion Paper No. 2147.*

DUNFORD, N., AND J. SCHWARTZ (1971): *Linear Operators*, vol. 3. Wiley.

GALLANT, A. R., AND D. W. NYCHKA (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

HU, Y., AND S. SCHENNACH (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76, 195–216.

IMBENS, G. (2007): "Nonadditive models with endogenous regressors," *ECONOMETRIC SOCIETY MONOGRAPHS*, 43, 17.

IMBENS, G., AND W. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

JAVORCIK, B. (2004): "Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages," *The American Economic Review*, 94(3), 605–627.

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

Liu, L. (1991): *ENTRY-EXIT AND PRODUCTIVITY CHANGE: AN EMPIRICAL ANALYSIS OF EFFICIENCY FRONTIERS (CHILE)*. THE UNIVERSITY OF MICHIGAN.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Pavcnik, N. (2002): "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *The Review of Economic Studies*, 69(1), 245.

Shen, X. (1997): "On Methods of Sieves and Penalization," *Annals of Statistics*, 25, 2555–2591.

Wooldridge, J. (2002): *Econometric analysis of cross section and panel data*.

Appendix A: Monte Carlo Simulation Set-up

In this appendix, we describe the model we used to simulate the random samples used in our Monte Carlo exercise. For the simulation, we assume the following Cobb-Douglas production for our simulation:

$$Y_t = \exp(\omega_t + \eta_t) L_t^{\beta_l} K_t^{\beta_k} M_t^{\beta_m} U_t^{\beta_u}$$

where M_t and U_t are the static intermediate inputs of material and electricity; ω_t follows an $AR(1)$ process

$$\omega_t = \rho\omega_{t-1} + \xi_t$$

; and η_t and ξ_t are i.i.d across time and firms. We set the equilibrium price for the output Y_t as 1, and the unit price of material and electricity input and wage as p_m, p_u and p_l, respectively. We let L_t be determined in period $t-1$, after K_t is determined, without the observing η_t and ξ_t. And M_t and U_t are determined simultaneously in period t after observing L_t and K_t. We let M_t, U_t and L_t to be affected by some error terms (unobserved by econometrician) in addition to by productivity. Denote $\Delta \equiv \log E(\exp(\eta_t))$. Then we have the following demand functions for material:

$$m_t^* = \frac{1}{1-\beta_m-\beta_u}\left[\Delta + (1-\beta_u)\log\frac{\beta_m}{p_m} + \beta_u\log\frac{\beta_u}{p_u} + \beta_l l_t + \beta_k k_t + \omega_t\right]$$

and for electricity:

$$u_t^* = \frac{1}{1-\beta_m-\beta_u}\left[\Delta + \beta_m\log\frac{\beta_m}{p_m} + (1-\beta_m)\log\frac{\beta_u}{p_u} + \beta_l l_t + \beta_k k_t + \omega_t\right]$$

The material and electricity inputs are measured with errors as follows:

$$\begin{aligned} m_t &= m_t^* + \varepsilon_t \\ u_t &= u_t^* + v_t \end{aligned}$$

Given the optimal material and electricity inputs, the expected output in the period t given the information in period $t-1$ would be:

$$E(Y_t|\omega_{t-1}) = \Delta_1 \exp(\rho\tilde{\omega}_{t-1}) K_t^{\tilde{\beta}_k} L_t^{\tilde{\beta}_l}$$

where

$$\Delta_1 \equiv \left(E(\exp(\eta_t))\left(\frac{\beta_m}{p_m}\right)^{\beta_m}\left(\frac{\beta_u}{p_u}\right)^{\beta_u}\right)^{\frac{1}{1-\beta_m-\beta_u}} E\left(\exp\left(\tilde{\xi}_t\right)\right)$$

and $\tilde{x} \equiv \frac{x}{1-\beta_m-\beta_u}$ for any parameter or variable x. So, the labor demand function can be derived as

$$l_t = \frac{1}{1-\tilde{\beta}_l}\left[\log\Delta_1 + \log\frac{\tilde{\beta}_l}{p_l} + \tilde{\beta}_k k_t + \rho\tilde{\omega}_{t-1}\right] + \epsilon_t$$

where ϵ_t is an unobserved factor affecting labor demand. The value-added (net of the costs of intermediate inputs) production function in log form is

$$\begin{aligned} y'_t &= \log\Delta_2 + \tilde{\beta}_l l_t + \tilde{\beta}_k k_t + \rho\tilde{\omega}_{t-1} + \tilde{\xi}_t + \log\left(\frac{\exp(\eta_t)}{E(\exp(\eta_t))} + \beta_m + \beta_u\right) \\ \Delta_2 &\equiv \left(E(\exp(\eta_t))\left(\frac{\beta_m}{p_m}\right)^{\beta_m}\left(\frac{\beta_u}{p_u}\right)^{\beta_u}\right)^{\frac{1}{1-\beta_m-\beta_u}} \end{aligned}$$

where y'_t is value-added. Finally, we assume the following investment equation:

$$i_{t-1} = \gamma_0 + \gamma_1 k_{t-1} + \gamma_2\omega_{t-1} + \zeta_{t-1}$$

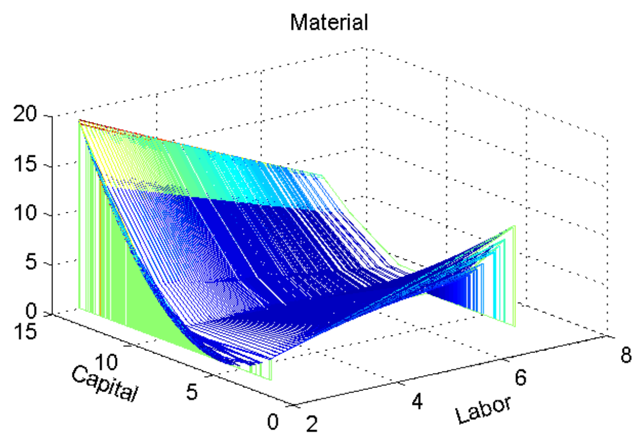
; and that the capital accumulates according to the following equation:

$$k_t = k_{t-1} + i_{t-1}$$

where i_{t-1} is investment; and ζ_{t-1} is an unobserved factor affecting investment. In the simulation, the error terms, $\eta_t, \varepsilon_t, v_t, \epsilon_t$ and ζ_t are mutally independent random variables from the standard Normal distributions. For estimation, we use m_{t-1} and u_{t-1} as the two proxy variables.

Appendix B: Figures

Figure 1: The Slope of Input Demand with Regard to the Latent Productivity, 1979-1981



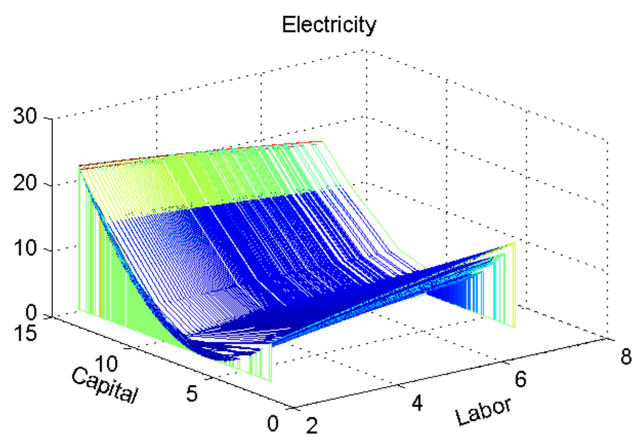


Figure 2: The Slope of Input Demand with Regard to the Latent Productivity, 1982-1983

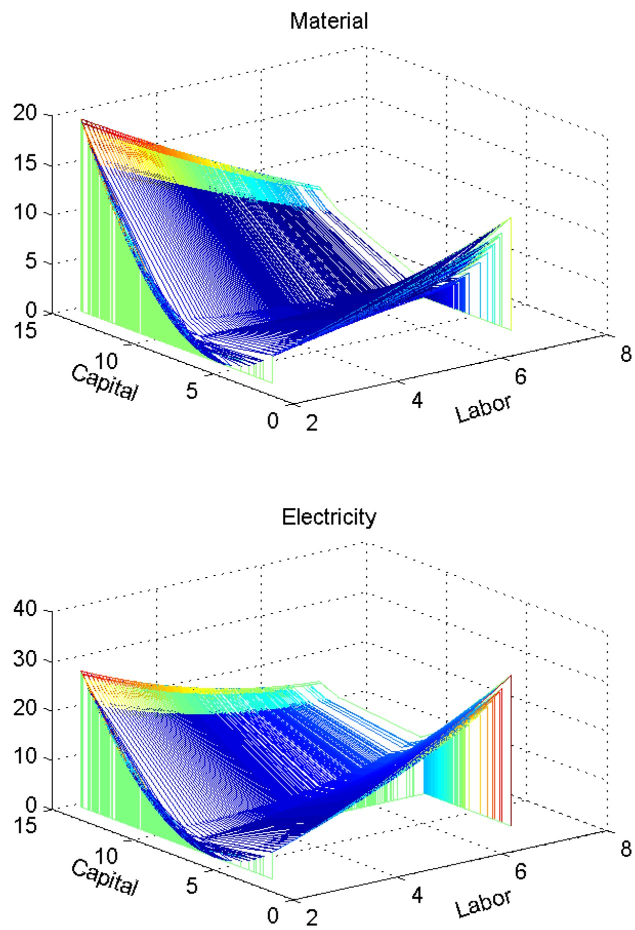


Figure 3: The Slope of Input Demand with Regard to the Latent Productivity, 1984-1986