Estimating Production Functions with Robustness Against Errors in the Proxy Variables

Yingyao Hu*, Guofang Huang† and Yuya Sasaki‡

Jan 2019

Abstract

This paper proposes a new approach to the identification and estimation of production functions. It extends the literature on the structural estimation of production functions, which dates back to 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 (e.g. the investment and the material input). The proposed generalized method of moment (GMM) estimator is flexible and straightforward to apply. The method is applied to study how rapidly firms in the Chilean food-product industry adjust their inputs in response to shocks to their productivity.

JEL Classifications: C14

Key Words: Production functions, proxy variables, instrumental variables, optimization errors, measurement errors.

*Johns Hopkins University

†Corresponding author, address: Krannert School of Management, Purdue University, 403 W. State Street West Lafayette, IN 47907; email: guofanghuang@gmail.com.

‡Vanderbilt University

1 Introduction

The literature on the structural estimation of production functions addresses two main problems: the simultaneity bias and sample selection. Both problems are caused by the unobserved productivity in production functions. Olley and Pakes (1996) (hereafter OP), in their seminal paper, suggest using investment as a proxy variable to control for the latent productivity. Their key insight is that if productivity, a scalar random variable, is the only unobserved factor affecting investment (i.e., the *scalar-unobservable assumption*), and investment is, ceteris paribus, a strictly monotonic function of the latent productivity (i.e, the monotonicity assumption), then one can consistently estimate the structural parameters in the production function by using a nonparametric function of investment and other covariates to control for the latent productivity. OP's approach and the later extensions of it have been widely applied in the IO and trade literature (e.g., Pavcnik (2002), Bernard et al. (2003), Javorcik (2004), Aw et al. (2008)). Building on the insights from the literature, this paper proposes a new approach to identifying and estimating production functions while relaxing the scalar-unobservable assumption. We focus on dealing with the simultaneity bias, following Levinsohn and Petrin (2003) and Ackerberg et al. (2015).

Important discussions and extensions of the OP's method include Levinsohn and Petrin (2003) (hereafter LP), Ackerberg et al. (2015) (hereafter ACF), among others. LP argue that static inputs, such as material and energy, may be better proxy variables for productivity because the primitive conditions that ensure the monotonicity condition for these proxy variables are easier to come by, and that they are normally much less lumpy and have fewer observations of zero. ACF point out an important identification problem with estimating the coefficient of the labor input in the first step of OP/LP's procedure. In particular, if the labor demand, like investment/intermediate inputs, is also a function of capital and productivity but of no other unobserved factors, then, after controlling for capital and productivity perfectly by a nonparametric function of capital and the proxy variable, there would be no independent variations in the labor input left to identify the coefficient of labor in the OP/LP's first step.

Though the scalar-unobservable assumption is a key to the above methods, it has been a concern since OP's original paper (p.1265). LP also point out that a major criterion in selecting their proxy variable is the avoidance of inputs that could be subject to the influence of other unobserved factors (LP, p.326). In general, some other unobserved factors, such as supply disruptions, optimization errors and measurement errors, could also affect the observed investment and inputs. If these other unobserved factors were important in practice, the

OP/LP/ACF procedures might not fully control for the latent productivity.[1] Furthermore, the scalar-unobservable assumption also forces us to give up some important sources of identification. This problem manifests itself most clearly through the identification problem, as ACF point out, in the estimation of the labor coefficient in LP's first step. Although researchers may not run into such an identification problem in practice, to maintain logical consistency, one does not want both to use the additional sources of variations in the labor input—due to cost shocks, for example—to identify the labor coefficient in OP/LP's first step *and* to exploit the single-unobservable assumption to use the investment or an intermediate input as a perfect proxy variable for the latent productivity. Related to this issue, Bond and Söderbom (2005) point out the difficulty of identifying the coefficients of fully flexible inputs when there are no variations 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 identify and estimate production functions, *allowing the proxy variables to be affected by other unobserved factors in addition to the latent productivity.* The idea of our method is as follows: because researchers normally have multiple proxy variables such as intermediate inputs and investment available for productivity, we may be able to find two such proxy variables that, conditional on productivity (and other covariates), are independent of each other in some reasonable cases. Then, we may intuitively view these two proxy variables as two contaminated measures of productivity, such that we may 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 in many important cases, even when the proxy variables do not satisfy the scalar-unobservable assumption.

Two key conditions are needed for our identification of production functions. The first one is the conditional independence condition alluded to above, and

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

the second is an injectivity condition that may be viewed as a generalization of the monotonicity condition of OP/LP after relaxing the scalar-unobservable assumption. As will be discussed in detail later, the conditions are reasonable in some important cases.

Our identification argument provides the foundation for alternative estimation methods that do not rely on the scalar-unobservable assumption about the proxy variables. A sieve maximum likelihood estimation (MLE) method follows directly from our identification result. The MLE method is feasible but harder to implement in practice than the methods of OP/LP, due to the functional nuisance parameters involved in the estimation. As a more practical alternative to the MLE method, we propose a GMM estimation approach, based on the same general identification idea of using two proxy variables for the latent productivity. To derive the GMM estimator, we impose several moment restrictions that are related to, but not implied by, the conditional independence condition mentioned above. Our GMM estimator may be viewed as an extension of the IV approach (Blundell and Bond (2000)) in that we do not restrict the AR(1) process for productivity transition to be linear.

We also provide a test of the econometric model of OP/LP to help applied researchers choose between OP/LP's model and our extension of their model. We base our test on the ACF critique of the OP/LP procedure, that is, the coefficient of the labor input (β_l) is not identified in the first step of the OP/LP estimation procedure. The lack of identification implies that we cannot reject such a null hypothesis as $H_0 : \beta_l = \beta_l^*$ for β_l^* being any fixed finite value in the first step of OP/LP's estimation procedure. Our specification test can thus be formulated as a test of the null of $H_0 : \beta_l = 0$, for example, in the first step of OP/LP's estimation procedure. Rejection of the null hypothesis $H_0 : \beta_l = 0$ thus also leads to the rejection of OP/LP's model. We develop the theory of the test using an inference procedure that is robust against possible non-identification of parameter(s). The test rejects OP/LP's model for the Chilean manufacturing data that we use in our empirical illustration.

To illustrate our GMM estimation method, we first compare its performance to those of the existing methods in Monte Carlo studies. The results show the robustness of our method—but not of the existing ones discussed above—against the existence of additional unobserved factors affecting the proxy variables. We then apply our method to the Chilean manufacturing data, as used by LP, to investigate how rapidly firms adjust the various inputs in response to the latest shocks to productivity. The empirical analysis shows that firms adjust the material input quickly to fit the latest level of productivity, but they are considerably slower in adjusting the labor and capital inputs, suggesting significant frictions in the corresponding input markets. The empirical results also help explain the differences in the estimates of production functions using the various methods.

The rest of the paper proceeds as follows. Section 2 shows the identification of production functions in a model that relaxes the scalar-unobservable assumption. Section 3 first proposes new estimation methods based on our identification result; then develops a test of the model of OP/LP and compares our methods to the existing ones using simulated data. Section 4 applies our method to the Chilean manufacturing-industry census data. Section 5 concludes.

2 Model and Identification

In this section, we study the identification 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; then, we set up a standard model of gross-output production function and show its identification. To save space, we defer our review and discussion of the related literature to the appendix, and refer readers to Ackerberg et al. (2007) and Ackerberg et al. (2015) for comprehensive reviews of the related literature.

Our main identification idea is to simultaneously use two proxy variables for productivity in the estimation of production functions. The two proxy variables can be thought of as two contaminated measures of the latent productivity. Intuitively, although one cannot directly invert the demand function of a proxy variable to fully control for the latent productivity, due to the presence of additional unobserved factors affecting the variable, we can use the other proxy variable as an instrument for the first one. Given this perspective of the model, we can employ the identification result from Hu and Schennach (2008) for nonclassical measurement-error models to show the identification of parameters in the production function. To illustrate the crux of our identification argument, suppose that we are interested in estimating the structural parameters β in the following equation of y_t,

$$y_t = W_{1t}\beta + \omega_t + \eta_t, \tag{1}$$

where W_{1t} is a vector of observed variables; and ω_t and η_t are unobserved scalar random variables. And suppose that we have two proxy variables for 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 $W_t = (W_{1t}, W_{2t})$ (W_{2t} indicates other covariates relevant to x_t and z_t)—i.e., $f(y_t|\omega_t, z_t, x_t, W_t) = f(y_t|\omega_t, W_t)$

and $g(z_t|\omega_t, x_t, W_t) = g(z_t|\omega_t, W_t)$, where $f(.)$ and $g(.)$ are conditional density functions; and 2) the integral operators defined by $f(y_t|\omega_t, W_t)$ and $h(\omega_t|x_t, W_t)$ are injective for any given W_t,[2] and $g(z_t|\omega_t, W_t)$ satisfies a mild technical assumption (to be clarified later). Then, it can be shown that the conditional density of $f(y_t|\omega_t, W_t)$, as well as $g(z_t|\omega_t, W_t)$ and $h(\omega_t|x_t, W_t)$, are identified through the following equation based on the observed conditional density of $f(y_t, z_t|x_t, W_t)$:[3]

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

As a result, the structural parameters β are identified given that $f(y_t|W_t, \omega_t)$ is identified.

Note that we impose the injectivity assumption on the integral operator defined by the conditional density related to one of the two proxy variables, and require the conditional density related to the other proxy variable to satisfy only a mild technical condition. We normally can find two such proxy variables in applications as we discuss in detail below.

2.1 Model

We assume that the general underlying structural framework is the same as that described by Olley and Pakes (1996), and follow the tradition of using uppercase letters to denote levels and lowercase letters to denote the logarithms of levels. We focus on the following Cobb-Douglas gross-output production function:[4]

$$y_t = \beta_l l_t + \beta_k k_t + \beta_m m_t + \beta_u u_t + \omega_t + \eta_t, \tag{2}$$

where y_t, l_t, k_t, m_t and u_t are, respectively, the logarithms of the output and the inputs of labor, capital, material and energy; ω_t is the logarithm of the latent productivity that subsumes the constant and is serially correlated; and η_t is the residual term with $\mathbb{E}(\eta_t|l_t, k_t, m_t, u_t, \omega_t) = 0$. The functional form assumption is made here for the ease of demonstration. The identification result that we

[2] For a conditional density function $f(x|z)$, the corresponding integral operator is defined as follows: $L_{x|z}(h(.))(x) = \int f(x|z)\, h(z)\, dz$.

[3] The equation is a result of the total law of probability and the conditional independence assumption.

[4] Gandhi et al. (2013) provides some compelling motivations for researchers to focus on gross-output, as opposed to value-added, production functions. We focus on the Cobb-Douglas production function in the paper because of its importance and popularity in applications. One may adapt our identification and estimation framework without much difficulty if one chooses to work with alternative production functions (e.g., CES production functions).

show in the following applies equally well to other common forms of production functions. Our interest here is to identify and estimate $(\beta_l, \beta_k, \beta_m, \beta_u)$, given that ω_t is correlated with (l_t, k_t, m_t, u_t) but is not observed by the econometrician. For productivity ω_t, let $\mathbb{E}(\omega_t|\mathcal{I}_{t-1})$ be the prediction of ω_t based on $\mathcal{I}_{t-1}$, the information available in period $t-1$, and $\xi_t = \omega_t - \mathbb{E}(\omega_t|\mathcal{I}_{t-1})$ is the prediction error. In the following, we assume that ω_t follows an exogenous first-order Markov process, such that $\mathbb{E}(\omega_t|\mathcal{I}_{t-1}) = \mathbb{E}(\omega_t|\omega_{t-1})$. We define $\rho(\omega_{t-1}) \equiv \mathbb{E}(\omega_t|\omega_{t-1})$. The more general case of ω_t following a controlled Markov process can be treated similarly 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 suit the specific industries under analysis. To be specific, we assume that decisions about inputs of l_t, m_t and u_t are made simultaneously in period t after observing ω_t and k_t, and that k_t is determined in period $t-1$ without observing the period-t innovation, ξ_t, of productivity.

More specifically, let the *optimal* choices of l_t, m_t and u_t for a firm be determined as the solution to the following profit-maximization problem:

$$\max_{L_t, M_t, U_t} \mathbb{E}\exp(\eta_t)\exp(\omega_t) L_t^{\beta_l} K_t^{\beta_k} M_t^{\beta_m} U_t^{\beta_u} - (p_{lt}L_t + p_{mt}M_t + p_{ut}U_t),$$

where the expectation is taken with respect to η_t; (p_{lt}, p_{mt}, p_{ut}) are the input prices of L_t, M_t and U_t respectively; and the output price is normalized to be 1 per unit. This problem yields linear reduced-form input choice rules, for $x = l, m, u$, as follows:

$$x_t = \alpha_{x0} + \alpha_{xk}k_t + \alpha_{x\omega}\omega_t + p_t\alpha_{xp},$$

where $p_t = (p_{lt}, p_{mt}, p_{ut})$ and α_{xp} is a vector of the corresponding parameters. The reduced-form parameters $(\alpha_{x0}, \alpha_{xk}, \alpha_{x\omega}, \alpha_{xp})$ are functions of $(\beta_l, \beta_k, \beta_m, \beta_u)$ and $\mathbb{E}\exp(\eta_t)$. Following the literature, we call l_t, m_t and u_t static inputs (except when we consider l_t being determined in period $t-1$).

As an important extension of the literature, we let the *observed* static inputs be determined, for $x = l, m, u$, as follows:

$$x_t = \alpha_{x0} + \alpha_{xk}k_t + \alpha_{x\omega}\omega_t + \epsilon_{xt}, \tag{3}$$

where ϵ_{xt} is a scalar random variable. The inclusion of ϵ_{xt} relaxes the scalar-unobservable assumption maintained by the previous papers (e.g. OP, LP and ACF) in the literature. In practice, ϵ_{xt} can capture $p_t\alpha_{xp}$ if the input prices are firm-specific and not observed by researchers, and/or other factors that cause the observed inputs to deviate from their optimal levels.[5] We defer more detailed discussions of possible empirical interpretations of ϵ_{xt} to section 2.4.1. In our identification argument, we use the following expression of x_{t+1}:

$$x_{t+1} = \alpha_{x0} + \alpha_{xk} k_{t+1} + \alpha_{x\omega}\rho(\omega_t) + \alpha_{x\omega}\xi_{t+1} + \epsilon_{xt+1}, \tag{4}$$

connecting x_{t+1} with ω_t.

It is worth noting that the optimal input demand functions derived from the first-order conditions are just natural candidates that we extend to illustrate how we may allow the input demand functions to depend on more than a single unobservable. It is straightforward to extend our identification and estimation to allow more flexible specifications. In particular, we can allow the following more flexible specifications for the static inputs:

$$x_t = \mu_{xt}(k_t, \omega_t) + \epsilon_{xt},$$

where $\mu_{xt}(k_t, \omega_t)$ are polynomials of k_t and ω_t. We can also extend our methods easily to the cases in which l_t depends on l_{t-1} or l_t is determined in period $t-1$.[6]

The data-generating process for the investment I_t is somewhat different from those of the above static inputs.[7] In practice, we often observe a significant portion of the firms in the data making 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:

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

where I_t is the observed investment; I_t^* is a latent index variable; and ζ_t captures unobservable factors, other than ω_t, that affect investment. The observed investment data are censored at zero.

To complete the model, let capital accumulates according to the following equation:

$$K_t = \kappa(K_{t-1}, I_{t-1}, \nu_{t-1}), \tag{5}$$

[5] We can simply add $p_t\alpha_{xp}$ back to equation (3) in cases in which researchers do observe firm-specific input prices.

[6] As will become clear later, the only adjustment that we need for these two cases is to include l_{t+1} in the equations of m_{t+1} and u_{t+1}

[7] The optimal investment is determined as the solution to firms' dynamic profit optimization problems (Olley and Pakes (1996)), which we omit here to avoid unnecessarily restricting us to a particular model specification.

where ν_{t-1} captures other factors affecting the capital accumulation process. As we explain later, by breaking the deterministic linear relationship between K_t and (K_{t-1}, I_{t-1}), the specification in (5) allows us to use I_t as one of the proxy variables for ω_t in identification. In practice, ν_{t-1} can include, for example, 1) lagged investments when some investments take more than one period to become productive capital (as argued by the influential paper of Kydland and Prescott (1982)); 2) shocks to the process of turning investment into productive capital; and 3) stochastic factors that affect capital depreciation. It is worth noting that a common specification for the capital accumulation process, as adopted by the papers that we discussed above, is $K_t = (1-\delta)K_{t-1} + I_{t-1}$, where δ is a depreciation factor. We interpret this particular deterministic process mainly as a parsimonious specification assumed to be consistent with the timing assumption of k_t being determined in period $t-1$ (without observing the period-t productivity innovation ξ_t), instead of literately to emphasize that the current-period investment becomes productive capital in the next period. Our data also show that the actual capital accumulation is a more nuanced process, and thus the above more flexible specification seems appropriate here. Lastly, we assume that $\eta_t, \alpha_{x\omega}\xi_{t+1} + \epsilon_{xt+1}, \zeta_t$ (where $x = m, u$) are mutually independent conditional on $(\omega_t, l_t, k_t, m_t, u_t)$, and that $\eta_t \perp\!\!\!\perp \nu_t | (\omega_t, l_t, k_t, m_t, u_t)$.

2.2 Identification

In the following, we base our identification discussion on three endogenous variables, (y_t, m_{t+1}, I_t), all of which depend on the unobserved productivity ω_t, in addition to the control variables and error terms. Let $W_t \equiv (l_t, k_t, m_t, u_t, k_{t+1})$ indicate the vector of control variables. We begin our identification argument by listing the conditions that we need to prove identification.

Condition 1. *(Conditional Independence)* $f(y_t | m_{t+1}, I_t, \omega_t, W_t) = f(y_t | \omega_t, W_t)$, *and* $g(I_t | m_{t+1}, \omega_t, W_t) = g(I_t | \omega_t, W_t)$, *for all* W_t, *where* f *and* g *are conditional density functions.*

Condition 2. *(Injectivity) i)* $\eta_t \perp\!\!\!\perp \omega_t | W_t$, *and* $(\alpha_{m\omega}\xi_{t+1} + \epsilon_{mt+1}) \perp\!\!\!\perp \omega_t | W_t$; *ii)* $\rho(\omega_t) = \mathbb{E}(\omega_{t+1} | \omega_t)$ *is strictly monotonic in* ω_t; *and iii) conditional characteristic functions of* $f(y_t | \omega_t, W_t)$ *and* $h(\omega_t | m_{t+1}, W_t)$ *do not vanish on the real line.*

The first equality in condition 1 states that m_{t+1} and I_t do not provide information about y_t beyond what is already contained in ω_t. The second equality in condition 1 says that the two proxy variables are independent of each other, conditional on ω_t and other control variables. The conditional independence assumptions follow from our model assumption that $\eta_t, \alpha_{m\omega}\xi_{t+1} + \epsilon_{mt+1}$ and ζ_t are mutually independent conditional on $(\omega_t, l_t, k_t, m_t, u_t)$, and they can be thought of intuitively as similar to the exclusion restrictions in the instrumental variable (IV) method. As a direct application of Proposition 2.4 in D'Haultfoeuille (2011), condition 2 guarantees that the integral operators defined by $f(y_t|\omega_t, W_t)$ and $h(\omega_t|m_{t+1}, W_t)$ are invertible. The injectivity assumption plays a role in our identification similar to that of the rank condition in the IV method. We also need the following two technical conditions for our identification.

Condition 3. *(Distinctive Eigenvalues) for any given W_t and any $\overline{\omega}_t \neq \widetilde{\omega}_t$, there exists a set A such that $g(I_t|\overline{\omega}_t, W_t) \neq g(I_t|\widetilde{\omega}_t, W_t)$ for all $I_t \in A$ and $\Pr(A) > 0$.*

Condition 4. *(Normalization) $\mathbb{E}(y_t - \beta_l l_t - \beta_k k_t - \beta_m m_t - \beta_u u_t|\omega_t, l_t, k_t, m_t, u_t) = \omega_t$; that is, $\mathbb{E}(\eta_t|\omega_t, l_t, k_t, m_t, u_t) = 0$.*

Condition 3 is a relatively mild condition — it requires only that, *ceteris paribus*, any change in a firm's productivity has to lead to some change in the distribution of the firm's investment decisions. The condition 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. It is also worth noting that condition 3 is feasible given the flexible capital accumulation process specified in (5). If we assumed $K_t = (1-\delta)K_{t-1} + I_{t-1}$, I_t would be completely determined by k_t and k_{t+1} (both of which are in W_t).[8] Condition 4 will be used to pin down the eigenfunctions for each given ω_t, which follows directly from our model assumption in equation (2). We will discuss the practical validity of the four conditions later in this section.

Given condition 1, we have:

$$\begin{aligned} & f(y_t, I_t|m_{t+1}, W_t) \\ = & \int f(y_t|I_t, m_{t+1}, \omega_t, W_t)\, g(I_t|m_{t+1}, \omega_t, W_t)\, h(\omega_t|m_{t+1}, W_t)\, d\omega_t \\ = & \int f(y_t|\omega_t, l_t, k_t, m_t, u_t)\, g(I_t|\omega_t, W_t)\, h(\omega_t|m_{t+1}, W_t)\, d\omega_t, \end{aligned}$$

where the first equality follows by the law of total probability; and the second equality follows from the conditional independence condition and our model assumption that η_t and ν_t are mutually independent conditional on $(\omega_t, l_t, k_t, m_t, u_t)$.

[8] Meanwhile, as we often observe several different types of investments, such as building, machinery and vehicles, one may use one of the different types of investments as a proxy variable so that condition 3 is feasible even if one assumes $K_t = (1-\delta)K_{t-1} + I_{t-1}$.

Copying the above equation for easier reference, we have:

$$f\left(y_t, I_t | m_{t+1}, W_t\right) = \int f\left(y_t | \omega_t, l_t, k_t, m_t, u_t\right) g\left(I_t | \omega_t, W_t\right) h\left(\omega_t | m_{t+1}, W_t\right) d\omega_t. \tag{6}$$

Now, the identification question is whether we can identify the latent conditional densities on the right-hand side of equation (6), especially $f\left(y_t | \omega_t, l_t, k_t, m_t, u_t\right)$, given the observed conditional density of $f\left(y_t, I_t | m_{t+1}, W_t\right)$.

Given the conditions above, Theorem 1 in Hu and Schennach (2008) can be applied to show that the latent densities $f\left(y_t | \omega_t, l_t, k_t, m_t, u_t\right)$, $g\left(I_t | \omega_t, W_t\right)$, and $h\left(\omega_t | m_{t+1}, W_t\right)$ 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 omit the control variables (W_t) in the proof for simpler notations. 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, define the integral operator $L_{x|z}$ based on the conditional density function $f(x|z)$ 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 $\mathcal{F}(\mathcal{X})$.

Now equation (6) can be equivalently written in corresponding integral operators as:

$$L_{I;y|m} = L_{y|\omega} \Delta_{I;\omega} L_{\omega|m}, \tag{7}$$

where $L_{I;y|m}$ is defined similarly to $L_{y|m}$ with $f(y|m)$ replaced by $f(I, y|m)$ for a given I, and where $\Delta_{I;\omega}$ is a "diagonal operator" mapping a function $h(\omega)$ to $f(I|\omega)\, h(\omega)$. Meanwhile, by integrating both sides of equation (7) with respect to I, we get $L_{y|m} = L_{y|\omega} L_{\omega|m}$, which is equivalent to:

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

Next, we substitute the above expression of $L_{\omega|m}$ into (7) and rearrange the operators based on observable densities to the left-hand side, and we get:

$$L_{I;y|m} L_{y|m}^{-1} = L_{y|\omega} \Delta_{I;\omega} L_{y|\omega}^{-1}. \tag{8}$$

The inverse of $L_{y|m}$ used in the above equation can be shown to exist because

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

$L_{y|\omega}$ and $L_{m|\omega}$ are invertible.

Equation (8) means that $L_{I;y|m}L_{y|m}^{-1}$ admits an eigenvalue-eigenfunction decomposition. The left-hand-side operator based on observed conditional densities is decomposed to obtain $f(y|\omega,.)$, and $g(I|\omega,.)$, 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 the density functions.

Lastly, conditions 3 and 4 together ensure the uniqueness of the ordering and indexing of the eigenvalues and eigenfunctions. By condition 3, the eigenvalue $g(I|\omega,.)$ is distinct for distinct values of ω. With condition 4, we uniquely determines both $f(y|\omega,.)$ and $g(I|\omega,.)$, by ordering them according to $\mathbb{E}(y_t-\beta_l l_t-\beta_k k_t - \beta_m m_t - \beta_u u_t|\omega_t, l_t, k_t, m_t, u_t)$.

The following Lemma summarizes our main result on the identification of $f(y_t|\omega_t, l_t, k_t, m_t, u_t)$.

Lemma 1. *Suppose that, for any fixed W_t, the joint density of $(y_t, I_t, m_{t+1}, \omega_t)$ conditional on W_t is bounded, and all marginal and conditional densities are also bounded. Then, under conditions 1, 2, 3, 4, the observed conditional density of $f(y_t, I_t|m_{t+1}, W_t)$ uniquely determines the latent conditional densities of $f(y_t|\omega_t, l_t, k_t, m_t, u_t)$, $g(I_t|\omega_t, W_t)$ and $h(\omega_t|m_{t+1}, W_t)$.*

Proof. The assumption of bounded densities corresponds to the Assumption 1 in Hu and Schennach (2008). Conditions 1-4 correspond to their Assumptions 2-5. Our theorem follows as a direct application of their Theorem 1. □

The independence and injectivity conditions play important roles in the above identification proof. The independence assumptions help reduce the dimensionality of the latent conditional densities to make the spectral decomposition possible. The injectivity assumptions ensure that the integration operators are invertible. This role played by the injectivity condition bears some similarity to that of the rank conditions for the IV method in the classical linear regression models.

Given the identification of the conditional densities and the assumptions of $\eta_t \perp\!\!\!\perp \omega_t|(l_t, k_t, m_t, u_t)$ and $\mathbb{E}(\eta_t|\omega_t, l_t, k_t, m_t, u_t) = 0$, the conditional density of η_t, $f_{\eta_t|(l_t,k_t,m_t,u_t)}$, and the structural parameters, $(\beta_l, \beta_k, \beta_m, \beta_u)$, in the production function are identified given enough variations in (l_t, k_t, m_t, u_t). We summarize the identification results in the following Theorem.

Theorem 1. *Let $V_t \equiv (l_t, k_t, m_t, u_t)'$ and $\beta \equiv (\beta_l, \beta_k, \beta_m, \beta_u)'$. Suppose that $\mathbb{E}(V_t V_t')$ is nonsingular, then under conditions 1, 2, 3 and 4, the observed conditional density $f(y_t, I_t|m_{t+1}, W_t)$ uniquely determines $(\beta_l, \beta_k, \beta_m, \beta_u)$, together* *with* $f_{\eta_t|(\omega_t,l_t,k_t,m_t,u_t)}$*,* $g(I_t|\omega_t,W_t)$ *and* $h(\omega_t|m_{t+1},W_t)$ *from the following equation:*

$$\begin{aligned} & f\left(y_t, I_t|m_{t+1}, W_t\right) \qquad (9) \\ = & \int_{-\infty}^{\infty} f_{\eta_t|(\omega_t,l_t,k_t,m_t,u_t)}\left(y_t - \beta_l l_t - \beta_k k_t - \beta_m m_t - \beta_u u_t - \omega_t\right) \times \\ & g\left(I_t|\omega_t, W_t\right) \times h\left(\omega_t|m_{t+1}, W_t\right) d\omega_t. \end{aligned}$$

Proof. The identification of $f\left(y_t|\omega_t,l_t,k_t,m_t,u_t\right)$ implies the identification of $\mathbb{E}(y_t|\omega_t,l_t,k_t,m_t,u_t)$. Meanwhile, our model has that $y_t = \beta_l l_t + \beta_k k_t + \beta_m m_t + \beta_u u_t + \omega_t + \eta_t$. Hence, we have $\mathbb{E}(y_t|0,l_t,k_t,m_t,u_t) = V_t'\beta$. Given that $\mathbb{E}(V_t V_t')$ is nonsingular by assumption, we get $\beta = \left(\mathbb{E}(V_t V_t')\right)^{-1}\mathbb{E}(\mathbb{E}(y_t|0,l_t,k_t,m_t,u_t)V_t)$. With $\mathbb{E}(V_t V_t')$ directly identified from data, β is identified.

Meanwhile, we have:

$$f_{\eta_t|(\omega_t,l_t,k_t,m_t,u_t)}\left(\tilde{\eta}\right) = f\left(y_t = \beta_l l_t + \beta_k k_t + \beta_m m_t + \beta_u u_t + \omega_t + \tilde{\eta}|\omega_t, l_t, k_t, m_t, u_t\right),$$

for any given $\tilde{\eta}$ and $(\omega_t,l_t,k_t,m_t,u_t)$. Since y_t and η_t share the same domain of the entire real line, the above equation identifies $f_{\eta_t|(\omega_t,l_t,k_t,m_t,u_t)}$. □

It is worth noting that the above identification arguments can also be made with (y_t, I_t, m_{t+1}) replaced by (y_t, I_t, u_{t+1}) or (y_t, I_t, y_{t+1}). However, we cannot make the same identification argument with (y_t, m_{t+1}, u_{t+1}), because the residual errors of both static inputs depend on ξ_{t+1} (equation (4)), directly contradicting condition 1. We cannot make the same identification argument with (y_t, I_t, m_t) either, because, in such a case, we cannot identify the coefficients of m_t and k_t in the production function due to collinearity.[10]

2.3 Extension

As an extension of the above identification result, we can allow ω_t to be endogenously determined. This extension is important for applications in which it is essential to assume that firms actively spend resources to improve productivity. For example, Doraszelski and Jaumandreu (2013) show that it is important to account for firms' R&D investment in explaining the evolution of firms' productivity in the Spanish manufacturing industry.[11]

[10] This is easy to check in the case with linear demand function for m_t. We can solve the m_t equation for ω_t as a function of m_t, k_t and ϵ_{mt}, and substitute it into the production function. Then, we can see that the coefficients of m_t and k_t in the production function cannot be separately identified from the coefficients of k_t and ω_t in the m_t equation.

[11] Doraszelski and Jaumandreu (2013) propose an alternative method to use a static input, as suggested by LP, to proxy for productivity in their estimation of production functions. Their key

Our method can 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:

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

where r_t is the R&D spending in period t (or some other control variable affecting the evolution of productivity),[12] and ϱ_t captures other unobserved factors affecting r_t. Under the alternative assumption, we have $\mathbb{E}(\omega_{t+1}|\mathcal{I}_t) = \mathbb{E}(\omega_{t+1}|\omega_t, r_t)$. Given that R&D spending is observed, our identification arguments above can be largely replicated as long as we replace the term $\mathbb{E}(\omega_{t+1}|\omega_t)$ in the m_{t+1} equations with $\mathbb{E}(\omega_{t+1}|\omega_t, r_t)$.

2.4 Discussion

We have shown above that the identification of production functions can be achieved even if we allow for additional unobservables in determining the proxy variables. In the following, we discuss the practical validity of the underlying conditions in order to assess the applicability of the above identification results for estimation. We discuss the key conditions in turn, assuming that the general underlying structural framework is the same as described in Olley and Pakes (1996).

2.4.1 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, I_t and m_{t+1} is equivalent to the assumption of mutual independence, conditional on the observable covariates, among the corresponding residual errors—i.e., $\eta_t, \zeta_t, \alpha_{m\omega}\xi_{t+1} + \epsilon_{mt+1}$. Whether it is reasonable to assume that the residual terms are mutually independent depends on the factors that they capture. The residual error of the output equation, η_t, may capture, for example, unanticipated technology shocks, such as the number of defective products and machine breakdowns, and/or measurement error of the output. And the residual errors in the equation of intermediate inputs and investment

insight is that, for some commonly used parametric production functions, one can easily solve for the optimal demand function for a static input and, thus, can get an explicit expression for the inverse function to back out the productivity. This observation, together with data on firm specific input costs, allows them to get around the identification problem, as pointed out by ACF, with LP's approach.

[12] r_t can also be other firm activities, such as exporting experience (De Loecker (2010)), that affect firms' productivity.

could be results of supply disruptions, optimization errors, idiosyncratic cost shocks, measurement errors, etc.. In the following, we discuss the conditional independence assumption for the main types of unobserved factors. We focus our discussion on the assumption of η_t, ζ_t and ϵ_{mt+1} being mutually independent, assuming that ξ_{t+1} is independent of (η_t, ζ_t) (which seems reasonable given typical interpretations of η_t and the possible interpretations of ζ_t that we discuss below).

Optimization error The conditional independence assumption seems reasonable if the residual errors in the equations of m_{t+1} and I_t are mainly optimization errors. Because m_{t+1} is a static input without dynamic implications (ACF), we expect no dependence between a firm's decisions on m_{t+1} and I_t. Furthermore, firms probably do not often observe their optimization errors; and if they find such an error, they are likely to respond by adjusting the inputs instead of their investment decisions. Thus, it seems reasonable to assume that the optimization errors in m_{t+1} and I_t are independent of each other, no matter whether the optimization error in m_{t+1} is independent across time or serially correlated. Meanwhile, the optimization errors of both m_{t+1} and I_t are unlikely to be related to the unanticipated technology shock or measurement errors of the output. Therefore, with ζ_t, ϵ_{mt+1} being optimization errors, it seems reasonable to assume that η_t, ζ_t and ϵ_{mt+1} are mutually independent.[13]

Unobserved idiosyncratic cost shocks Unobserved firm-level idiosyncratic cost shocks for static inputs can be important in some applications. In this case, the residual error in the m_{t+1} equation can capture the idiosyncratic cost shocks for static inputs. Our model assumptions and identification conditions would still hold in this case if the cost shocks do not affect firms' investment decisions or enter η_t, the residual error of the output equation. Thus, for this case, our identification requires that the idiosyncratic cost shocks are independent across time and data on the actual inputs are available to researchers.

Measurement error Measurement errors in the output, inputs and investment can arise in a number of ways. They can be caused simply by recording errors and/or by researchers' imperfect ways of computing the actual output/inputs/investment. For example, measurement errors in inputs can arise

[13]Gandhi et al. (2013) propose a method for identifying and estimating production functions by exploiting the first-order conditions in firms' static profit-optimization problems. Our paper complements theirs in providing an alternative method for applications in which firms' optimization errors might be important.

when we observe only the expenditures on the inputs, but not the actual inputs, and the input prices vary across firms; the measurement error in the capital input can arise due to the imperfect ways that we use to deal with capital stock depreciation and aggregating different types of capital inputs.[14] In addition, as LP point out, some intermediate inputs —such as materials and fuels—may be storable, and, thus, measurement errors can occur if the econometrician can observe only the new purchases of such inputs instead of the actual usage of them.

The conditional independence assumption may still hold for (y_t, I_t, m_{t+1}) if the residual errors capture only measurement errors that are *independent across time*. In this case, the residual error in the production function equation, η_t, can capture measurement errors in the output as well as in the inputs, which seem unlikely to be related to the measurement error in I_t. Suppose that the conditional independence conditions continue to hold despite the measurement errors in the inputs, what we identify through equation 6 is $f(y_t|k_t, l_t, m_t, u_t, \omega_t)$ (and, thus, $E(y_t|k_t, l_t, m_t, u_t, \omega_t)$) with (k_t, l_t, m_t, u_t) being the observed inputs instead of the actual inputs as in the structural production functions.

Thus, although our identification argument still holds if the mutually independent measurement errors are limited to the output or investment, it is not sufficient for the case with measurement errors in the input variables. The measurement errors in the input variables bias the estimates of their coefficients toward zero. Meanwhile, as we show in the next section, we may deal with the measurement errors in the inputs by incorporating instruments for the mismeasured inputs into our flexible GMM estimation method, assuming that the measurement errors in the inputs are independent across time.

In summary, the conditional independence assumption seems reasonable for (y_t, I_t, m_{t+1}) in many important cases. We get similar conclusions if there are multiple intermediate inputs or if we replace (y_t, I_t, m_{t+1}) with (y_t, I_t, u_{t+1}) or (y_t, I_t, y_{t+1}). In practice, researchers should pay close attention to the interpretations of the residual errors when assessing the validity of the assumptions.

2.4.2 The injectivity assumption

The part i) of condition 2 can be restrictive. For example, the distribution of ξ_{t+1} may depend on ω_t, because we have only $\mathbb{E}(\xi_{t+1}|\omega_t) = 0$ in our model. However, the conditioning on the covariates W_t makes the requirement less restrictive,

[14] We thank one of the referees for pointing out these important issues. Note that our identification works for the case in which the firm-specific input prices are missing but the actual inputs (instead the expenditures on the inputs) are available.

because it allows the distributions (and hence, for example, the variances) of η_t and $(\alpha_{m\omega}\xi_{t+1} + \epsilon_{mt+1})$ to depend on the covariates W_t.

The part ii) requirement of $\rho(\omega_t)$ being strictly monotonic says that a higher productivity today leads to a higher expected productivity tomorrow, which seems reasonable. In the case of $m_{t+1} = \mu_{mt+1}(k_{t+1}, \omega_{t+1}) + \epsilon_{mt+1}$ with $\mu_{mt+1}(.)$ being nonlinear, we also require that $\mu_{mt+1}(k_{t+1}, \omega_{t+1})$, equivalently $\mathbb{E}(m_{t+1}|k_{t+1}, \omega_{t+1})$, is strictly monotonic in ω_{t+1} for any given k_{t+1}. The condition is, in practice, less restrictive than the assumption of m_{t+1} being strictly monotonic in ω_{t+1} for any given k_{t+1}.

The part iii) is a technical assumption, which is equivalent to the conditional characteristic functions of η_t and $\alpha_{m\omega}\xi_{t+1} + \epsilon_{mt+1}$, residual errors in the y_t and m_{t+1} equations respectively, being nonvanishing on the real line. These conditions seem reasonable given that y_t and m_{t+1} are continuous variables.

It is worth noting that condition 2 is sufficient, but not necessary, to ensure the injectivity of the corresponding integral operators. Unfortunately, we are not aware of weaker primitive conditions that can guarantee the injectivity that we need in our identification proof.

2.4.3 The distinctive eigenvalues and the normalization

The distinctive eigenvalue condition requires that, for any fixed W_t and $\bar{\omega}_t \neq \tilde{\omega}_t$, $\iota(\bar{\omega}_t, W_t, \zeta_t)$ and $\iota(\tilde{\omega}_t, W_t, \zeta_t)$ have different distributions. This condition is relatively mild, because all it requires is that, ceteris paribus, any change in a firm's productivity has to lead to some change in the distribution of the firm's investment decisions. A sufficient, but not necessary, condition that implies the distinctive eigenvalue condition is $\mathbb{E}(I_t|\omega_t, W_t)$ being strictly increasing in ω_t for any given W_t (which is less restrictive than requiring that I_t itself being strictly increasing in ω_t for any given W_t). Lastly, the normalization assumption of $\mathbb{E}(\eta_t|\omega_t, l_t, k_t, m_t, u_t) = 0$ is standard in the literature.

3 Estimation

In light of the identification results above, one possible method of estimating the production function in equation (2) is Maximum Likelihood Estimation (MLE). Due to the presence of many functional nuisance parameters, the MLE approach is feasible but harder to implement in practice than the methods of OP/LP/ACF. We briefly describe the MLE approach in the Appendix, and refer interested readers to an earlier version of this paper (Huang and Hu (2011)) for more details on the approach. Our focus in this section will be on a straightforward GMM estimator that we propose. The GMM estimator is based on the same identification idea of using two proxy variables for the latent productivity, although the moment conditions that we use to derive the GMM estimator do not follow directly from the identification conditions introduced in Section 2.2.

Let us first rewrite the gross-output Cobb-Douglas production function in logs as:

$$\tilde{y}_t(\beta) = \omega_t + \eta_t, \tag{10}$$

where $\beta = (\beta_l, \beta_k, \beta_m, \beta_u)'$ and $\tilde{y}_t(\beta) \equiv y_t - (\beta_l l_t + \beta_k k_t + \beta_m m_t + \beta_u u_t)$. Likewise, we write the reduced-form demand functions for the static inputs of $x = m$ and u in logs as:

$$\tilde{x}_{t+1}(\alpha_x) = \alpha_{x\omega}\omega_{t+1} + \epsilon_{xt+1}, \tag{11}$$

where $\alpha_x = (\alpha_{x0}, \alpha_{xk})'$ and $\tilde{x}_{t+1}(\alpha_x) \equiv x_{t+1} - (\alpha_{x0} + \alpha_{xk}k_{t+1})$.

Assume that the productivity ω_t transitions according to the following AR(1) process:

$$\omega_t = \rho(\omega_{t-1}) + \xi_t = \sum_{p=1}^{P} \rho_p \omega_{t-1}^p + \xi_t. \tag{12}$$

Furthermore, we require the following moment conditions to derive our GMM estimator.

Condition 5. *The following moment independence conditions:*

$$\mathbb{E}\left(\begin{pmatrix} \epsilon_{xt+1} \\ \xi_{t+1} \end{pmatrix} | \omega_t, \tilde{z}_{td}\right) = 0, \tag{13}$$

$$\mathbb{E}(\eta_t^q | \omega_t, \tilde{z}_{td}) = \mathbb{E}(\eta_t^q), q = 1, ...P, \tag{14}$$

are satisfied for $\tilde{z}_{td} = I_t, l_t, k_t, m_t,$ *and* u_t.

The moment conditions in (13) implies that $\mathbb{E}(\alpha_{x\omega}\xi_{t+1} + \epsilon_{xt+1}|\omega_t, \tilde{z}_{td}) = 0$, for $\tilde{z}_{td} = I_t, l_t, k_t, m_t$, and u_t, which is a condition that we use directly in deriving our GMM estimator and is similar to the moment condition (2.12) in Wooldridge (2009). In the following, we use the moment conditions in (13) and (14) to derive our GMM estimator.

Let us denote $\tilde{z}_t \equiv (I_t, l_t, k_t, m_t, u_t)$. Then, by (10), (11), (12), (13) and (14), we have that, for $p = 1, ..., P$,

$$\text{cov}(\tilde{y}_t(\beta)^p, \tilde{z}_t) = \sum_{q=1}^{p} \binom{p}{q} \mathbb{E}(\eta_t^{p-q}) \text{cov}(\omega_t^q, \tilde{z}_t), \tag{15}$$

$$\text{and} \quad \text{cov}(\tilde{x}_{t+1}(\alpha_x), \tilde{z}_t) = \sum_{q=1}^{P} \varphi_q \text{cov}(\omega_t^q, \tilde{z}_t), \tag{16}$$

where $\varphi_q = \alpha_{x\omega}\rho_q$ for each $q = 1, \cdots, P$. We note that, because $\mathbb{E}(\eta_t) = 0$, for the cases of $P = 1$ or $P = 2$, (15) can be simply written as

$$\text{cov}(\tilde{y}_t(\beta)^p, \tilde{z}_t) \quad = \quad \text{cov}(\omega_t^p, \tilde{z}_t), \qquad \text{for } p = 1, 2,$$

which can be substituted into (16) to obtain moment restrictions of the following form:

$$\mathbb{E}\left[\ \tilde{z}_t \left(\tilde{x}_{t+1}(\alpha_x) - \textstyle\sum_{p=1}^{P} \varphi_p \tilde{y}_t(\beta)^p \right) \ \right] = 0, \qquad \text{for } P \leq 2$$

for any proxy $x = m$ or u. Taking these moment conditions to the GMM framework provides an estimate of $\theta = (\alpha_x', \alpha_{x\omega}, \beta', \varphi_1, \cdots, \varphi_P)'$.

In cases of $P > 2$, (15) can still be explicitly solved for $\text{cov}(\omega_t^q, \tilde{z}_t)$ for each $q = 1, \cdots, P$. Note that equation (15) can be written in matrix form as follows:

$$\begin{pmatrix} \text{cov}(\tilde{y}_t(\beta), \tilde{z}_t) \\ \text{cov}(\tilde{y}_t(\beta)^2, \tilde{z}_t) \\ \vdots \\ \text{cov}(\tilde{y}_t(\beta)^P, \tilde{z}_t) \end{pmatrix} = M(t, P) \begin{pmatrix} \text{cov}(\omega_t, \tilde{z}_t) \\ \text{cov}(\omega_t^2, \tilde{z}_t) \\ \vdots \\ \text{cov}(\omega_t^P, \tilde{z}_t) \end{pmatrix},$$

where $M(t, P)$ is a $P \times P$ lower triangular matrix defined as follows:

$$M(t, P) = \begin{pmatrix} 1 & & & \\ \binom{2}{1}\mathbb{E}[\eta_t] & 1 & & \\ \vdots & \vdots & \ddots & \\ \binom{P}{1}\mathbb{E}[\eta_t^{P-1}] & \binom{P}{2}\mathbb{E}[\eta_t^{P-2}] & \cdots & 1 \end{pmatrix}. \tag{17}$$

Since its diagonal elements are all non-zero, $M(t, P)$ is invertible. Let $M(t, P)^{-1}$ denote the inverse of $M(t, P)$, and let $[M(t, P)^{-1}]_{(q,p)}$ denote its (q, p)-th element. Then, we can solve for $\text{cov}(\omega_t^q, \tilde{z}_t)$ as follows:

$$\text{cov}(\omega_t^q, \tilde{z}_t) \ = \ \sum_{p=1}^{P} [M(t, P)^{-1}]_{(q,p)} \text{cov}(\tilde{y}_t(\beta)^p, \tilde{z}_t) \qquad \text{for each } q = 1, \cdots, P.$$

The above solution can be substituted into equation (16) to obtain moment restrictions with the following form:

$$\mathbb{E}\left[\ \tilde{z}_t \left(\tilde{x}_{t+1}(\alpha_x) - \textstyle\sum_{p=1}^{P} \tilde{\varphi}_p \tilde{y}_t(\beta)^p \right) \ \right] = 0, \tag{18}$$

for any proxy $x = m$ or u and instrument vector $\tilde{z}_t$, where $\tilde{\varphi}_p := \sum_{q=1}^{P} \varphi_q [M(t, P)^{-1}]_{(q,p)}$ for $p = 1, \cdots, P$. Taking the moment condition in (18) to the GMM framework provides an estimate of $\theta = (\alpha_x', \alpha_{x\omega}, \beta', \tilde{\varphi}_1, \cdots, \tilde{\varphi}_P)'$. The following are a few examples of the relationship between the scaled AR parameters $\varphi_p = \alpha_{x\omega}\rho_p$ and the reduced-form parameters $\tilde{\varphi}_p$:

1. $P = 1$: $\tilde{\varphi}_1 = \varphi_1$.
2. $P = 2$: $(\tilde{\varphi}_1, \tilde{\varphi}_2) = (\varphi_1, \varphi_2)$.
3. $P = 3$: $(\tilde{\varphi}_1, \tilde{\varphi}_2, \tilde{\varphi}_3) = (\varphi_1 - 3\sigma^2_{\eta_t}\varphi_3, \varphi_2, \varphi_3)$.

We make three observations here about the GMM estimator that we derive above. First, in deriving the moment restrictions above, although we need the additional moment-independence condition in (14), we make no use of the conditional independence assumptions of $\eta_t \perp\!\!\!\perp \omega_t | W_t$ and $(\alpha_{x\omega}\xi_{t+1} + \epsilon_{xt+1}) \perp\!\!\!\perp \omega_t | W_t$ (which are part of condition 2 we used in proving identification in Section 2). The mean-independence assumptions in (13) are standard in the literature. We test the robustness of our estimation method to minor violations of condition (14) through Monte Carlo experiments. Second, under the alternative assumption of the labor input being determined one period before the static inputs, we need simply include l_{t+1} in the x_{t+1} (with $x = m$ or u) equation and add l_{t+1} to $\tilde{z}_t$, the vector of instruments, in the estimation by GMM. Lastly, the above estimation method may be viewed as an extension of the IV approach (Blundell and Bond (2000)) in that we do not restrict the AR(1) process for productivity transition to be linear.

3.1 Extensions

A convenient feature of the above GMM approach is that we can add moment conditions if doing so improves statistical power. We get moment conditions similar to equation (18) if we replace (y_t, I_t, m_{t+1}) with (y_t, I_t, y_{t+1}). To improve efficiency, we may add the moment restrictions based on y_{t+1} in our estimation, given the following mean-independence condition:[15]

Condition 6. $\mathbb{E}(\xi_{t+1} + \eta_{t+1} | \omega_t, \tilde{z}_{td}) = 0$ *holds for* $\tilde{z}_{td} = I_t, l_t, k_t, m_t,$ *and* u_t.

Recall that $\tilde{z}_t \equiv (I_t, l_t, k_t, m_t, u_t)$. Then, we can use the following augmented set of moment restrictions in estimation:

$$\mathbb{E}\left[\begin{array}{c} \tilde{z}_t\left(\tilde{y}_{t+1}(\alpha_x) - \sum_{p=1}^{P} \alpha_{x\omega}^{-1}\tilde{\varphi}_p \tilde{y}_t(\beta)^p\right) \\ \tilde{z}_t\left(\tilde{x}_{t+1}(\alpha_x) - \sum_{p=1}^{P} \tilde{\varphi}_p \tilde{y}_t(\beta)^p\right) \end{array}\right] = 0,$$

where $\tilde{\varphi}_p := \sum_{q=1}^{P} \varphi_q [M(t,P)^{-1}]_{(q,p)}$.

[15]The mean-Independence condition is common in the literature (e.g., condition (2.12) in Wooldridge (2009)).

Furthermore, if one is concerned about classical measurement errors in the inputs and the measurement errors are *independent across time*, we may replace the mis-measured inputs with their one-period lagged values in $\tilde{z}_t$ to get consistent estimates of the production-function parameters.[16] For example, suppose that the main concern is classical measurement errors in the material input. Let m_t denotes the observed material input, which measures the actual material input m_t^* with error—that is, $m_t = m_t^* + \tilde{\epsilon}_{mt}$, where $\tilde{\epsilon}_{mt}$ is the measurement error.[17] In this case, maintaining all our original notations, we would have both the residual error ϵ_{mt} in the m_t equation and η_t in the y_t equation (partly) capture the measurement error $\tilde{\epsilon}_{mt}$. As a result, the moment conditions in (14) and the covariance equation (15) do not hold if $\tilde{z}_t$ includes m_t. However, if the measurement error in m_t is independent across time, then the moment conditions in (13) and (14) and the covariance equations (15) and (16) would hold if we replace the m_t in $\tilde{z}_t$ with m_{t-1}. Thus, we may estimate the model parameters using the following moment conditions:

$$\mathbb{E}\left[\ \tilde{z}_t\left(\tilde{x}_{t+1}(\alpha_x) - \sum_{p=1}^P \tilde{\varphi}_p \tilde{y}_t(\beta)^p\right)\ \right] = 0,$$

where $\tilde{z}_t = (I_t, l_t, k_t, m_{t-1}, u_t)$ (instead of $\tilde{z}_t = (I_t, l_t, k_t, m_t, u_t)$).[18]

Lastly, the above estimation method can also be extended to accomondate the following first-order controlled Markov process for productivity:

$$\omega_{t+1} = \sum_{p=1}^P \rho_{1p}\omega_t^p + \sum_{p=1}^P \rho_{2p} r_t^p + \sum_{p=1}^P \sum_{q=1}^P \rho_{3pq}\omega_t^p r_t^q + \xi_{t+1},$$

where r_t is the firm's expenditure on research and development (R&D) in period t. In this case, we have:

$$\begin{aligned}\text{cov}\left(\tilde{x}_{t+1}(\alpha_x), \tilde{z}_t\right) = \alpha_{x\omega}\sum_{p=1}^P \rho_{1p}\text{cov}\left(\omega_t^p, \tilde{z}_t\right) + \alpha_{x\omega}\sum_{p=1}^P \rho_{2p}\text{cov}\left(r_t^p, \tilde{z}_t\right) + \\ \alpha_{x\omega}\sum_{p=1}^P\sum_{q=1}^P \rho_{3pq}\text{cov}\left(\omega_t^p r_t^q, \tilde{z}_t\right).\end{aligned} \tag{19}$$

[16]il Kim et al. (2016) allow for measurement errors in the inputs in the method they propose for estimating production functions within the modeling framework of Olley and Pakes (1996) and Levinsohn and Petrin (2003). Their method combines sieve MLE in the first step and GMM as in Wooldridge (2009) in the second step.

[17]We get the specification of measurement error if, for example, we observe $log(M_t P_{mt})$ but not $log(M_t)$.

[18]It is worth emphasizing here that replacing the mis-measured inputs in $\tilde{z}_t$ with their lagged values would not work if the measurement errors in the inputs are serially correlated.

Suppose that $\mathbb{E}(\eta_t^q|\omega_t,\tilde{z}_t,r_t) = \mathbb{E}(\eta_t^q)$, for $q = 1,...,P$. Then, for any given $p,q \leq P$, we have:

$$\text{cov}\left(\tilde{y}_t(\beta)^p r_t^q, \tilde{z}_t\right) = \sum_{j=0}^{p} \begin{pmatrix} p \\ j \end{pmatrix} \mathbb{E}\left(\eta_t^{p-j}\right) \text{cov}\left(\omega_t^j r_t^q, \tilde{z}_t\right).$$

We can solve the above equation for $\text{cov}\left(\omega_t^p r_t^q, \tilde{z}_t\right)$ as follows:

$$\text{cov}\left(\omega_t^p r_t^q, \tilde{z}_t\right) = \sum_{j=1}^{P} \left[M\left(t,P\right)^{-1}\right]_{(p,j)} \left(\text{cov}\left(\tilde{y}_t(\beta)^j r_t^q\right) - \mathbb{E}\left(\eta^j\right) \text{cov}\left(r_t^q, \tilde{z}_t\right)\right),$$

where $M\left(t,P\right)$ is just the invertible matrix defined above in (17). In addition, recall that $\text{cov}\left(\omega_t^p, \tilde{z}_t\right) = \sum_{q=1}^{P} \left[M\left(t,P\right)^{-1}\right]_{(p,q)} \text{cov}\left(\tilde{y}_t(\beta)^q, \tilde{z}_t\right)$. Substituting the solutions for $\text{cov}\left(\omega_t^p r_t^q, \tilde{z}_t\right)$ and $\text{cov}\left(\omega_t^p, \tilde{z}_t\right)$ into equation (19), we get:

$$\begin{aligned}\text{cov}\left(\tilde{x}_{t+1}(\alpha_x), \tilde{z}_t\right) = \sum_{q=1}^{P} \tilde{\rho}_{1q} \text{cov}\left(\tilde{y}_t(\beta)^q, \tilde{z}_t\right) + \sum_{q=1}^{P} \tilde{\rho}_{2q} \text{cov}\left(r_t^q, \tilde{z}_t\right) + \\ \sum_{j=1}^{P}\sum_{q=1}^{P} \tilde{\rho}_{3jq} \text{cov}\left(\tilde{y}_t(\beta)^j r_t^q, \tilde{z}_t\right),\end{aligned}$$

where

$$\begin{aligned}\tilde{\rho}_{1q} &= \alpha_{x\omega} \sum_{p=1}^{P} \rho_{1p} \left[M\left(t,P\right)^{-1}\right]_{(p,q)}, \\ \tilde{\rho}_{2q} &= \alpha_{x\omega} \left(\rho_{2q} - \sum_{p=1}^{P} \rho_{3pq} \sum_{j=1}^{P} \left[M\left(t,P\right)^{-1}\right]_{(p,j)} \mathbb{E}\left(\eta^j\right)\right), \\ \tilde{\rho}_{3jq} &= \alpha_{x\omega} \sum_{p=1}^{P} \rho_{3pq} \left[M\left(t,P\right)^{-1}\right]_{(p,j)}.\end{aligned}$$

Then, we can transform the above covariance equality into the following moment condition:

$$\mathbb{E}\left(\tilde{z}_t \left(\tilde{x}_{t+1}(\alpha_x) - \sum_{q=1}^{P} \tilde{\rho}_{1q} \tilde{y}_t(\beta)^q - \sum_{q=1}^{P} \tilde{\rho}_{2q} r_t^q - \sum_{j=1}^{P}\sum_{q=1}^{P} \tilde{\rho}_{3jq} \tilde{y}_t(\beta)^j r_t^q\right)\right) = 0,$$

which we can use to estimate the production-function parameters in the GMM framework if we observe r_t.

3.2 The GMM Estimator and Its Asymptotic Properties

We prove the asymptotic properties of our GMM estimator in the subsection. The moment restrictions (18) may be written as $\mathbb{E}[g_t(\theta)]$, where

$$g_t(\theta) = \tilde{z}_t \left(\tilde{x}_{t+1}(\alpha_x) - \sum_{p=1}^{P} \tilde{\varphi}_p \tilde{y}_t(\beta)^p \right)$$

For a suitable weighting matrix $\hat{W}$, the generalized method of moments (GMM) estimator $\hat{\theta}$ for the true parameter vector θ_0 is defined by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{2} \mathbb{E}_n[g_t(\theta)]' \hat{W} \mathbb{E}_n[g_t(\theta)]$$

where $\mathbb{E}_n$ denotes the cross-sectional sample mean operator. The variance of $\sqrt{n}(\hat{\theta} - \theta_0)$ is approximated by

$$\hat{V} = (\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{\Sigma}\hat{W}\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}$$

where $\hat{G} = \mathbb{E}_n[D_\theta g_t(\hat{\theta})]$ is an estimator for $G = \mathbb{E}[D_\theta g_t(\theta_0)]$ and $\hat{\Sigma} = \mathbb{E}_n[g_t(\hat{\theta}) g_t(\hat{\theta})']$ is an estimator for $\hat{\Sigma} = \mathbb{E}[g_t(\theta_0) g_t(\theta_0)']$. To guarantee that the GMM estimator and its variance estimator behave well in large sample, we make the following assumption.

Assumption 1. *(i) The sample is i.i.d. (ii) $\hat{W} \xrightarrow{p} W$, which is positive definite. (iii) θ_0 is in the interior of Θ, which is compact. (iv) $\tilde{z}_t$, x_{t+1}, and k_{t+1} have bounded second moments, and y_t, l_t, k_t, m_t, and u_t have bounded $2P$-th moments. (v) $\tilde{z}_t$, x_{t+1}, and k_{t+1} have bounded fourth moments, and y_t, l_t, k_t, m_t, and u_t have bounded $4P$-th moments.*

Theorem 2. *If Assumption 1 (i), (ii), (iii), (iv) is satisfied, then the following result holds:*
(I) $\hat{\theta} \xrightarrow{p} \theta_0$.
If Assumption 1 (i), (ii), (iii), (v) is satisfied, then the following results hold:
(II) $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1})$; and
(III) $\hat{V} \xrightarrow{p} (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1}$.

Proof. We prove the theorem by checking the conditions of Newey and McFadden (NM, 1994).

(I) The identification and Assumption 1 (ii) satisfy condition (i) of Theorem 2.6 in NM. Assumption 1 (iii) satisfies condition (ii) of Theorem 2.6 in NM. The functional form of our g_t and Assumption 1 (iii) satisfy condition (iii) of Theorem 2.6 in NM. By Hölder's inequality, the functional form of our g_t and Assumption 1 (iii), (iv) satisfy condition (iv) of Theorem 2.6 in NM. Therefore, $\hat{\theta} \xrightarrow{p} \theta_0$ by Theorem 2.6 in NM.

(II) Assumption 1 (iii) satisfies condition (i) of Theorem 3.4 in NM. The functional form of our g_t and Assumption 1 (iii) satisfy condition (ii) of Theorem 3.4 in NM. By Hölder's inequality, the functional form of our g_t and Assumption 1 (iii), (v) satisfy conditions (iii) and (iv) of Theorem 3.4 in NM. The identification and Assumption 1 (ii) satisfy condition (v) of Theorem 3.4 in NM. Therefore, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1})$ by Theorem 3.4 in NM.

(III) By Hölder's inequality, the functional form of our g_t and Assumption 1 (iii), (v) satisfy the condition of Theorem 4.5 in NM in addition to those of Theorem 3.4 in NM. □

For convenience of readers, we present the estimation and inference procedure based on the above theory as an algorithm.

Algorithm 1.

1. *Compute the first-step estimate* $\hat{\theta}_I = \arg\min_{\theta \in \Theta} \mathbb{E}_n[g_t(\theta)]'\mathbb{E}_n[g_t(\theta)]$.
2. *Compute the estimated variance matrix* $\hat{\Sigma}_I = \mathbb{E}_n[g_t(\hat{\theta}_I)g_t(\hat{\theta}_I)']$
3. *Compute the second-step estimate* $\hat{\theta}_{II} = \arg\min_{\theta \in \Theta} \mathbb{E}_n[g_t(\theta)]'\hat{\Sigma}_I^{-1}\mathbb{E}_n[g_t(\theta)]$.
4. *Compute the estimated second-step variance matrix* $\hat{V}_{II} = (\hat{G}_{II}'\hat{\Sigma}_I^{-1}\hat{G}_{II})^{-1}$ $\hat{G}_{II}'\hat{\Sigma}_I^{-1}\hat{\Sigma}_{II}\hat{\Sigma}_I^{-1}\hat{G}_{II}$ $(\hat{G}_{II}'\hat{\Sigma}_I^{-1}\hat{G}_{II})^{-1}$*, where* $\hat{G}_{II} = \mathbb{E}_n[D_\theta g_t(\hat{\theta}_{II})]$ *and* $\hat{\Sigma}_{II} = \mathbb{E}_n[g_t(\hat{\theta}_{II})g_t(\hat{\theta}_{II})']$.
5. *Report estimation and inference results based on* $\hat{\theta}_{II}$ *and* $\hat{V}_{II}$.

3.3 A Test of the Model of OP/LP/Wooldridge

We propose a test of the model of OP/LP to help researchers choose between the model of OP/LP and the extended model that we propose in this paper. LP use a nonparametric function $\mu_{mt}^{-1}(m_t, k_t)$ to control for ω_t (see the appendix for the OP/LP procedure in detail). Thus, the first-step estimating equation of LP is:

$$y_t = \beta_l l_t + \tilde{\phi}_t(m_t, k_t) + \eta_t,$$

where $\tilde{\phi}_t(m_t, k_t) \equiv \beta_k k_t + \beta_m m_t + \mu_{mt}^{-1}(m_t, k_t)$ is unknown and needs to be specified nonparametrically. One can similarly consider the first-step estimating equation of OP, where I_t is used as a control variable. From the ACF critique, β_l

in this first-step estimating equation is unidentified in the model of OP/LP due to functional dependence: l_t is determined as a function of (ω_t, k_t) and, hence, a function of (m_t, k_t), because $\omega_t = \mu_{mt}^{-1}(m_t, k_t)$. Thus, the standard 95% confidence set for β_l should contain all real values. We propose to use this argument based on the ACF critique to construct a test of the functional dependence, an implication of the model of OP/LP.

Following the convention (LP/Woodridge), we use an ι-dimensional linear-in-parameter approximation to the control function $\tilde{\phi}_t$ by a parameter vector ν, i.e., $\tilde{\phi}_t(m_t, k_t) = \nu' v(m_t, k_t)$ for some basis $v(m_t, k_t)$. The first-step moment function is written as

$$f_i(\beta_l, \nu) = (l_{it}, v(m_{it}, k_{it})')'(y_{it} - \beta_l l_{it} - \nu' v(m_{it}, k_{it})),$$

which is $(1 + \iota) \times 1$ dimensional vector-valued. Let the vectorized gradient of $f_i(\beta_l, \nu)$ be denoted by:

$$q_i(\beta_l, \nu) = \begin{pmatrix} -(l_{it}, v(m_{it}, k_{it})')' l_{it} \\ -(l_{it}, v(m_{it}, k_{it})')' \nu_1(m_{it}, k_{it}) \\ \vdots \\ -(l_{it}, v(m_{it}, k_{it})')' \nu_\iota(m_{it}, k_{it}) \end{pmatrix},$$

which is $(1 + \iota)^2 \times 1$ dimensional vector-valued. The joint variance matrix of $(f_i(\beta_l, \nu)', q_i(\beta_l, \nu)')'$ is denoted by

$$\begin{pmatrix} V_{ff}(\beta_l, \nu) & V_{f\theta}(\beta_l, \nu) \\ V_{\theta f}(\beta_l, \nu) & V_{\theta\theta}(\beta_l, \nu) \end{pmatrix} = \operatorname{Var}\begin{pmatrix} f_i(\beta_l, \nu) \\ q_i(\beta_l, \nu) \end{pmatrix}.$$

The projected score is denoted by

$$\hat{D}_n(\beta_l, \nu) = \Bigg[n^{-1} \sum_{i=1}^{n} \Big(q_{i,1}(\beta_l, \nu) - \hat{V}_{\theta f,1}(\beta_l, \nu) \hat{V}_{ff}(\beta_l, \nu)^{-1} f_i(\beta_l, \nu) \Big) \\ ..., n^{-1} \sum_{i=1}^{n} \Big(q_{i,\iota+1}(\beta_l, \nu) - \hat{V}_{\theta f,\iota+1}(\beta_l, \nu) \hat{V}_{ff}(\beta_l, \nu)^{-1} f_i(\beta_l, \nu) \Big) \Bigg].$$

With these notations, we can write the concentrated K-statistic (Kleibergen

(2005)) as

$$K(\beta_l^*) = n\left(n^{-1}\sum_{i=1}^{n} f_i(\beta_l^*, \nu(\beta_l^*))'\hat{V}_{ff}(\beta_l^*, \nu(\beta_l^*))^{-1}\hat{D}_n(\beta_l^*, \nu(\beta_l^*))\right)$$
$$\times\left(\hat{D}_n(\beta_l^*, \nu(\beta_l^*))'\hat{V}_{ff}(\beta_l^*, \nu(\beta_l^*))^{-1}\hat{D}_n(\beta_l^*, \nu(\beta_l^*))\right)^{-1}$$
$$\times\left(n^{-1}\sum_{i=1}^{n} f_i(\beta_l^*, \nu(\beta_l^*))'\hat{V}_{ff}(\beta_l^*, \nu(\beta_l^*))^{-1}\hat{D}_n(\beta_l^*, \nu(\beta_l^*))\right)'.$$

Under the following assumption, this statistic can be used to test the hypothesis $H_0 : \beta_l = \beta_l^*$ *robustly without assuming that the labor coefficient* β_l *is identified.*

Assumption 2. *(i)* ν *belongs to the interior of a compact parameter set. (ii)* $(y_{it}, l_{it}, v(m_{it}, k_{it})')'$ *has bounded fourth moments. (iii)* $((l_{it}, v(m_{it}, k_{it})')'\nu_1(m_{it}, k_{it})$, ..., $(l_{it}, v(m_{it}, k_{it})')'\nu_\iota(m_{it}, k_{it}))$ *has a full rank* ι*.*

Theorem 3. *If Assumption 1 (i) and Assumption 2 are satisfied, then* $K(\beta_l^*) \xrightarrow{d} \chi^2(1)$ *under the null hypothesis* $H_0 : \beta_l = \beta_l^*$*.*

Proof. Assumption 1 (i) and Assumption 2 (i)–(ii) imply that Assumption 1 of Kleibergen (2005) is satisfied by Lindeberg-Lévy CLT. Similarly, Assumption 1 (i) and Assumption 2 (i)–(ii) imply that Assumption 2 of Kleibergen (2005) is satisfied by the weak law of large numbers. Assumption 2 (iii) implies that Assumption 2 of Kleibergen (2005) is satisfied by the definition of f_i. □

Consider the null hypothesis $H_0 : \beta_l = \beta_l^*$ where β_l^* is set to a negative number (for example). If the model of OP/LP is true, then the functional dependence property pointed out by ACF implies that this null hypothesis cannot be rejected. Therefore, if the test based on the K statistic rejects such a null hypothesis, then we can take it as an evidence against the model of OP/LP. We apply the test in our Monte Carlo experiments and empirical application ahead. Based on this testing procedure, we indeed find that the empirical data does *not* support the model of OP/LP (see Section 4.1).

3.4 Monte Carlo Experiments

We consider the following data-generating process (DGP) for simulating data. The gross-output Cobb-Douglas production function in logs is given by

$$y_t = \beta_l l_t + \beta_k k_t + \beta_m m_t + \beta_u u_t + \omega_t + \eta_t, \qquad \eta_t \sim N(0, s_\eta^2),$$

where $(\beta_l, \beta_k, \beta_m, \beta_u) = (0.4, 0.3, 0.2, 0.1)$, and $s_\eta = 1$. The productivity level ω_t follows a linear AR(1) process

$$\omega_{t+1} = \rho_1 \omega_t + \xi_{t+1}, \qquad \xi_{t+1} \sim N(0, s_\xi^2),$$

where $\rho_1 = 1.00$ and $s_\xi = 0.05$. The capital accumulates according to the following law of motion:

$$K_{t+1} = (1 - \delta)K_t + 0.5I_t + 0.5I_{t-1},$$

where $\delta = 0.1$ and the reduced-form investment policy is specified as:

$$log(I_t) = -0.02k_t - 0.01i_{t-1} + 1.00\omega_t + \zeta_t, \qquad \zeta \sim N(0, s_\zeta^2),$$

for $s_\zeta = 1.00$. The static input choices are determined as the solution to the profit-maximization problem:

$$\max_{L_t, M_t, U_t} \mathbb{E} \exp(\eta_t) \exp(\omega_t) L_t^{\beta_l} K_t^{\beta_k} M_t^{\beta_m} U_t^{\beta_u} - (p_l L_t + p_m M_t + p_u U_t),$$

where the input prices $(p_l, p_m, p_u) = (0.3, 0.2, 0.1)$. This problem yields linear reduced-form input choice rules as in (11). For $x_t = m_t$ for example, it holds with the reduced-form parameters $\alpha_{mk} = \frac{\beta_k}{1-\beta_l-\beta_m-\beta_u}$ and $\alpha_{m\omega} = \frac{1}{1-\beta_l-\beta_m-\beta_u}$. Thus, the scaled AR parameter takes the value of $\varphi_1 = \frac{\rho_1}{1-\beta_l-\beta_m-\beta_u} = 3\frac{1}{3}$. We will refer to the DGP described here later as the baseline DGP when differentiating DGPs that deviate from it.

To avoid arbitrary initial conditions, we simulate the above model for ten periods and use the last two periods to estimate the parameters (following ACF). Estimation results with $x_{t+1} = m_{t+1}$ based on 2,500 simulated random samples are reported in Table 1. Similarly, Table 2 shows estimation results with $x_{t+1} = y_{t+1}$, and Table 3 shows results using both m_{t+1} and y_{t+1}. The latter two settings allow us to directly identify the AR parameter ρ_1 instead of only the reduced-form parameters $\tilde{\varphi}_1 (= \varphi_1)$ (as in the first setting). The difference is because ω_{t+1} enters the y_{t+1} equation directly, but enters the m_{t+1} equation linearly as $\alpha_{m\omega}\omega_{t+1}$.

The estimates in Tables 1-3 show that we can obtain consistent estimates using the moment conditions based on either (y_t, I_t, m_{t+1}) or (y_t, I_t, y_{t+1}); and using moment conditions based on $(y_t, I_t, m_{t+1}, y_{t+1})$ produces consistent estimates with smaller variances than either of the two prior cases (and we get essentially the same results when we use higher-order polynomials for $\rho(\omega_t)$). Overall, all these simulation results show the effectiveness of our estimation strategies.

We next compare our estimates with those produced using existing methods, including the ordinary least squares (OLS) method and the GMM method proposed by Wooldridge (2009) (which efficiently implements the estimation strategy of OP and LP).[19] Table 4 reports the estimates, based on the same simulated data as above, using the methods of OLS, Wooldridge's (W), LP's and ours (HHS). The OLS estimates have the largest root mean square errors (RMSEs), which do not diminish with sample size. Wooldridge's and LP's methods reduce the RMSEs relative to OLS, but the RMSEs do not decrease with sample size either. The estimates of OLS, Wooldridge's and LP's show upward bias for the coefficients of static inputs, (l_t, m_t, u_t), and OLS (Wooldridge's and LP's) estimates show downward (small upward) biases for the coefficient of capital, k_t. In contrast, the RMSEs of our estimates diminish toward zero with sample size.[20] The RMSEs of the estimates of OLS and Wooldridge's change little as the sample size increases, because the standard deviations are significantly smaller than the magnitude of the biases and RMSEs are dominated by the persistent biases in the estimates. Therefore, in the case in which the scalar-unobservable assumption is violated, the methods of OP/LP/Wooldridge reduce biases in the estimates but do not eliminate them, whereas our method is able to produce consistent estimates by exploiting two proxy variables.

Recall that we impose the moment restriction (14) for our GMM estimator, which is not used in the existing methods of production function estimation. To examine how our estimator behaves under a violation of this condition, and compare the bias of our estimator with those of the existing methods, we simulate data using a DGP that is the same as the baseline DGP above, except that η is now generated heterosketastically according to $\eta_t \sim N(0, s_\eta^2(1 + (\omega_t/2)^2))$. Such a DGP entails a violation of the moment restriction (14). Table 5 reports the estimates using the methods of OLS, Wooldridge's (W), LP's and ours (HHS). The biases of our estimator are significantly smaller than those of the OLS, Wooldridge's and LP's, showing evidence that our estimator still performs better than the existing estimators even under a violation of the assumption (14).

We next illustrate our method for the case with a nonlinear transition equation for the productivity ω_t. In particular, let us modify the baseline DGP by

[19] As in our method, we use $x = m$ as a proxy. Polynomial sieve control function of degree three is employed—i.e., $\gamma' c_t = (\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{20}, \gamma_{11}, \gamma_{02}, \gamma_{30}, \gamma_{22}, \gamma_{12}, \gamma_{03})(1, k_t, m_t, k_t^2, k_t m_t, m_t^2, k_t^3, k_t^2 m_t, k_t m_t^2, m_t^3)'$. Following Wooldridge (2009), we use the restrictions $\mathbb{E}[(1, l_t, u_t, l_{t-1}, u_{t-1}, c_t, c_{t-1})'(\tilde{y}(\beta) - \gamma' c_t)] = 0$ and $\mathbb{E}[(1, k_t, l_{t-1}, u_{t-1}, c_{t-1})'(\tilde{y}(\beta) - \rho_0 - \rho_1 \gamma' c_{t-1})] = 0$. The two-step GMM is used for estimation.

[20] For the estimates in Table 1, the $\sqrt{N}$ convergence rate for large samples does not seem to start until $N = 4000$.

assuming the following quadratic AR(1) process for productivity:

$$\omega_{t+1} = \rho_1 \omega_t + \rho_2 \omega_t^2 + \xi_{t+1}, \qquad \xi_{t+1} \sim N(0, s_\xi^2),$$

where $\rho_1 = 1.000$, $\rho_2 = -0.025$, and $s_\xi = 0.050$. We simulate the model similarly for ten periods and use the last two periods for estimation. Table 6 reports the estimates using moment conditions based on $(y_t, I_t, m_{t+1}, y_{t+1})$. It shows that the estimates assuming $P = 1$ have persistent biases even under a large sample, whereas those with $P = 2, 3$ have biases vanishing as the sample size increases. Although the biases in the case with $P = 1$ are in the same direction as OLS estimates, the magnitude of the biases are significantly smaller than those of the OLS estimates. Note that, with the mis-specification of $P = 1$, the root-mean-square error (RMSE) (unlike the standard deviation) does not converge at the rate of $\sqrt{N}$ because of the bias. Hence, with a sufficiently flexible specification for $\rho(\omega_t) = \mathbb{E}(\omega_{t+1}|\omega_t)$, our method produces consistent estimates; and, even with a linear specification for $\rho(\omega_t)$, our method still helps reduce bias in the estimates relative to the OLS estimates.

The baseline DGP focuses on the unit-root process for the productivity, but our method does not rely on the unit-root process. To demonstrate the robustness of our method against alternative AR(1) specifications, we present Monte Carlo simulation results under sub-unit-root AR(1) process of the transition of the productivity. Specifically, we set $\rho_1 = 0.95$, as opposed to $\rho_1 = 1.00$ as in the baseline DGP. Table 7 reports the estimates. The estimates demonstrate the consistency of our GMM estimator under this alternative DGP, showing that our method does not rely on the unit root assumption for the productivity transition process.

Finally, we report Monte Carlo studies for the test of identifiability of β_l in the first step of OP/LP's estimation procedure, following the theory proposed in Section 3.3. The test of identifiability is based on the null hypothesis $H_0 : \beta_l = \beta_l^*$, with β_l^* being a fixed finite value. Figure 1 shows power curves for the nominal size of 0.05 over $-0.6 \le \beta_l^* \le 0.6$ under various sample sizes of $N = 100$, 200, 400 and 800. The simulation size approaches the nominal size 0.05 around $\beta_l^* = 0.49$, consistent with the estimates of β_l reported for LP method in Table 4. Besides this location of β_l^*, the power curves increase to one as the sample size becomes larger, showing the consistency of the test. Therefore, the proposed test rejects the null of $H_0 : \beta_l = \beta_l^*$, for β_l^* being any negative value (for example), and thus correctly rejects the model of OP/LP.

4 Empirical Application: The Input Decisions and Productivity Shocks

In the following, we apply our method to the Chilean manufacturing data that LP use in their paper. We first present our estimates to illustrate the performance of our method with real data. Then, we study empirically how quickly firms adjust their inputs in response to the latest changes in their productivity. The analysis helps us to understand how efficiently firms in the industry operate and to identify potential frictions in the input markets. Methodologically, the analysis can provide guidance for choosing proxies for the latent productivity and help explain differences in the estimates using various methods.

4.1 Estimates of the Production Function

We apply our estimation method to industry ISIC 311 (the industry of food products), which has the most observations, in the Chilean manufacturing data. Following LP, we estimate a gross-output production function. The inputs include two types of labor inputs (high-skill and low-skill, l_t^s and l_t^u, respectively), capital (k_t), material (m_t), electricity (e_t) and fuel (u_t). We use the moment conditions based on $(y_t, I_t, x_{t+1}, y_{t+1})$, where x is one of the three inputs of (m, e, u), in our estimation. Following LP (2003), we include fixed effects for the three time periods, 1979–1981, 1982–1983, and 1984–1986, and we use d_t^1 and d_t^2 to denote time-period dummies for the latter two of the three periods. This setup yields six structural parameters $\beta = (\beta_{lu}, \beta_{ls}, \beta_k, \beta_m, \beta_e, \beta_u)'$ and two coefficients $\beta_d = (\beta_{d^1}, \beta_{d^2})'$ of the time dummies (d_t^1, d_t^2) for the production function equation (10); five reduced-form parameters in $(\alpha_x', \alpha_{x\omega})' = (\alpha_{x0}, \alpha_{xk}, \alpha_{xls}, \alpha_{xlu}, \alpha_{x\omega})'$ and two coefficients $\alpha_{xd} = (\alpha_{xd^1}, \alpha_{xd^2})'$ of the time dummies (d_t^1, d_t^2) for equation (11) of x_{t+1}, and $P+1$ reduced-form parameters $\tilde{\varphi} = (\tilde{\varphi}_0, ..., \tilde{\varphi}_P)$ for the AR(1) transition process of ω.[21] Thus, we have a total of $16 + P$ unknown parameters.

Before proceeding with estimation of the production function, we first test the model validity of OP/LP based on the method presented in Section 3.3. To this end, we compute the K-statistic on a grid of points for the parameter subvector $(\beta_{lu}^*, \beta_{ls}^*)$ of labor coefficient values. Figure 2 illustrates the region where the test fails to reject the null hypothesis $H_0 : (\beta_{lu}, \beta_{ls}) = (\beta_{lu}^*, \beta_{ls}^*)$. Recall that the model of OP/LP entails the functional dependence in the first steps of their estimation procedures, as pointed out by ACF. Hence, if their model were true, then the test would fail to reject such a null hypothesis globally. However, the

[21] We include a constant in the transition equation for ω here to make the specification more flexible for the real data.

results illustrated in Figure 2 imply otherwise. We take this finding as an evidence against the validity of the model of OP/LP for the food-product industry in the Chilean manufacturing data.

We now turn to our estimation procedure. We use the two-step GMM procedure, as described at the end of Section 3.2, in estimation. Separately for $x = m$, e, and u, we obtain the following 20 moment restrictions:

$$\mathbb{E}\left[\begin{array}{l} \tilde{z}_t\left(\tilde{y}_{t+1}(\beta', \beta'_d) - \sum_{p=0}^{P} \alpha_{x\omega}^{-1}\tilde{\varphi}_p\tilde{y}_t(\beta', \beta'_d)^p\right) \\ \tilde{z}_t\left(\tilde{x}_{t+1}(\alpha'_x, \alpha'_{xd}) - \sum_{p=0}^{P} \tilde{\varphi}_p\tilde{y}_t(\beta', \beta'_d)^p\right) \end{array}\right] = 0, \tag{20}$$

where $\tilde{z}_t = (1, I_t, l_t^s, l_t^u, k_t, m_t, e_t, u_t, d_t^1, d_t^2)'$, $\tilde{y}_t((\beta', \beta'_d)') = y_t - (\beta_{lu}l_t^u + \beta_{ls}l_t^s + \beta_k k_t + \beta_m m_t + \beta_e e_t + \beta_u u_t + \beta_{d^1}d_t^1 + \beta_{d^2}d_t^2)$ and $\tilde{x}_{t+1}((\alpha'_x, \alpha'_{xd})') = x_{t+1} - (\alpha_{x0} + \alpha_{xk}k_{t+1} + \alpha_{xls}l_{t+1}^s + \alpha_{xlu}l_{t+1}^u + \alpha_{xd^1}d_{t+1}^1 + \alpha_{xd^2}d_{t+1}^2)$. Note that, although we work with the assumption of the labor inputs being static inputs, we allow them to be possibly dynamic inputs by including the labor inputs, l_{t+1}^s and l_{t+1}^u, in the x_{t+1} ($x = m, e$ and u) equations. The vector of instruments $\tilde{z}_t$ does not include l_{t+1}^s and l_{t+1}^u, because, under the working assumption, l_{t+1}^s and l_{t+1}^u are correlated with ξ_{t+1} in the x_{t+1} equations.

We consider the cases of $P = 1$, 2, and 3 for the AR(1) process for ω. The standard errors are computed using the covariance formula of the asymptotic distribution of the two-step GMM procedure. To deal with potential problems of local optimums, we use 125 different initial points for numerical optimization and report the optimal interior estimates. More flexible specifications of the AR(1) process (i.e., with $P > 3$) do not produce any significant changes in the estimates of the structural parameters.[22] In a note on implementing the LP's estimation procedure, Petrin et al. (2004) (p.116) also suggests choosing $P = 3$ for the AR(1) process.

Table 8 presents our estimation results for different choices of x_{t+1} and P. As a reference, we copy, in the table, the estimates from Table 3 in LP, for which they use materials as the proxy for productivity. We also report the estimates using the GMM approach of Wooldridge (2009) and one of the static inputs (m_t, e_t and u_t) as a proxy for productivity (W-LP). The estimate of return-to-scale (RTS) is computed as the sum of the estimates of all the β coefficients in the production function.

For each choice of x_{t+1}, the differences in the estimates of the production-

[22] A common practice in applied research is to choose P by increasing it one by one and stopping when further increasing P does not bring significant changes in the estimates of the structural parameters. Hu and Schennach (2008) (p.206) also suggests similar informal guidelines for determining the smoothing parameters.

function parameters with $P = 1, 2$ and 3 are small, and the estimates of the $\tilde{\varphi}_2$ and $\tilde{\varphi}_3$ are also relatively small (except for the case with $x_{t+1} = m_{t+1}$ and $P = 3$).

The main difference in the estimates with the three different choices of x_{t+1} is in β_m and, consequently, in the RTS. With $x_{t+1} = m_{t+1}$, the point estimates of β_m range from 0.354 to 0.369, and those of the RTS range from 0.892 to 0.978. In comparison, with $x_{t+1} = e_{t+1}$ or u_{t+1}, the point estimates of β_m range from 0.636 to 0.673, and those of the RTS range from 1.153 to 1.386. Meanwhile, our estimates of β_e and β_u are similar across the different choices of x_{t+1} and P, and none of our estimates of the two parameters is statistically significant. A possible explanation for these results is that the demand for electricity and fuel is determined mainly by the levels of the other inputs—i.e., labor, capital and materials—but rarely by the latest level of a firm's productivity. As a result, e_{t+1} and u_{t+1} make poor proxies for productivity in our method.

There are also differences between our estimates using m_{t+1} and LP's. We focus on comparing with LP's original estimates given that the W-LP estimates are close to LP's original ones. In particular, our estimates of β_m and β_k (β_{ls} and β_{lu}) are noticeably smaller (larger) than LP's corresponding estimates, but our point estimates of β_e and β_u are similar to those of LP.

As LP point out in Section 2 of their paper, it is generally impossible to sign the simultaneity biases in the OLS estimates when there are multiple inputs, and the sign of biases depends on how the inputs covary with each other and with the latent productivity. Their analysis suggests that, without control for firms' productivity, the estimated coefficients of the most-variable inputs are likely biased upward, whereas those of the least-variable inputs can be biased downward if the inputs are positively correlated. Therefore, to better understand the causes of the difference in the estimates using the various methods, we need to know how variable the different inputs are and how rapidly they adjust with the latest productivity shocks (Marschak and Andrews (1944)).

The empirical analysis that we present in the following subsection shows that only m_{t+1}, but not e_{t+1} or u_{t+1}, depends, statistically significantly, on ξ_{t+1} and ξ_t, the innovations in productivity in the two latest periods. With ξ_t being a part of ω_t, these findings help explain why e_{t+1} and u_{t+1} seem poor proxies for ω_t in our method, and why only the estimate of β_m, but not those of β_e and β_u, are significantly inflated in OLS estimation and when we use either e_{t+1} or u_{t+1} as one of the proxies for ω_t in our method. In addition, we find that neither l^s_{t+1} nor l^u_{t+1} depends, statistically significantly, on ξ_{t+1} or ξ_t, showing that the labor inputs adjust considerably more slowly than the material input. Thus, these findings also offer a potential explanation for the difference between our estimates and LP's: the estimates of the coefficients of more- (less-) variable inputs may be biased upward (downward) due to imperfect control of the latent productivity under LP's method.

4.2 Inputs and Productivity Shocks

To study how quickly firms adjust their inputs to the latest changes in their productivity, we note that, for each period-$(t+1)$ input $z_{t+1} = l^s_{t+1}$, l^u_{t+1}, k_{t+1}, m_{t+1}, e_{t+1}, and u_{t+1}, we have:

$$\text{cov}(\omega_{t+1}, z_{t+1}) = \text{cov}(\tilde{y}_{t+1}(\beta), z_{t+1}), \tag{21}$$

and

$$\text{cov}(\xi_{t+1}, z_{t+1}) = \text{cov}\left(\tilde{y}_{t+1}(\beta) - \sum_{p=0}^{P} \alpha_{x\omega}^{-1} \tilde{\varphi}_p \tilde{y}_t(\beta)^p, z_{t+1}\right). \tag{22}$$

Thus, for each input $z_{t+1} = l^s_{t+1}$, l^u_{t+1}, k_{t+1}, m_{t+1}, e_{t+1}, and u_{t+1}, we may estimate $\text{cov}(\omega_{t+1}, z_{t+1})$ by $\text{cov}(\tilde{y}_{t+1}(\hat{\beta}), z_{t+1})$ and $\text{cov}(\xi_{t+1}, z_{t+1})$ by $\text{cov}\left(\tilde{y}_{t+1}(\hat{\beta}) - \sum_{p=0}^{P} \hat{\alpha}_{x\omega}^{-1} \hat{\tilde{\varphi}}_p \tilde{y}_t(\hat{\beta})^p, z_{t+1}\right)$, where $\hat{\beta}$ denotes the vector of estimated production-function parameters. To account for the effect of estimating β by $\hat{\beta}$ on the standard errors of the estimates for these covariances, we separately add each moment equality for these covariances (equations (21) and (22)) as one additional moment restriction to the moment conditions in (20) to estimate the covariance together with θ by the two-step GMM. Because using m_{t+1}, in comparison to e_{t+1} or u_{t+1}, as one of the two proxies for ω_t seems to perform better, we let $\tilde{x}_{t+1} = \tilde{m}_{t+1}$ in the moment restrictions in (20) in our estimation.

Table 9 shows estimates of the covariances. The covariance between productivity ω_{t+1} and the inputs, shown in part (A) of the table, are significantly positive for all inputs under all the different specifications of P. This shows that each input choice is affected, directly or indirectly, by a firm's current productivity. On the other hand, the covariances between technological innovation ξ_{t+1} and the inputs, shown in part (B) of the table, are all positive, but statistically insignificant, for all the cases of P that we consider. Among them, the covariance between ξ_{t+1} and material input m_{t+1} is closer to being statistically significantly positive. To gain statistical power, we reestimate the covariances by using the longer panel data (1979–1996) available, along with biennial time fixed effects. We report the results in part (C) of the table. With the increase in sample size, we find a statistically significant and positive covariance between ξ_{t+1} and m_{t+1}, but not between ξ_{t+1} and any other input. These results show that only the material input m adjusts in response to the latest innovation in productivity. The other inputs, including the two types of labor inputs, adjust more slowly to changes in productivity.[23]

To further analyze how inputs adjust with productivity, we also estimate $\text{cov}(\omega_t, z_{t+1})$ by $\text{cov}(\tilde{y}_t(\hat{\beta}), z_{t+1})$ and $\text{cov}(\xi_t, z_{t+1})$ by $\text{cov}\left(\tilde{y}_t(\hat{\beta}) - \sum_{p=0}^{P} \hat{\alpha}_{m\omega}^{-1}\hat{\tilde{\varphi}}_p \tilde{y}_{t-1}(\hat{\beta})^p, z_{t+1}\right)$, for each input $z_{t+1} = l_{t+1}^s$, l_{t+1}^u, k_{t+1}, m_{t+1}, e_{t+1}, and u_{t+1}. The analysis of these covariances also help explain the differences in our estimates under different choices of x_{t+1}. Estimation and computation of standard errors follow the same procedure as above. Table 10 reports the estimates of the covariances.

The covariance between the one-period lag productivity ω_t and each input, shown in part (A) of the table, is significantly positive for all the specifications of P that we consider. The covariance between lag technological innovation ξ_t and the inputs, shown in part (B) of the table, is significantly positive for the material input m_{t+1} for the cases of $P = 2, 3$, but it is statistically insignificant for all the other inputs. To gain statistical power, we again reestimate the covariances by using the longer panel (1979-1996) available, along with biennial time fixed effects. We report the results in part (C) of the table. With the increase in sample size, we obtain qualitatively the same results as those with the shorter panel, except that the positive covariance between the lag technological innovation ξ_t and material input m_{t+1} is statistically significant, at either the 5% or the 10% level, for all three choices of P.

In sum, we find that, although all the inputs show positive covariance with the current and one-period lagged productivity, only the material input shows statistically significant and positive covariance with the current and one-period lag productivity shock. This suggests that, although firms generally determine the levels of their inputs in accordance with their productivity, they rapidly adjust only the material input to the latest change in their productivity. The slower adjustments of the labor inputs are likely due to frictions in the labor market: hiring and firing costs may prevent firms from adjusting their labor inputs rapidly to respond to shocks to their productivity. Meanwhile, the adjustment in the capital input is also slow, which is not surprising given the time needed to put new capital in place. Therefore, in light of these findings, policies that aim to reduce the frictions in the labor markets have the potential to improve efficiency in the industry.

[23] The covariance between ξ_{t+1} and k_{t+1} is negative and statistically significant at the 5% level for the case of $P = 2$. The negative correlation may be due to a nonlinear relationship between ω and k.

5 Conclusions

In this paper we propose a new approach for structural identification and estimation of production functions, relaxing the well-known scalar-unobservable assumption maintained by the existing methods of OP/LP/ACF. The new approach is more robust when there are important unobservables in addition to the latent productivity. It also frees up some important identification sources that were not applicable under the scalar-unobservable assumption. We introduce a straightforward GMM procedure for estimating structural parameters in production functions, following our identification results. The estimation procedure is straightforward to apply and can be adjusted to allow for potential measurement errors in the input variables as long as the measurement errors are independent across time.

We apply our method to studying how rapidly firms respond in their input decisions to the latest changes in their productivity. Based on the estimates of the covariances between the inputs and the latest shocks to productivity, we find that firms are quick to adjust the material input, but much slower to adjust the labor and capital inputs.

It worth pointing out that, although our method does not produce point estimates of firm-level productivity, its applicability does not seem significantly limited by the issue. For example, it can be used to essentially replicate OP's empirical analysis of deregulation's effects in the telecommunications equipment industry. In view of the large number of applications based on the previous methods, we believe that our contribution to this literature can be of value to future studies of various issues centered around firm productivity and production functions.

References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007). Econometric tools for analyzing market outcomes. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Vol. 6A*. North-Holland.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

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*, 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.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. Volume 6 of *Handbook of Econometrics*, Chapter 76, pp. 5549 – 5632. Elsevier.

Chen, X. and H. White (1999). Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Transactions on Information Theory 45*(2), 682–691.

De Loecker, J. (2010). A note on detecting learning by exporting. Technical report, National Bureau of Economic Research.

D'Haultfoeuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory 27*(03), 460–471.

Doraszelski, U. and J. Jaumandreu (2013). R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies 80*(4), 1338–1383.

Dunford, N. and J. Schwartz (1971). *Linear Operators*, Volume 3. Wiley.

Gallant, A. R. and D. W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica 55*, 363–390.

Gandhi, A., S. Navarro, and D. Rivers (2013). On the identification of production functions: How heterogeneous is productivity?

Hu, Y. and S. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*, 195–216.

Huang, G. and Y. Hu (2011). Estimating production functions with robustness against errors in the proxy variables. *Available at SSRN 1805213*.

il Kim, K., A. Petrin, and S. Song (2016). Estimating production functions with control functions when capital is measured with error. *Journal of Econometrics 190*(2), 267–279.

Imbens, G. and W. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Imbens, G. W. (2007). Nonadditive models with endogenous regressors. *Econometric Society Monographs 43*, 17.

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.

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica 73*(4), 1103–1123.

Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70*, 317–342.

Marschak, J. and W. H. Andrews (1944). Random simultaneous equations and the theory of production. *Econometrica, Journal of the Econometric Society*, 143–205.

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.

Petrin, A., B. P. Poi, J. Levinsohn, et al. (2004). Production function estimation in stata using inputs to control for unobservables. *Stata journal 4*, 113–123.

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

Wooldridge, J. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.

Tables

Table 1: Monte Carlo results with $x_{t+1} = m_{t+1}$

N	P		β_l	β_m	β_u	β_k	$\tilde{\varphi}_1$
		True	0.400	0.200	0.100	0.300	3.333
1000	1	Simulation Mean	0.364	0.164	0.064	0.405	3.227
		Simulation St. Dev.	(0.070)	(0.070)	(0.070)	(0.249)	(1.898)
		Theoretical St. Err.	[0.087]	[0.087]	[0.087]	[0.271]	[3.243]
		Simulation RMSE	(0.079)	(0.079)	(0.079)	(0.270)	(1.901)
		Simulation 95% Cover	0.999	0.997	0.999	0.994	0.941
2000	1	Simulation Mean	0.382	0.182	0.082	0.354	3.535
		Simulation St. Dev.	(0.054)	(0.053)	(0.054)	(0.182)	(1.794)
		Theoretical St. Err.	[0.061]	[0.061]	[0.061]	[0.191]	[2.293]
		Simulation RMSE	(0.057)	(0.056)	(0.057)	(0.190)	(1.806)
		Simulated 95% Cover	0.996	0.998	0.997	0.992	0.934
4000	1	Simulation Mean	0.393	0.193	0.093	0.321	3.677
		Simulation St. Dev.	(0.041)	(0.041)	(0.041)	(0.132)	(1.544)
		Theoretical St. Err.	[0.043]	[0.043]	[0.043]	[0.135]	[1.622]
		Simulation RMSE	(0.042)	(0.042)	(0.041)	(0.133)	(1.582)
		Simulation 95% Cover	0.994	0.995	0.994	0.986	0.927
8000	1	Simulation Mean	0.397	0.198	0.097	0.308	3.597
		Simulation St. Dev.	(0.030)	(0.030)	(0.030)	(0.094)	(1.211)
		Theoretical St. Err.	[0.031]	[0.031]	[0.031]	[0.096]	[1.147]
		Simulation RMSE	(0.030)	(0.030)	(0.030)	(0.094)	(1.240)
		Simulation 95% Cover	0.986	0.983	0.982	0.978	0.935

Table 2: Monte Carlo results with the $x_{t+1} = y_{t+1}$

N	P		β_l	β_m	β_u	β_k	ϕ_1
		True	0.400	0.200	0.100	0.300	1.000
1000	1	Simulation Mean	0.400	0.199	0.099	0.313	1.002
		Simulation St. Dev.	(0.044)	(0.044)	(0.044)	(0.231)	(0.036)
		Theoretical St. Err.	[0.043]	[0.043]	[0.043]	[0.205]	[0.025]
		Simulation RMSE	(0.044)	(0.044)	(0.044)	(0.231)	(0.037)
		Simulation 95% Cover	0.926	0.923	0.929	0.934	0.926
2000	1	Simulation Mean	0.400	0.200	0.100	0.304	1.003
		Simulation St. Dev.	(0.032)	(0.032)	(0.031)	(0.153)	(0.022)
		Theoretical St. Err.	[0.030]	[0.030]	[0.030]	[0.145]	[0.018]
		Simulation RMSE	(0.032)	(0.032)	(0.031)	(0.153)	(0.022)
		Simulation 95% Cover	0.916	0.920	0.922	0.945	0.922
4000	1	Simulation Mean	0.400	0.200	0.101	0.302	1.001
		Simulation St. Dev.	(0.022)	(0.022)	(0.022)	(0.107)	(0.014)
		Theoretical St. Err.	[0.021]	[0.021]	[0.021]	[0.102]	[0.012]
		Simulation RMSE	(0.022)	(0.022)	(0.022)	(0.107)	(0.015)
		Simulation 95% Cover	0.933	0.930	0.932	0.942	0.925
8000	1	Simulation Mean	0.400	0.200	0.100	0.304	1.000
		Simulation St. Dev.	(0.015)	(0.016)	(0.016)	(0.073)	(0.009)
		Theoretical St. Err.	[0.015]	[0.015]	[0.015]	[0.072]	[0.009]
		Simulation RMSE	(0.015)	(0.016)	(0.016)	(0.073)	(0.009)
		Simulation 95% Cover	0.944	0.941	0.939	0.952	0.942

Table 3: Monte Carlo results with both m_{t+1} and y_{t+1}

N	P		β_l	β_m	β_u	β_k	ϕ_1	ϕ_2	ϕ_3
		True	0.400	0.200	0.100	0.300	1.000	0.000	0.000
1000	1	Mean	0.388	0.187	0.088	0.338	1.000		
		St. Dev.	(0.041)	(0.041)	(0.041)	(0.160)	(0.027)		
		RMSE	(0.043)	(0.043)	(0.043)	(0.164)	(0.027)		
		95% Cover	0.915	0.918	0.911	0.930	0.937		
2000	1	Mean	0.393	0.192	0.093	0.325	0.999		
		St. Dev.	(0.030)	(0.030)	(0.030)	(0.113)	(0.019)		
		RMSE	(0.030)	(0.031)	(0.031)	(0.116)	(0.019)		
		95% Cover	0.917	0.911	0.916	0.929	0.934		
4000	1	Mean	0.397	0.197	0.097	0.307	1.000		
		St. Dev.	(0.021)	(0.021)	(0.021)	(0.077)	(0.013)		
		RMSE	(0.021)	(0.021)	(0.021)	(0.077)	(0.013)		
		95% Cover	0.922	0.927	0.925	0.943	0.939		
8000	1	Mean	0.398	0.198	0.098	0.306	1.000		
		St. Dev.	(0.014)	(0.014)	(0.014)	(0.054)	(0.009)		
		RMSE	(0.014)	(0.015)	(0.014)	(0.054)	(0.009)		
		95% Cover	0.935	0.936	0.941	0.943	0.942		
1000	2	Mean	0.385	0.182	0.084	0.350	1.000	-0.000	
		St. Dev.	(0.038)	(0.040)	(0.039)	(0.156)	(0.043)	(0.029)	
		RMSE	(0.041)	(0.044)	(0.043)	(0.164)	(0.043)	(0.029)	
		95% Cover	0.932	0.920	0.920	0.934	0.951	0.989	
2000	2	Mean	0.390	0.189	0.090	0.334	0.999	0.000	
		St. Dev.	(0.028)	(0.028)	(0.028)	(0.112)	(0.021)	(0.013)	
		RMSE	(0.030)	(0.031)	(0.030)	(0.117)	(0.021)	(0.013)	
		95% Cover	0.940	0.930	0.939	0.925	0.946	0.982	
4000	2	Mean	0.396	0.195	0.096	0.313	1.000	0.000	
		St. Dev.	(0.021)	(0.021)	(0.021)	(0.078)	(0.014)	(0.008)	
		RMSE	(0.021)	(0.021)	(0.021)	(0.079)	(0.014)	(0.008)	
		95% Cover	0.933	0.932	0.930	0.941	0.949	0.978	
8000	2	Mean	0.398	0.198	0.098	0.306	1.000	0.000	
		St. Dev.	(0.015)	(0.014)	(0.014)	(0.055)	(0.009)	(0.005)	
		RMSE	(0.015)	(0.015)	(0.015)	(0.055)	(0.009)	(0.005)	
		95% Cover	0.931	0.931	0.938	0.937	0.950	0.974	
1000	3	Mean	0.385	0.184	0.085	0.354	1.000	-0.000	-0.000
		St. Dev.	(0.036)	(0.036)	(0.036)	(0.163)	(0.113)	(0.025)	(0.038)
		RMSE	(0.039)	(0.040)	(0.039)	(0.171)	(0.113)	(0.025)	(0.038)
		95% Cover	0.935	0.944	0.939	0.922	0.952	0.990	0.993
2000	3	Mean	0.391	0.190	0.091	0.329	1.000	-0.000	0.000
		St. Dev.	(0.027)	(0.028)	(0.027)	(0.111)	(0.076)	(0.014)	(0.026)
		RMSE	(0.028)	(0.029)	(0.029)	(0.115)	(0.076)	(0.014)	(0.026)
		95% Cover	0.937	0.930	0.941	0.931	0.949	0.983	0.994
4000	3	Mean	0.395	0.194	0.095	0.316	1.000	-0.000	-0.000
		St. Dev.	(0.020)	(0.020)	(0.020)	(0.078)	(0.045)	(0.008)	(0.014)
		RMSE	(0.021)	(0.021)	(0.021)	(0.079)	(0.045)	(0.008)	(0.014)
		95% Cover	0.940	0.932	0.936	0.940	0.955	0.974	0.987
8000	3	Mean	0.398	0.198	0.098	0.306	0.999	-0.000	0.000
		St. Dev.	(0.014)	(0.014)	(0.014)	(0.053)	(0.032)	(0.006)	(0.009)
		RMSE	(0.014)	(0.014)	(0.014)	(0.053)	(0.032)	(0.006)	(0.009)
		95% Cover	0.947	0.948	0.942	0.956	0.950	0.972	0.984

Table 4: A comparison of Monte Carlo results across the ordinary least squares (OLS), the method of Wooldridge (W) with polynomial sieve control function of degree three, the method of Levinsohn and Petrin (LP) with polynomial sieve control function of degree three, and our method (HHS) for the case of $P = 1$ copied from Table 3.

	N		β_l	β_m	β_u	β_k		β_l	β_m	β_u	β_k
		True	0.400	0.200	0.100	0.300		0.400	0.200	0.100	0.300
OLS	1000	Mean	0.489	0.289	0.189	0.142					
		St. Dev.	(0.014)	(0.014)	(0.014)	(0.045)					
		RMSE	(0.090)	(0.090)	(0.090)	(0.165)					
		95% Cover	0.000	0.000	0.000	0.000					
OLS	2000	Mean	0.489	0.289	0.188	0.143					
		St. Dev.	(0.010)	(0.010)	(0.010)	(0.032)					
		RMSE	(0.090)	(0.089)	(0.089)	(0.160)					
		95% Cover	0.000	0.000	0.000	0.000					
OLS	4000	Mean	0.489	0.289	0.189	0.142					
		St. Dev.	(0.007)	(0.007)	(0.007)	(0.022)					
		RMSE	(0.089)	(0.089)	(0.089)	(0.159)					
		95% Cover	0.000	0.000	0.000	0.000					
OLS	8000	Mean	0.489	0.289	0.189	0.142					
		St. Dev.	(0.005)	(0.005)	(0.005)	(0.016)					
		RMSE	(0.089)	(0.089)	(0.089)	(0.159)					
		95% Cover	0.000	0.000	0.000	0.000					
W	1000	Mean	0.445	0.248	0.147	0.292	LP	0.487	0.233	0.188	0.322
		St. Dev.	(0.009)	(0.017)	(0.016)	(0.103)		(0.010)	(0.055)	(0.010)	(0.285)
		RMSE	(0.046)	(0.051)	(0.049)	(0.103)		(0.088)	(0.064)	(0.088)	(0.286)
		95% Cover	0.006	0.000	0.001	0.003		—	—	—	—
W	2000	Mean	0.444	0.247	0.146	0.301	LP	0.487	0.233	0.188	0.299
		St. Dev.	(0.006)	(0.015)	(0.012)	(0.078)		(0.007)	(0.036)	(0.007)	(0.232)
		RMSE	(0.045)	(0.049)	(0.047)	(0.078)		(0.088)	(0.049)	(0.088)	(0.232)
		95% Cover	0.002	0.000	0.000	0.001		—	—	—	—
W	4000	Mean	0.445	0.245	0.144	0.310	LP	0.488	0.236	0.188	0.308
		St. Dev.	(0.003)	(0.009)	(0.006)	(0.041)		(0.005)	(0.022)	(0.005)	(0.169)
		RMSE	(0.045)	(0.046)	(0.045)	(0.042)		(0.088)	(0.042)	(0.088)	(0.169)
		95% Cover	0.000	0.000	0.000	0.000		—	—	—	—
W	8000	Mean	0.445	0.245	0.144	0.314	LP	0.487	0.237	0.188	0.313
		St. Dev.	(0.001)	(0.003)	(0.003)	(0.016)		(0.004)	(0.014)	(0.004)	(0.120)
		RMSE	(0.045)	(0.045)	(0.044)	(0.022)		(0.088)	(0.040)	(0.088)	(0.121)
		95% Cover	0.000	0.000	0.000	0.000		—	—	—	—
HHS	1000	Mean	0.388	0.187	0.088	0.338					
		St. Dev.	(0.041)	(0.041)	(0.041)	(0.160)					
		RMSE	(0.043)	(0.043)	(0.043)	(0.164)					
		95% Cover	0.915	0.918	0.911	0.930					
HHS	2000	Mean	0.393	0.192	0.093	0.325					
		St. Dev.	(0.030)	(0.030)	(0.030)	(0.113)					
		RMSE	(0.030)	(0.031)	(0.031)	(0.116)					
		95% Cover	0.917	0.911	0.916	0.929					
HHS	4000	Mean	0.397	0.197	0.097	0.307					
		St. Dev.	(0.021)	(0.021)	(0.021)	(0.077)					
		RMSE	(0.021)	(0.021)	(0.021)	(0.077)					
		95% Cover	0.922	0.927	0.925	0.943					
HHS	8000	Mean	0.398	0.198	0.098	0.306					
		St. Dev.	(0.014)	(0.014)	(0.014)	(0.054)					
		RMSE	(0.014)	(0.015)	(0.014)	(0.054)					
		95% Cover	0.935	0.936	0.941	0.943					

Table 5: Monte Carlo results under heteroskedasticity, $\eta_t \sim N(0, s_\eta^2(1 + (\omega_t/2)^2))$.

	N		β_l	β_m	β_u	β_k		β_l	β_m	β_u	β_k
		True	0.400	0.200	0.100	0.300		0.400	0.200	0.100	0.300
OLS	1000	Mean	0.489	0.289	0.189	0.144					
		St. Dev.	(0.018)	(0.018)	(0.018)	(0.062)					
		RMSE	(0.090)	(0.090)	(0.091)	(0.168)					
		95% Cover	0.000	0.000	0.000	0.004					
OLS	2000	Mean	0.489	0.289	0.188	0.141					
		St. Dev.	(0.013)	(0.013)	(0.013)	(0.044)					
		RMSE	(0.090)	(0.090)	(0.089)	(0.165)					
		95% Cover	0.000	0.000	0.000	0.000					
OLS	4000	Mean	0.489	0.289	0.189	0.142					
		St. Dev.	(0.009)	(0.009)	(0.009)	(0.031)					
		RMSE	(0.089)	(0.089)	(0.089)	(0.161)					
		95% Cover	0.000	0.000	0.000	0.000					
OLS	8000	Mean	0.489	0.289	0.189	0.142					
		St. Dev.	(0.006)	(0.006)	(0.006)	(0.022)					
		RMSE	(0.089)	(0.089)	(0.089)	(0.160)					
		95% Cover	0.000	0.000	0.000	0.000					
W	1000	Mean	0.445	0.247	0.146	0.301	LP	0.487	0.241	0.187	0.455
		St. Dev.	(0.005)	(0.015)	(0.014)	(0.086)		(0.013)	(0.089)	(0.013)	(0.428)
		RMSE	(0.045)	(0.049)	(0.048)	(0.086)		(0.088)	(0.098)	(0.088)	(0.456)
		95% Cover	0.017	0.002	0.002	0.002		—	—	—	—
W	2000	Mean	0.445	0.246	0.145	0.310	LP	0.487	0.236	0.188	0.350
		St. Dev.	(0.003)	(0.010)	(0.009)	(0.052)		(0.009)	(0.059)	(0.009)	(0.317)
		RMSE	(0.045)	(0.047)	(0.046)	(0.053)		(0.088)	(0.069)	(0.088)	(0.321)
		95% Cover	0.005	0.000	0.000	0.0004		—	—	—	—
W	4000	Mean	0.445	0.245	0.144	0.314	LP	0.487	0.233	0.188	0.320
		St. Dev.	(0.001)	(0.003)	(0.002)	(0.012)		(0.006)	(0.038)	(0.006)	(0.249)
		RMSE	(0.045)	(0.045)	(0.044)	(0.019)		(0.088)	(0.050)	(0.088)	(0.250)
		95% Cover	0.000	0.000	0.000	0.000		—	—	—	—
W	8000	Mean	0.445	0.245	0.144	0.315	LP	0.488	0.235	0.187	0.303
		St. Dev.	(0.001)	(0.001)	(0.001)	(0.000)		(0.004)	(0.023)	(0.004)	(0.198)
		RMSE	(0.045)	(0.045)	(0.044)	(0.015)		(0.088)	(0.042)	(0.087)	(0.198)
		95% Cover	0.000	0.000	0.000	0.000		—	—	—	—
HHS	1000	Mean	0.375	0.174	0.074	0.382					
		St. Dev.	(0.055)	(0.055)	(0.055)	(0.212)					
		RMSE	(0.060)	(0.061)	(0.061)	(0.228)					
		95% Cover	0.901	0.888	0.894	0.923					
HHS	2000	Mean	0.387	0.186	0.087	0.340					
		St. Dev.	(0.040)	(0.041)	(0.040)	(0.148)					
		RMSE	(0.042)	(0.043)	(0.042)	(0.153)					
		95% Cover	0.906	0.903	0.909	0.928					
HHS	4000	Mean	0.394	0.193	0.093	0.320					
		St. Dev.	(0.029)	(0.029)	(0.030)	(0.104)					
		RMSE	(0.030)	(0.030)	(0.030)	(0.106)					
		95% Cover	0.906	0.913	0.896	0.931					
HHS	8000	Mean	0.397	0.197	0.097	0.311					
		St. Dev.	(0.020)	(0.020)	(0.020)	(0.072)					
		RMSE	(0.020)	(0.020)	(0.020)	(0.073)					
		95% Cover	0.918	0.921	0.920	0.939					

Table 6: Monte Carlo results with both m_{t+1} and y_{t+1} for a quadratic AR(1) process of the transition of productivity, where the model with $P = 1$ is mis-specified.

N	P			β_l	β_m	β_u	β_k	ϕ_1	ϕ_2	ϕ_3
			True	0.400	0.200	0.100	0.300	1.000	-0.025	0.000
1000	1	Mis-specified	Mean	0.410	0.208	0.109	0.313	1.056		
			St. Dev.	(0.036)	(0.036)	(0.036)	(0.177)	(0.028)		
			RMSE	(0.037)	(0.037)	(0.037)	(0.178)	(0.062)		
			95% Cover	0.905	0.914	0.905	0.911	0.352		
2000	1	Mis-specified	Mean	0.419	0.218	0.119	0.276	1.057		
			St. Dev.	(0.024)	(0.025)	(0.025)	(0.123)	(0.020)		
			RMSE	(0.031)	(0.031)	(0.031)	(0.125)	(0.061)		
			95% Cover	0.842	0.843	0.840	0.909	0.149		
4000	1	Mis-specified	Mean	0.422	0.222	0.123	0.260	1.057		
			St. Dev.	(0.017)	(0.017)	(0.017)	(0.083)	(0.015)		
			RMSE	(0.028)	(0.028)	(0.028)	(0.093)	(0.059)		
			95% Cover	0.694	0.696	0.692	0.898	0.034		
8000	1	Mis-specified	Mean	0.424	0.224	0.124	0.254	1.056		
			St. Dev.	(0.012)	(0.012)	(0.012)	(0.061)	(0.011)		
			RMSE	(0.027)	(0.027)	(0.027)	(0.076)	(0.058)		
			95% Cover	0.474	0.462	0.451	0.835	0.003		
1000	2		Mean	0.382	0.182	0.083	0.357	0.998	-0.024	
			St. Dev.	(0.042)	(0.043)	(0.042)	(0.168)	(0.046)	(0.021)	
			RMSE	(0.046)	(0.047)	(0.045)	(0.177)	(0.046)	(0.021)	
			95% Cover	0.914	0.914	0.918	0.918	0.938	0.795	
2000	2		Mean	0.389	0.188	0.089	0.343	0.999	-0.024	
			St. Dev.	(0.032)	(0.032)	(0.031)	(0.123)	(0.028)	(0.012)	
			RMSE	(0.034)	(0.034)	(0.033)	(0.130)	(0.028)	(0.012)	
			95% Cover	0.906	0.909	0.918	0.912	0.935	0.806	
4000	2		Mean	0.394	0.194	0.095	0.324	1.000	-0.024	
			St. Dev.	(0.022)	(0.022)	(0.022)	(0.084)	(0.018)	(0.007)	
			RMSE	(0.023)	(0.023)	(0.023)	(0.087)	(0.018)	(0.007)	
			95% Cover	0.916	0.930	0.918	0.926	0.934	0.853	
8000	2		Mean	0.397	0.196	0.096	0.319	1.000	-0.024	
			St. Dev.	(0.015)	(0.015)	(0.015)	(0.057)	(0.012)	(0.005)	
			RMSE	(0.016)	(0.016)	(0.016)	(0.060)	(0.012)	(0.005)	
			95% Cover	0.932	0.929	0.933	0.930	0.932	0.866	
1000	3		Mean	0.389	0.186	0.088	0.344	0.993	-0.025	0.000
			St. Dev.	(0.028)	(0.029)	(0.028)	(0.137)	(0.050)	(0.036)	(0.007)
			RMSE	(0.030)	(0.032)	(0.031)	(0.144)	(0.051)	(0.036)	(0.007)
			95% Cover	0.980	0.972	0.975	0.961	0.976	0.930	1.000
2000	3		Mean	0.393	0.193	0.094	0.326	0.997	-0.023	0.000
			St. Dev.	(0.021)	(0.021)	(0.021)	(0.097)	(0.029)	(0.018)	(0.003)
			RMSE	(0.022)	(0.023)	(0.022)	(0.101)	(0.029)	(0.018)	(0.003)
			95% Cover	0.978	0.984	0.977	0.963	0.962	0.919	0.999
4000	3		Mean	0.395	0.195	0.095	0.319	1.001	-0.022	0.000
			St. Dev.	(0.016)	(0.016)	(0.016)	(0.070)	(0.017)	(0.009)	(0.002)
			RMSE	(0.017)	(0.017)	(0.017)	(0.072)	(0.017)	(0.010)	(0.002)
			95% Cover	0.976	0.972	0.975	0.964	0.958	0.898	0.996
8000	3		Mean	0.397	0.197	0.097	0.313	1.001	-0.022	0.000
			St. Dev.	(0.012)	(0.011)	(0.011)	(0.049)	(0.011)	(0.007)	(0.001)
			RMSE	(0.012)	(0.012)	(0.012)	(0.051)	(0.011)	(0.007)	(0.001)
			95% Cover	0.976	0.978	0.976	0.960	0.952	0.881	0.977

Table 7: Monte Carlo results with both m_{t+1} and y_{t+1} for a sub-unit-root AR(1) process of the transition of productivity.

N	P		β_l	β_m	β_u	β_k	ϕ_1	ϕ_2	ϕ_3
		True	0.400	0.200	0.100	0.300	0.950	0.000	0.000
1000	1	Mean	0.391	0.190	0.090	0.346	0.950		
		St. Dev.	(0.037)	(0.037)	(0.037)	(0.163)	(0.042)		
		RMSE	(0.038)	(0.039)	(0.038)	(0.169)	(0.042)		
		95% Cover	0.891	0.899	0.900	0.913	0.900		
2000	1	Mean	0.396	0.195	0.095	0.320	0.951		
		St. Dev.	(0.026)	(0.026)	(0.026)	(0.110)	(0.031)		
		RMSE	(0.026)	(0.026)	(0.026)	(0.112)	(0.031)		
		95% Cover	0.912	0.910	0.915	0.926	0.911		
4000	1	Mean	0.398	0.197	0.098	0.312	0.950		
		St. Dev.	(0.018)	(0.018)	(0.018)	(0.075)	(0.020)		
		RMSE	(0.018)	(0.019)	(0.018)	(0.076)	(0.020)		
		95% Cover	0.918	0.926	0.931	0.932	0.939		
8000	1	Mean	0.398	0.198	0.098	0.307	0.949		
		St. Dev.	(0.012)	(0.012)	(0.012)	(0.051)	(0.014)		
		RMSE	(0.012)	(0.012)	(0.013)	(0.052)	(0.014)		
		95% Cover	0.943	0.941	0.942	0.946	0.944		
1000	2	Mean	0.385	0.184	0.085	0.355	0.926	-0.005	
		St. Dev.	(0.037)	(0.038)	(0.037)	(0.158)	(0.106)	(0.091)	
		RMSE	(0.040)	(0.041)	(0.040)	(0.167)	(0.108)	(0.091)	
		95% Cover	0.922	0.916	0.926	0.924	0.937	0.940	
2000	2	Mean	0.392	0.191	0.092	0.326	0.940	-0.001	
		St. Dev.	(0.026)	(0.026)	(0.026)	(0.105)	(0.046)	(0.042)	
		RMSE	(0.027)	(0.028)	(0.027)	(0.108)	(0.047)	(0.042)	
		95% Cover	0.935	0.935	0.931	0.938	0.942	0.936	
4000	2	Mean	0.396	0.195	0.097	0.312	0.944	0.000	
		St. Dev.	(0.018)	(0.019)	(0.018)	(0.073)	(0.024)	(0.024)	
		RMSE	(0.019)	(0.019)	(0.019)	(0.074)	(0.025)	(0.024)	
		95% Cover	0.934	0.937	0.937	0.947	0.947	0.948	
8000	2	Mean	0.398	0.198	0.098	0.307	0.947	0.001	
		St. Dev.	(0.013)	(0.013)	(0.013)	(0.050)	(0.015)	(0.015)	
		RMSE	(0.013)	(0.013)	(0.013)	(0.050)	(0.015)	(0.015)	
		95% Cover	0.940	0.941	0.940	0.953	0.950	0.944	
1000	3	Mean	0.384	0.183	0.085	0.372	0.973	-0.008	-0.019
		St. Dev.	(0.034)	(0.034)	(0.034)	(0.171)	(0.382)	(0.081)	(0.173)
		RMSE	(0.037)	(0.038)	(0.038)	(0.185)	(0.382)	(0.081)	(0.174)
		95% Cover	0.948	0.947	0.945	0.934	0.974	0.957	0.9744
2000	3	Mean	0.392	0.190	0.092	0.333	0.957	-0.002	-0.007
		St. Dev.	(0.024)	(0.024)	(0.024)	(0.111)	(0.272)	(0.037)	(0.131)
		RMSE	(0.025)	(0.026)	(0.025)	(0.115)	(0.272)	(0.037)	(0.132)
		95% Cover	0.962	0.956	0.960	0.944	0.968	0.952	0.969
4000	3	Mean	0.396	0.195	0.096	0.315	0.951	-0.000	-0.002
		St. Dev.	(0.017)	(0.017)	(0.018)	(0.076)	(0.185)	(0.023)	(0.092)
		RMSE	(0.018)	(0.018)	(0.018)	(0.077)	(0.185)	(0.023)	(0.092)
		95% Cover	0.961	0.947	0.948	0.947	0.952	0.936	0.961
8000	3	Mean	0.398	0.197	0.098	0.308	0.946	-0.000	0.002
		St. Dev.	(0.013)	(0.013)	(0.013)	(0.052)	(0.131)	(0.015)	(0.066)
		RMSE	(0.013)	(0.013)	(0.013)	(0.053)	(0.131)	(0.015)	(0.066)
		95% Cover	0.945	0.940	0.952	0.948	0.946	0.946	0.955

Table 8: Estimates of the gross output production function for ISIC 311.

	x_{t+1}	P	Production Function β_{ls}	β_{lu}	β_m	β_e	β_u	β_k	RTS	AR Coefficients $\tilde{\varphi}_1$	$\tilde{\varphi}_2$	$\tilde{\varphi}_3$
OLS			0.111	0.231	0.659	0.060	0.002	0.058	1.121			
			(0.008)	(0.009)	(0.008)	(0.006)	(0.004)	(0.005)	(0.006)			
LP			0.051	0.139	0.500	0.085	0.023	0.240	1.037			
			(0.009)	(0.010)	(0.078)	(0.007)	(0.004)	(0.053)	(0.059)			
W-LP			0.067	0.156	0.491	0.066	0.000	0.255	1.035			
(m)			(0.031)	(0.029)	(0.091)	(0.104)	(0.096)	(0.049)	(0.326)			
W-LP			0.076	0.166	0.494	0.079	0.001	0.262	1.077			
(e)			(0.031)	(0.029)	(0.091)	(0.091)	(0.096)	(0.049)	(0.315)			
W-LP			0.058	0.147	0.487	0.066	0.010	0.247	1.015			
(u)			(0.031)	(0.029)	(0.090)	(0.104)	(0.090)	(0.049)	(0.324)			
HHS	m_{t+1}	1	0.090	0.209	0.354	0.094	0.033	0.112	0.892	1.008		
(1)	y_{t+1}		(0.036)	(0.033)	(0.111)	(0.136)	(0.123)	(0.057)	(0.408)	(0.027)		
HHS	m_{t+1}	2	0.077	0.235	0.369	0.101	0.032	0.165	0.978	0.983	0.183	
(2)	y_{t+1}		(0.038)	(0.034)	(0.110)	(0.136)	(0.124)	(0.059)	(0.413)	(0.025)	(0.019)	
HHS	m_{t+1}	3	0.090	0.234	0.368	0.099	0.030	0.141	0.962	1.070	0.115	-0.070
(3)	y_{t+1}		(0.038)	(0.034)	(0.111)	(0.137)	(0.125)	(0.059)	(0.414)	(0.026)	(0.019)	(0.022)
HHS	e_{t+1}	1	0.069	0.219	0.636	0.065	0.015	0.153	1.155	1.068		
(4)	y_{t+1}		(0.040)	(0.037)	(0.126)	(0.153)	(0.138)	(0.063)	(0.459)	(0.022)		
HHS	e_{t+1}	2	0.098	0.207	0.659	0.087	0.022	0.218	1.291	0.943	0.042	
(5)	y_{t+1}		(0.039)	(0.036)	(0.128)	(0.157)	(0.141)	(0.062)	(0.465)	(0.032)	(0.034)	
HHS	e_{t+1}	3	0.062	0.205	0.636	0.099	0.030	0.121	1.153	0.855	0.077	0.199
(6)	y_{t+1}		(0.038)	(0.034)	(0.117)	(0.143)	(0.130)	(0.061)	(0.432)	(0.019)	(0.015)	(0.013)
HHS	u_{t+1}	3	0.071	0.198	0.652	0.118	0.014	0.167	1.220	1.014		
(7)	y_{t+1}		(0.038)	(0.035)	(0.122)	(0.149)	(0.134)	(0.060)	(0.441)	(0.033)		
HHS	u_{t+1}	3	0.094	0.179	0.673	0.115	0.020	0.304	1.386	0.976	0.036	
(8)	y_{t+1}		(0.038)	(0.035)	(0.123)	(0.151)	(0.136)	(0.060)	(0.449)	(0.030)	(0.033)	
HHS	u_{t+1}	3	0.134	0.215	0.669	0.135	0.031	0.127	1.311	0.642	-0.304	0.510
(9)	y_{t+1}		(0.036)	(0.033)	(0.111)	(0.137)	(0.125)	(0.060)	(0.418)	(0.023)	(0.020)	(0.020)

Note: Standard errors in parentheses.

Table 9: The covariance between ω_{t+1} (ξ_{t+1}) and inputs

(A) Covariance between technology ω_{t+1} and each input: 1979–1986

P	Cov(ω_{t+1}, l^s_{t+1})	Cov(ω_{t+1}, l^u_{t+1})	Cov(ω_{t+1}, m_{t+1})	Cov(ω_{t+1}, e_{t+1})	Cov(ω_{t+1}, u_{t+1})	Cov(ω_{t+1}, k_{t+1})
1	0.260***	0.264***	0.511***	0.406***	0.302***	0.233***
	(0.022)	(0.022)	(0.015)	(0.013)	(0.013)	(0.015)
2	0.182***	0.196***	0.422***	0.328***	0.207***	0.207***
	(0.023)	(0.024)	(0.016)	(0.014)	(0.014)	(0.015)
3	0.232***	0.257***	0.474***	0.360***	0.255***	0.192***
	(0.022)	(0.023)	(0.016)	(0.014)	(0.014)	(0.015)

(B) Covariance between technological innovation ξ_{t+1} and each input: 1979–1986

P	Cov(ξ_{t+1}, l^s_{t+1})	Cov(ξ_{t+1}, l^u_{t+1})	Cov(ξ_{t+1}, m_{t+1})	Cov(ξ_{t+1}, e_{t+1})	Cov(ξ_{t+1}, u_{t+1})	Cov(ξ_{t+1}, k_{t+1})
1	0.004	0.009	0.039	0.021	0.011	-0.001
	(0.058)	(0.071)	(0.038)	(0.033)	(0.033)	(0.049)
2	-0.005	0.003	0.025	0.005	-0.000	-0.018
	(0.062)	(0.076)	(0.039)	(0.035)	(0.035)	(0.048)
3	-0.003	0.006	0.037	0.014	0.006	-0.008
	(0.061)	(0.076)	(0.039)	(0.035)	(0.035)	(0.050)

(C) Covariance between technological innovation ξ_{t+1} and each input: 1979–1996

P	Cov(ξ_{t+1}, l^s_{t+1})	Cov(ξ_{t+1}, l^u_{t+1})	Cov(ξ_{t+1}, m_{t+1})	Cov(ξ_{t+1}, e_{t+1})	Cov(ξ_{t+1}, u_{t+1})	Cov(ξ_{t+1}, k_{t+1})
1	-0.010	-0.018	0.043**	0.019	-0.013	-0.031
	(0.027)	(0.029)	(0.018)	(0.016)	(0.015)	(0.019)
2	-0.017	-0.009	0.045**	0.017	-0.015	-0.037*
	(0.026)	(0.029)	(0.018)	(0.015)	(0.015)	(0.019)
3	-0.017	-0.009	0.036*	0.016	-0.015	-0.020
	(0.027)	(0.030)	(0.019)	(0.016)	(0.015)	(0.020)

Note: 1) Standard errors in parentheses; 2) $* \; p < 0.1$, $** \; p < 0.05$, $*** \; p < 0.01$.

Table 10: The covariance between ω_t (ξ_t) and inputs

(A) Covariance between lag technology ω_t and each input: 1979–1986

P	Cov(ω_t, l^s_{t+1})	Cov(ω_t, l^u_{t+1})	Cov(ω_t, m_{t+1})	Cov(ω_t, e_{t+1})	Cov(ω_t, u_{t+1})	Cov(ω_t, k_{t+1})
1	0.337***	0.308***	0.614***	0.492***	0.404***	0.406***
	(0.023)	(0.025)	(0.015)	(0.015)	(0.014)	(0.016)
2	0.096***	0.156***	0.313***	0.274***	0.076***	0.028
	(0.035)	(0.030)	(0.023)	(0.018)	(0.022)	(0.025)
3	0.087***	0.095***	0.309***	0.233***	0.118***	0.053***
	(0.030)	(0.030)	(0.020)	(0.017)	(0.018)	(0.018)

(B) Covariance between lag technological innovation ξ_t and each input: 1979–1986

P	Cov(ξ_t, l^s_{t+1})	Cov(ξ_t, l^u_{t+1})	Cov(ξ_t, m_{t+1})	Cov(ξ_t, e_{t+1})	Cov(ξ_t, u_{t+1})	Cov(ξ_t, k_{t+1})
1	0.002	0.013	0.048	0.013	0.003	-0.002
	(0.057)	(0.057)	(0.037)	(0.033)	(0.032)	(0.034)
2	0.010	0.020	0.069**	0.026	0.009	-0.006
	(0.056)	(0.056)	(0.035)	(0.032)	(0.031)	(0.034)
3	-0.004	0.012	0.056*	0.023	-0.002	-0.006
	(0.051)	(0.052)	(0.034)	(0.027)	(0.024)	(0.034)

(C) Covariance between lag technological innovation ξ_t and each input: 1979–1996

P	Cov(ξ_t, l^s_{t+1})	Cov(ξ_t, l^u_{t+1})	Cov(ξ_t, m_{t+1})	Cov(ξ_t, e_{t+1})	Cov(ξ_t, u_{t+1})	Cov(ξ_t, k_{t+1})
1	-0.014	-0.014	0.066***	-0.001	-0.029	-0.033
	(0.030)	(0.032)	(0.018)	(0.018)	(0.019)	(0.018)
2	-0.009	-0.004	0.050**	0.009	-0.011	-0.032
	(0.033)	(0.034)	(0.020)	(0.019)	(0.020)	(0.019)
3	-0.021	-0.011	0.038*	-0.014	-0.025	-0.055***
	(0.034)	(0.036)	(0.021)	(0.019)	(0.021)	(0.020)

Note: 1) Standard errors in parentheses; 2) $*\ p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

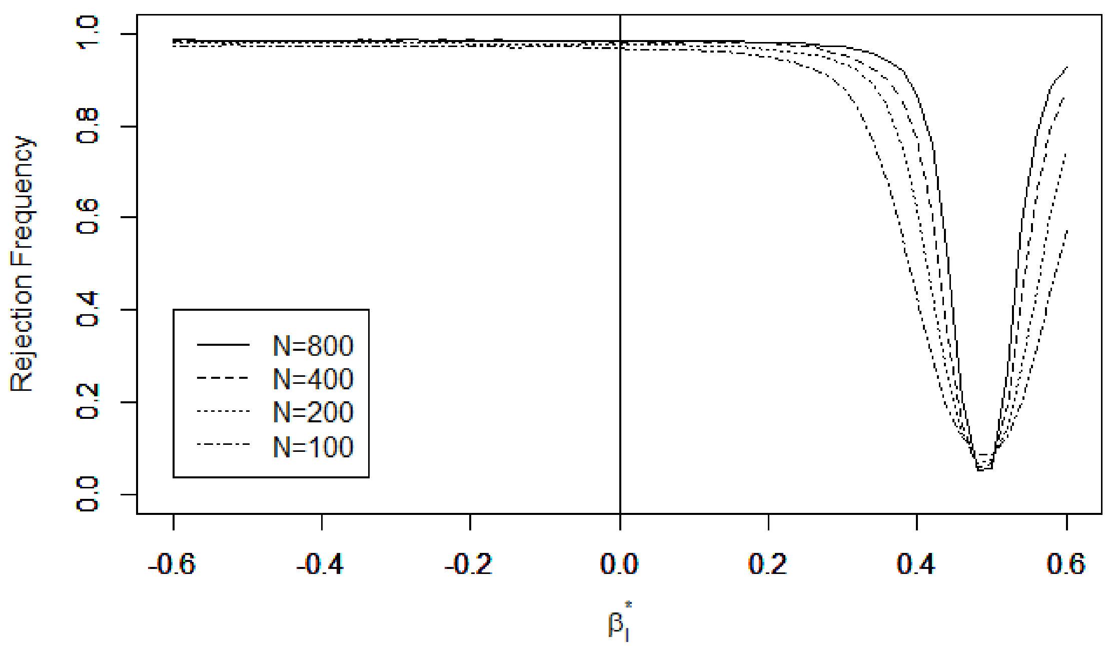

Figure 1: Monte Carlo results of rejection frequencies for the null hypothesis of H_0 : $\beta_l = \beta_l^*$ based on the K-statistics for a test of the Model of OP/LP/Wooldridge.

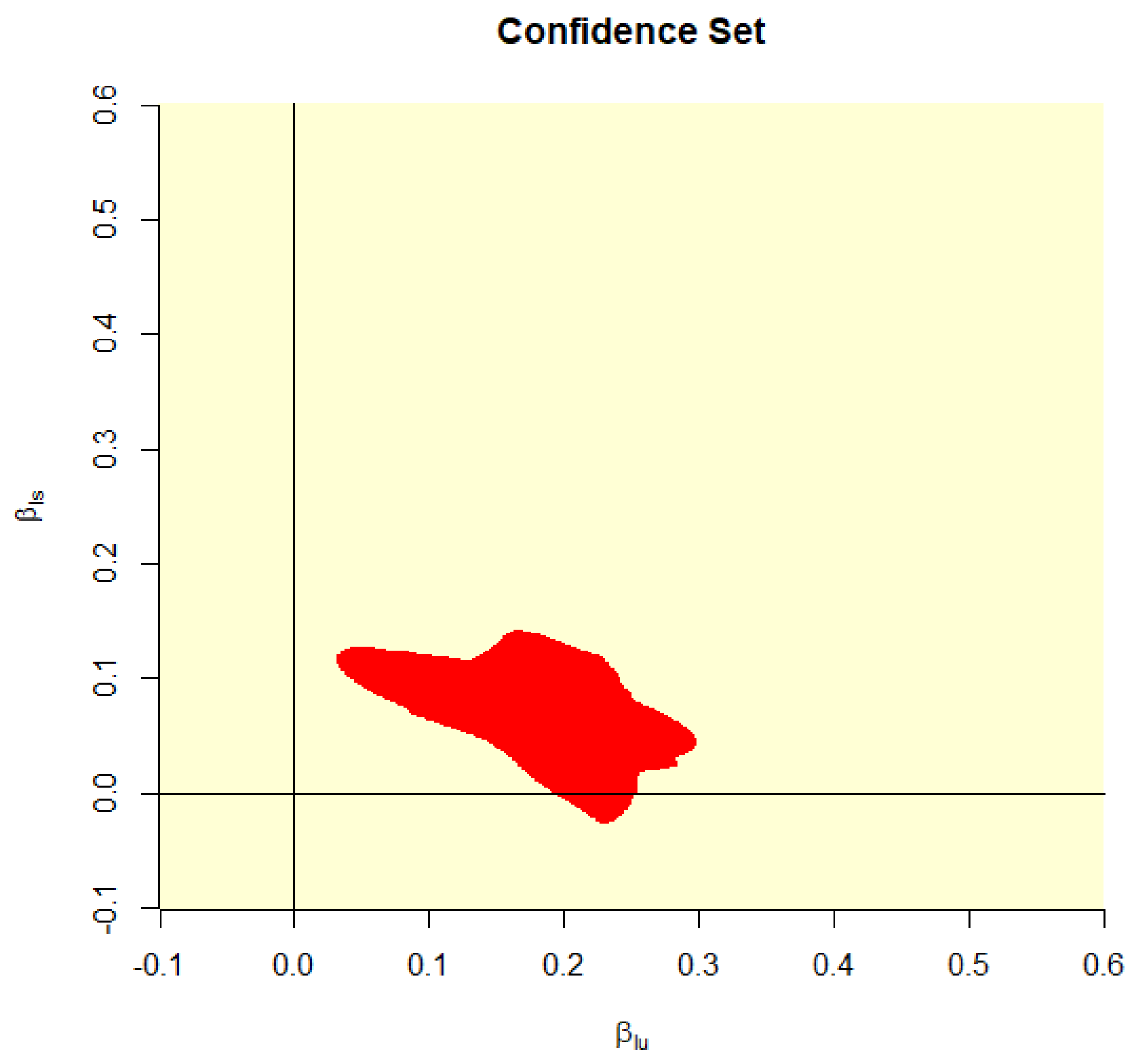

Figure 2: The 95% confidence set of (β_{ls}, β_{lu}) based on K-statistics for a test of the Model of OP/LP/Wooldridge.

Appendix A: Literature Review

We start by laying out the model used by OP/LP/ACF. 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 $\mathbb{E}\left(\eta_t|\omega_t, l_t, k_t\right) = 0$. The productivity ω_t follows an exogenous first-order Markov process:

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

where ξ_t is mean-independent of ω_{t-1}. The capital accumulates according to the 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:

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

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

$$m_t = \mu_{mt}\left(\omega_t, k_t\right),$$

which 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.

Olley and Pakes (1996)

OP propose a structural approach to estimate the production function. Their key observation is that we can use investment as a proxy for ω_t. More specifically, if the investment demand function $\iota_t\left(\omega_t, k_t\right)$ is strictly increasing in ω_t and we use $\iota^{-1}\left(., k_t\right)$ to indicate the inverse function of $\iota_t\left(\omega_t, k_t\right)$ for any fixed k_t, we have $\omega_t = \iota_t^{-1}\left(I_t, k_t\right)$. 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 $\phi_t (i_t, k_t) = \beta_0 + \beta_k k_t + \iota_t^{-1} (i_t, k_t)$ is estimated nonparametrically. We get an estimate of β_l and ϕ_{t-1} in this step.

Step 2: semiparametrically estimate:

$$y_t - \hat{\beta}_l l_t = \beta_0 + \beta_k k_t + \rho \left(\hat{\phi}_{t-1} - \beta_0 - \beta_k k_{t-1} \right) + \xi_t + \eta_t,$$

where $\rho (\omega_{t-1}) \equiv \mathbb{E} (\omega_t | \omega_{t-1})$ is specified nonparametrically. Here, one gets a consistent estimate of β_k using the condition that k_t, k_{t-1} and ω_{t-1} are mean-independent of ξ_t.

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, we have $\omega_t = \mu_{mt}^{-1}(m_t, k_t)$, where $\mu_{mt}^{-1}(., k_t)$ denotes the inverse function of $\mu_{mt}(\omega_t, k_t)$ for any fixed k_t. Then, following OP's idea, we can use a nonparametric function of k_t and m_t to control for ω_t when estimating the production function. Based on this insight, LP uses a two-step procedure, similar to OP's, to estimate the production function.

The LP method has two advantages over the original OP method. First, one does not have to eliminate 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.

Ackerberg, Caves and Frazer (2015)

The critique of ACF is that the first steps in OP's and LP's procedures are actually not identified because l_t would have no independent variations 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:

$$l_t = \psi_t (\omega_t, k_t) .$$

And for LP's method, one has $\omega_t = \mu_t^{-1} (m_t, k_t)$. Thus, $l_t = \psi_t \left(\mu_t^{-1} (m_t, k_t), k_t \right)$ is also a function of (m_t, k_t) and would be collinear with the nonparametric terms used to approximate the unknown function of $\tilde{\phi}_t(m_t, k_t) \equiv \beta_0 + \beta_k k_t + \mu_{mt}^{-1}(m_t, k_t)$.

ACF assume that the decision on l_t is made before that of m_t, and thus the intermediate input demand function would be $m_t = \mu_{mt}(\omega_t, k_t, l_t)$, where μ_{mt} is assumed to be strictly increasing in ω_t for any given (k_t, l_t). So, after substituting in the expression of $\omega_t = \mu_{mt}^{-1}(m_t, k_t, l_t)$, the production function can be written as follows:

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \mu_t^{-1}(m_t, k_t, l_t) + \eta_t.$$

To get around the identification problem of l_t in the first step of LP's procedure, ACF suggest estimating the coefficients of both l_t and k_t in the second step. 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 $\varphi_t(m_t, l_t, k_t) = \beta_0 + \beta_l l_t + \beta_k k_t + \mu_t^{-1}(m_t, k_t, l_t)$. This step produces estimates of φ_t and φ_{t-1}.

Step 2. Estimate $\beta \equiv (\beta_0, \beta_l, \beta_k)$ using the following set of two moment conditions,

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

where $\xi_t = \omega_t - \mathbb{E}(\omega_t | \omega_{t-1})$ is estimated by $\hat{\xi}_t = \hat{\varphi}_t - \beta_0 - \beta_l l_t - \beta_k k_t - \hat{\rho}(\hat{\varphi}_{t-1} - \beta_0 - \beta_l l_{t-1} - \beta_k k_{t-1})$, and $\rho(\omega_{t-1}) \equiv \mathbb{E}(\omega_t | \omega_{t-1})$ is specified nonparametrically.

Wooldridge (2009)

Wooldridge (2009) points out that we can actually implement the above methods with a GMM approach. In particular, we may stack up the moment conditions from the two steps of the above methods and estimate them together using the GMM framework. The approach is more efficient and allows one to use standard formulas to compute the asymptotic standard errors for the estimates.

Discussion

All of the above methods rely critically on the key assumption that the latent productivity is the only unobservable affecting the intermediate inputs and investment. So, when the observed intermediate inputs and investment are also affected by supply disruptions, optimization errors, measurement errors, 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 in the following:

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

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, now the equation 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, the part of ω_t that is a linear combination of m_t and ϵ_t would always be missed. Therefore, in this case, the above methods would not be able to completely eliminate the simultaneity bias. We will demonstrate this issue via simulations.

Appendix B: Semi-Nonparametric MLE Approach

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 can be 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 data of two periods. Thus we adopt the partial likelihood framework (c.f. Wooldridge (2002)).

For estimation, we first spell out the observed density, $f(y_t, m_{t+1}, I_t, K_{t+1}, m_t, u_t|l_t, k_t)$, as a mixture of the product of several latent conditional densities as follows:

$$
\begin{aligned}
& f\left(y_t, m_{t+1}, I_t, k_{t+1}, m_t, u_t|l_t, k_t\right) \\
= & \int g_{m'}\left(m_{t+1}|y_t, I_t, k_{t+1}, l_t, k_t, m_t, u_t, \omega_t\right) g_{k'}\left(k_{t+1}|y_t, I_t, l_t, k_t, m_t, u_t, \omega_t\right) \\
& g_y\left(y_t|I_t, l_t, k_t, m_t, u_t, \omega_t\right) g_I\left(I_t|l_t, k_t, m_t, u_t, \omega_t\right) g_m\left(m_t|l_t, k_t, u_t, \omega_t\right) \\
& g_u\left(u_t|l_t, k_t, \omega_t\right) g_\omega\left(\omega_t|l_t, k_t\right) d\omega_t \\
= & \int g_{m'}\left(m_{t+1}|k_{t+1}, \omega_t\right) g_{k'}\left(k_{t+1}|I_t, k_t\right) g_y\left(y_t|l_t, k_t, m_t, u_t, \omega_t\right) \\
& g_I\left(I_t|k_t, \omega_t\right) g_m\left(m_t|l_t, k_t, \omega_t\right) g_u\left(u_t|l_t, k_t, \omega_t\right) g_\omega\left(\omega_t|l_t, k_t\right) d\omega_t \\
= & \int g_{m'}\left(m_{t+1}|k_{t+1}, \omega_t\right) g_y\left(y_t|l_t, k_t, m_t, u_t, \omega_t\right) g_I\left(I_t|k_t, \omega_t\right) \\
& g_m\left(m_t|l_t, k_t, \omega_t\right) g_u\left(u_t|l_t, k_t, \omega_t\right) g_\omega\left(\omega_t|l_t, k_t\right) d\omega_t \cdot g_{k'}\left(k_{t+1}|I_t, k_t\right)
\end{aligned}
$$

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

$$\begin{aligned}
& (\hat{\beta}, \hat{g}_{m'}, \hat{g}_{\eta_t|.}, \hat{g}_I, \hat{g}_m, \hat{g}_m, \hat{g}_\omega) \\
= \; & \arg \max_{\left(\beta, g_{m'}, g_{\eta_t|.}, g_I, g_m, g_u, g_\omega\right)} \frac{1}{J} \sum_{j=1}^{J} \sum_{t=1}^{T} \ln \int_{-\infty}^{\infty} g_{m'}\left(m_{jt+1} | k_{jt+1}, \omega_{jt}\right) \\
& g_{\eta_t|(\omega_t, l_t, k_t, m_t, u_t)}\left(y_{jt} - \beta_l l_{jt} - \beta_k k_{jt} - \beta_m m_{jt} - \beta_u u_{jt} - \omega_{jt}\right) g_I\left(I_{jt} | k_{jt}, \omega_{jt}\right) \\
& g_m\left(m_{jt} | l_{jt}, k_{jt}, \omega_{jt}\right) g_u\left(u_{jt} | l_{jt}, k_{jt}, \omega_{jt}\right) g_\omega\left(\omega_{jt} | l_{jt}, k_{jt}\right) d\omega_{jt}.
\end{aligned}$$

Note that the sum of per-period likelihoods over t for each firm j is not the likelihood of the observation of firm j.

We refer readers to Chen (2007) for a comprehensive treatment of concrete procedures of the SNPML estimator. We can use artificial neural networks (Chen and White (1999)) to approximate the conditional density functions, and use Hermitian series (Gallant and Nychka (1987)) to approximate some of the conditional density functions if they are assumed to be independent of the variables in the conditioning set. The main cost of implementing the SNPMLE is the computational time. Our GMM estimator is a computationally less costly alternative to the SNPML estimator.