Estimating Production Functions with Fixed Effects*

Jose Miguel Abito

January 10, 2020

Abstract

I propose an estimation procedure that can accommodate fixed effects in the widely used proxy variable approach to estimating production functions. The proposed procedure allows unobserved productivity to have a permanent component in addition to a (nonlinear) Markov shock. In contrast to dynamic panel methods, the procedure does not rely on differencing out the fixed effect and thus is not limited to within-firm variation for identification. Finally, implementation is straightforward since it only entails adding a two stage least squares step using internal instruments.

*Abito: University of Pennsylvania, Wharton School, Business Economics and Public Policy, abito@wharton.upenn.edu. I would like to thank Uli Doraszelski, Joao Granja, Joe Harrington, John Lazarev, Arvind Magesan, Eric Mbakop, Lucija Muehlenbachs, Aviv Nevo, André Trindade and Alex Whalley for useful comments. I am indebted to Devesh Raval and Dan Ackerberg for the Chile data. All errors are mine.

1 Introduction

Estimation of production functions is a staple in several fields in economics including agriculture and resource economics, trade, macroeconomics, and industrial organization. An important econometric problem that has spurred a "search for identification" (Griliches and Mairesse, 1998) is the problem of simultaneity (Marschak and Andrews, 1944). Observed input choices of firms are not under the control of the econometrician but instead reflect optimal behavior of the firm. Thus, the relationship between inputs and output may depend on factors that are observed by the firm but not by the econometrician. Not accounting for these unobserved factors—often referred to as unobserved productivity—creates bias in estimates of the production function.

Different methods essentially impose different assumptions on unobserved productivity, with the econometrician facing a tradeoff in choosing the method most appropriate for the data and institutional setting. In this paper, I develop an estimation procedure that combines the strengths of the two most popular estimation methods and thus relaxes the tradeoff between the two.

Just as in the *dynamic panel approach* (Arellano and Bond, 1995; Blundell and Bond, 1998, 2000), I allow unobserved productivity to have both a permanent component (i.e. a fixed effect) and a time-varying component that follows a Markov process. However, I do not restrict the Markov process to be linear and the estimation method I propose in the paper does not require differencing of the fixed effect. My approach assumes the existence of an input (i.e. a proxy variable) that can be used to provide an expression for unobserved productivity as a function of observables, following the *proxy variable approach* to estimating production functions (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves and Frazer, 2015). However since proxy variable approaches do not allow for a permanent component in unobserved productivity, the main contribution of the paper is to offer a novel estimation procedure that accommodates this case.

A robust observation about productivity is that it is persistent and exhibits large dispersion even in narrowly defined industries (see Syverson (2011)). It seems natural to use methods that are robust to inclusion of a fixed effect. Current methods rely on differencing out the fixed effect which can lead to attenuation bias (Griliches and Hausman, 1986) and often exhibit poor finite sample performance (Blundell and Bond, 1998; Blundell and Bond, 2000).

The method I propose does not rely on differencing. Instead, I construct instruments from the data that are orthogonal to the fixed effect. This strategy is similar to the additional moment conditions suggested by Blundell and Bond (1998, 2000) which uses the level equation with first differences of inputs as instruments. However, instead of relying on an auxiliary assumption that the fixed effect is orthogonal to the growth in inputs—which are endogenous decision variables of the firm—orthogonality is between time-varying productivity shocks and the fixed effect, both of which are assumed exogenous in the model.

The proxy variable approach assumes that there exists a strictly monotonic mapping between unobserved productivity and a firm's decision variable (the proxy). The econometrician can then invert this mapping to express unobserved productivity as a function of observables, including the proxy variable. There are however two reasons why the proxy variable approach, as is often formulated, cannot accommodate a permanent fixed component in unobserved productivity.

First, as Ackerberg, Caves and Frazer (2015) and Ackerberg (2016) show, the proxy variable approach essentially rely on timing assumptions for identification. Basically the econometrician exploits the gap between when an action is taken and when unobservables (to the econometrician) such as productivity enter the firm's information set. When unobserved productivity includes a permanent component such as a fixed effect, then all actions, regardless on when they are made, will be correlated with the unobservable. To the best of my knowledge, there is currently no method that allows this case without sacrificing the attractive features of proxy variable methods.[1]

Second, if unobserved productivity also includes a fixed effect, then the inversion may fail. This is known as the scalar unobservable assumption. Similar to an extension to Gandhi, Navarro and Rivers (2017, Appendix O5-1), I assume that there is a proxy that is a strictly increasing function of the *sum* of the two components of unobserved productivity. Although the two components cannot be separately expressed as functions of observables, one can still invert the relationship to express the sum of the components as a function of observables.

The key step then is to think about the sum of the two components of unobserved productivity as having a measurement error form: the latent variable is the time-varying Markov component while the fixed component is the measurement error. This measurement error is not classical since it is correlated with input choices including proxy variables. Thus, I cannot use double measurements such as multiple proxies subject to independent measurement errors as instruments for identification (e.g. Hu, Huang and Sasaki, (2019)).

Instead, I rely on a different instrumental variable strategy that uses a *Berskon* instrument whereby the measurement error is correlated with the latent variable but independent of the instrument (Schennach, 2007). I construct Berkson instruments by taking first differences of the control function involving the proxy variable, which is equal to the sum of

[1] Gandhi, Rivers and Navarro (2017) allow for a fixed effect as an extension of their estimation procedure. However, they assume that the Markov component evolves as a linear (AR(1)) process and rely on differencing the fixed effect as in the dynamic panel approach.

More recently, Lee, Stoyanov and Zubanov (2019) consider a model where a fixed effect enters the *evolution* of productivity, i.e. $\omega_{it} = g(\omega_{it-1}) + c_i + \xi_{it}$ assuming the inversion of the proxy variable gives ω_{it}. This is similar to the set up in Theorem 1 of Asker, Collard-Wexler and De Loecker (2014) whereby the fixed effect eventually appears linearly in the estimating equation which can then be differenced out.

the permanent and Markov components of unobserved productivity and is a function of just observables. Finally, under the assumption that the Markov component is mean independent of the permanent component, this difference is also orthogonal to the fixed effect which allows construction of the necessary moment conditions for identification without needing to difference the estimating equation to eliminate the fixed effect.

An important additional benefit of the procedure is that it is easy to implement. Estimation of capital and labor output elasticities only involves adding a two stage least squares step to the standard proxy variable estimation procedure. Moreover, as mentioned earlier, the estimation procedure does not require additional proxies for instruments. Instead, "internal" instruments are constructed from the proxy's control function. Therefore the method is readily accessible to econometricians who are already planning to use some form of the proxy variable approach.

Using Monte Carlo experiments, I compare the performance of the proposed estimation procedure with some of the widely used estimators in the literature. Proxy variable approaches yield inconsistent estimators when there is a fixed effect hence the bias persists even in relativity large samples. In contrast, the estimation procedure that uses a Berkson instrument is robust to inclusion of the fixed effect and performs as well as the standard proxy variable approach when there is no fixed effect. Compared to methods that rely on differencing, the results point to better finite sample performance in terms of bias and precision. Finally, the system GMM approach (Arellano and Bover, 1995; Blundell and Bond, 1998, 2000) only works if the fixed effect is orthogonal to the growth in inputs. Since the growth in capital is driven by investment, this essentially requires that the way investment affects growth is orthogonal to a state variable such as a fixed effect. Indeed I do observe biased estimates when the role of the fixed effect in determining investment is larger.

Finally, I assess the empirical performance of the proposed estimation procedure using data from Chilean manufacturing plants as in Levinsohn and Petrin (2003) and Ackerberg, Caves and Frazer (2006). I find that estimates can significantly differ compared to existing methods. In particular, I find that estimates of the capital coefficient tend to be higher using the proposed estimation procedure.

The paper is related to the literature that relaxes the scalar unobservable assumption in proxy variable methods, specifically to the strand of the literature that develops methods to handle measurement error in inputs and proxies. Hu, Huang and Sasaki (2019) establish identification of such a model using results from the nonlinear errors-in-variables literature (i.e. Hu and Schennach (2008)). Their strategy is to use multiple proxies, which they assume as having independent measurement errors.[2] In contrast to this double measurement strategy, I use an instrumental variables strategy with a Berkson instrument (Schennach, 2007). Finally, Kim, Petrin and Song (2016) and Collard-Wexler and De Loecker (2017) both focus on the case where capital is measured with error. Similar to Hu, Huang and Sasaki (2019), Kim, Petrin and Song (2016) use results form the nonlinear errors-in-variables literature and suggest an estimation procedure based on sieves. In contrast, Collard-Wexler and De Loecker (2017) leverage on a Cobb-Douglas production function to pose the problem as a linear errors-in-variables model, and use investment as an instrument for mismeasured capital. They emphasize the attractiveness of methods that are easy to implement and are readily accessible.

The next section presents the model and key assumptions maintained throughout the paper. Section 3 discusses both identification and estimation. Section 4 contains the Monte Carlo experiments and section 5 applies the proposed estimation procedure to Chilean data.

[2] Ackerberg, Benkard, Berry and Pakes (2007) also suggest the general idea of bringing in additional proxies to relax the scalar unobservable assumption. The idea is to use information from the additional proxies to control for the other unobservables.

Section 6 concludes. For ease of exposition, Table 1 lists the abbreviations of some of the key papers I repeatedly refer to.

2 Model

Consider the following production function:[3]

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + w_{it} + \varepsilon_{it}.$$

The variable y_{it} is firm i's log output at time t, and l_{it} and k_{it} are the labor and capital inputs (in logs) respectively. These inputs and output are observed by the econometrician. In contrast, the econometrician does not observe ε_{it} and w_{it}. The unobservable ε_{it} represents shocks to the production function that are also unobserved by the firm at the time it decides on its time t inputs. This unobservable can represent measurement error in output or other factors that affect output (holding inputs fixed) that the firm cannot predict. On the other hand, w_{it} represents factors that are observed (or predicted) by the firm and taken into account in its input choices. The unobservable w_{it} is often thought of as a firm's "productivity" which is observed by the firm but not by the econometrician.

The goal of the econometrician is to estimate parameters β_l and β_k, which are the output elasticities for labor and capital respectively. In this paper, I focus on the problem of simultaneity (Marschak and Andrews, 1944). Since the firm takes productivity w_{it} into account when choosing its inputs, an OLS regression of output on inputs will generate biased estimates.

As a solution to the simultaneity problem, the proxy variable approach (e.g. OP, LP and ACF) exploits a "structural" function that maps a firm's productivity to an endogenous

[3]I discuss the approach in the context of value-added production functions to keep the discussion as close to OP, LP and ACF.

decision. In OP, investment is assumed to be strictly increasing in productivity and thus can be inverted to express it as a function of observables (e.g. investment, capital stock and age). That is, if I_{it} refers to investment and

$$I_{it} = f(w_{it}, k_{it}, age_{it}),$$

where f is strictly increasing in ω_{it}, then

$$w_{it} = f^{-1}(I_{it}, k_{it}, age_{it})) = h(x_{it})$$

where $x_{it} = (I_{it}, k_{it}, age_{it})$. LP suggest using intermediate inputs as a proxy variable since data usually contain a lot of observations with zero investment. As in OP, LP assumes that the intermediate input is a strictly increasing function of productivity, and thus can be inverted to control for the latter.

An important assumption in proxy variable methods is that unobserved productivity evolves as a Markov process. Formally, unobserved productivity w_{it} is assumed to be equal to

$$\omega_{it} = E[\omega_{it}|\mathcal{I}_{it-1}] = E[\omega_{it}|\omega_{it-1}] + \xi_{it} = g(\omega_{it-1}) + \xi_{it}$$

where $\mathcal{I}_{it-1}$ is firm i's information set at time $t-1$. This Markov assumption implies $E[\xi_{it}|\mathcal{I}_{it-1}] = 0$, and thus one can generate moment conditions based on the timing of decisions and the information used when making these decisions (ACF; Ackerberg, 2016). For example, since past decisions on the labor input belongs to $\mathcal{I}_{it-1}$, then $E[\xi_{it}|l_{it-j}] = 0$ for all $j \geq 1$. These timing assumptions are powerful since not only do they provide moment conditions for estimation, they also allow the econometrician to generate and use "internal" instruments, e.g. lagged input choices.

The timing assumption breaks down when unobserved productivity has a permanent component. To see this, suppose $w_{it} = \omega_{it} + a_i$ where ω_{it} satisfies the Markov assumption as above. Although $E[\xi_{it}|\mathcal{I}_{it-1}] = 0$ remains valid, $E[\xi_{it} + a_i|\mathcal{I}_{it-1}] = 0$ is not. In fact, when unobserved productivity has a permanent component, all input choices will be correlated with a_i regardless of how far in the past the decision was made.

One potential solution is to eliminate a_i from the estimating moment conditions. This is possible under the proxy variable approach if either a_i does not enter the proxy equation, or if it does, $g(\cdot)$ is linear (GNR). If a_i does not enter the proxy equation, then $\omega_{it-1} = h(x_{it-1})$ and we can rewrite the production function as

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + g(h(x_{it-1})) + (\xi_{it} + a_i + \varepsilon_{it}).$$

We can then eliminate a_i from the estimating equation by taking first differences.

Suppose instead that a_i enters the proxy variable equation. This violates the so-called scalar unobservable assumption hence one can no longer express ω_{it-1} as a function of *only* observables. In Appendix O5-1 of GNR, they extend their model to the case where unobserved productivity has a fixed effect component. They assume that the proxy variable is a strictly increasing function of the sum $(\omega_{it} + a_i)$, and thus $h(x_{it}) = \omega_{it} + a_i$ in my notation. Such an assumption is tenable if the proxy is static, otherwise the proxy is likely to be a function of the state variables ω_{it} and a_i separately, and not only their sum (Ackerberg, 2016). Following GNR, suppose this assumption holds, e.g. we use intermediate inputs that are assumed to be static. The production function becomes

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + g(h(x_{it-1}) - a_i) + (\xi_{it} + a_i + \varepsilon_{it}).$$

In this case, we need $g(\cdot)$ to be linear in order to difference out a_i, similar to the dynamic panel approach (Arellano and Bond, 1991; BB98; BB00). Because of differencing, a drawback of the dynamic panel approach is that identification relies on within-firm variation, which may not be a good source of variation when output and inputs are highly serially correlated

(Griliches and Hausman, 1986; BB98).[4]

In the next section, I show how to identify and estimate output elasticities in a model that largely follows the proxy variable approach but allows for a fixed effect in unobserved productivity. Unless stated otherwise, I maintain the following assumptions throughout the paper:

1. **Unobserved productivity:**

$$w_{it} = \omega_{it} + a_i$$

2. **Markov:**

$$\omega_{it} = E[\omega_{it}|\mathcal{I}_{it-1}] + \xi_{it} = g(\omega_{it-1}) + \xi_{it}$$

where $g(\cdot)$ can be nonlinear.

3. **Timing:**

$$E[\xi_{it}|k_{it-j}] = 0 \text{ for all } j \geq 0$$

4. **Proxy variable:**

$$\omega_{it} + a_i = h(x_{it})$$

where x_{it} is a vector of observables including the proxy variable.

5. **Mean independence:**

$$E[a_i|\xi_{it}] = 0 \text{ for all } t.$$

The first assumption assumes the same error structure as in BB00 and GNR. The second and third assumptions are standard in the proxy variable literature, while the fourth assumption follows GNR as discussed above. Finally the fifth assumption consists of normalizing the

[4] GNR (Table O6.5) find lower and noisier capital estimates for Colombia using an extension of their method that allows for fixed effects.

mean of the fixed effect to zero, and also the assumption that unexpected innovations ξ_{it} are not informative of a_i. Note that mean independence is sufficient as opposed to full independence. The purpose of the fifth assumption will be clearer in the next section.

3 Identification and estimation

Using the previous assumptions, we can rewrite the production function as

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + g(\omega_{it-1}) + (\xi_{it} + a_i + \varepsilon_{it}) \tag{1}$$

There are two issues in estimating the coeffcients β_l and β_k. First, even if we observe ω_{it-1}, input choices (including the proxy) are still endogenous because a_i is unobserved and correlated with the inputs. Second, ω_{it-1} is, in fact, unobserved and cannot be simply replaced by $h(x_{it})$ as in the standard proxy variable case where $a_i = 0$.

To solve the first issue, I use a vector of instruments z such that

$$E[\xi_{it} + a_i + \varepsilon_{it}|z] = 0.$$

Consider the decomposition

$$\begin{aligned} l_{it} &= E[l_{it}|z] + \eta_{lit} = \widehat{l}_{it} + \eta_{lit} \\ k_{it} &= E[k_{it}|z] + \eta_{kit} = \widehat{k}_{it} + \eta_{kit} \end{aligned}$$

where by construction, $E[\eta_{lit}|z] = E[\eta_{kit}|z] = 0$. I can then rewrite the estimating equation as

$$y_{it} = \beta_l \widehat{l}_{it} + \beta_k \widehat{k}_{it} + g(\omega_{it-1}) + (\beta_l \eta_{lit} + \beta_k \eta_{kit} + \xi_{it} + a_i + \varepsilon_{it}). \tag{2}$$

Thus, if ω_{it-1} were observed, then one can estimate β's by applying OLS to equation 2.

As for the second issue, although ω_{it-1} is unobserved, we do observe the sum $\omega_{it-1} + a_i$ up to some unknown function $h(\cdot)$ of observables:

$$h(x_{it-1}) = \omega_{it-1} + a_i.$$

We can think of ω_{it-1} as a latent variable that we observe subject to some measurement error a_i. A common solution to handle nonlinear errors-in-variables models is to use a second *independent* measurement as an instrument. In the context of production function estimation, Hu, Huang and Sasaki (2019)), for example, assume the econometrician has multiple proxies with independent measurement errors. However in my case, the measurement error a_i is potentially correlated with all proxies and inputs, and thus the strategy of using multiple proxies will not work.

Schennach (2007) establishes nonparametric identification of a nonlinear errors-in-variables model using an instrument that has a *Berkson-error* form (Berkson, 1950; Chen, Hong and Nekipelov, 2011). Instead of assuming independence between the latent variable and the measurement error, a measurement with Berkson-error satisfies the "causal equation"

$$\omega_{it-1} \quad = \quad m(z) + v$$

where the instrument z and the error v are independent, $m(\cdot)$ is some function to be estimated[5], and $E[\xi_{it} + \varepsilon_{it}|z, v] = 0$ and $E[a_i|z, v, \xi_{it}] = 0$. In contrast to classical measurement error, the latent variable is correlated with the measurement error v. The independence assumption instead is between $m(z)$ and the measurement error. Given these assumptions,

[5]The function $m(\cdot)$ is identified and can be estimated using data on $h(x_{it-1})$ and z. To see this note

$$h(x_{it-1}) = \omega_{it-1} + a_i = m(z) + (v + a)$$

where $E[v + a|z] = 0$ by assumption.

Schennach (2007) shows that β_l, β_k and $g(\cdot)$ are identified.[6]

Nonlinear errors-in-variables models are often complicated to estimate and implementation is often still challenging (Collard-Wexler and De Loecker, 2016). It turns out that if we assume $g(\cdot)$ is a polynomial, then estimating the coefficients β_l and β_k is pretty straightforward. To show this, I rely on results from Hausman, Newey, Ichimura and Powell (1991), where they show identification using Berkson instruments when $g(\cdot)$ is a polynomial of known degree P:

$$g(\omega) = \sum_{j=0}^{P} \beta_j \omega^j.$$

Under the polynomial assumption, the estimating equation (2) becomes

$$y_{it} = \beta_l \widehat{l}_{it} + \beta_k \widehat{k}_{it} + \sum_{j=0}^{P} \gamma_j \widehat{\omega}_{it-1}^j + e_{it}$$

where $\widehat{\omega}_{it-1} = m(z)$,

$$\gamma_j \equiv \left(\sum_{q=j}^{P} \beta_j \binom{q}{j} E(v_{it}^q) \right)$$

and

$$e_{it} \equiv \sum_{j=0}^{P} \widehat{\omega}_{it-1}^j \left(\sum_{q=j}^{P} \beta_j \binom{q}{j} \left[v_{it}^q - E(v_{it}^q) \right] \right) + (\beta_l \eta_{lit} + \beta_k \eta_{kit} + \xi_{it} + a_i + \varepsilon_{it}).$$

As long as the instrument vector z satisfies $E[e_{it}|z] = 0$, then β_l and β_k can be identified using the moment condition

$$E[e_{it}|\widehat{l}_{it}, \widehat{k}_{it}, \widehat{\omega}_{it-1}] = 0. \tag{3}$$

Therefore we can estimate the output elasticities by running an OLS regression of y_{it} on $\widehat{l}_{it}$, $\widehat{k}_{it}$, and powers of $\widehat{\omega}_{it-1}$.

[6] Schennach's (2007) main model does not include regressors such as l and k. A model similar to equation 2 is mentioned as a simple extension. See Schennach (2007, p. 222).

3.1 Internal Berkson instruments

The feasibility of the previous identification strategy hinges on the existence of a set of instruments that satisfy the properties of a Berkson instrument. In the context of our model, we need

1. $E[\eta_{lit}|z] = E[\eta_{kit}|z] = 0$,

2. $E[\xi_{it} + a_i + \varepsilon_{it}|z] = 0$, and

3. $E[v_{it}^q|z] = E[v_{it}^q]$ for all $q = 1, 2, ..., P$.

The first condition follows by construction since

$$\eta_{lit} = l_{it} - E[l_{it}|z] \text{ and } \eta_{kit} = k_{it} - E[k_{it}|z]$$

and therefore I focus on the second and third conditions.

I propose using changes in unobserved productivity and its lags as instruments:

$$\Delta\omega_{it-j} = \Delta h(x_{it-j})$$

for $j \geq 1$. The advantage of these instruments is that they are "internal" in that they can be constructed from the data at hand. These instruments are similar to the instruments proposed by Blundell and Bond (1998, 2000) to improve the finite sample performance of the dynamic panel approach. There, a crucial assumption is that changes in endogenous variables, i.e. Δl_{it} and Δk_{it}. are mean independent with a_i. In my case, mean independence is between a_i and a function of ξ_{it-j}. In addition, since $E[\xi_{it}|\mathcal{I}_{it-1}] = 0$ and $\Delta\omega_{it-j} \in \mathcal{I}_{it-1}$, then indeed $E[\xi_{it} + a_i + \varepsilon_{it}|\Delta\omega_{it-j}] = 0$.

Finally, we need the instruments to satisfy $E[v_{it}^q|z] = E[v_{it}^q]$ for all $q = 1, 2, ..., P$. This is implied by the structure of the Berskon-error, i.e. z ($m(z)$) is independent of

$$v = \omega_{it-1} - m(z).$$

Mean independence between z and v is satisfied by construction since I estimate $m(z) = E[h(x_{it-1})|z] = E[\omega_{it-1}|z]$. What is then needed is mean independence between z and higher moments of v (up to the P-th moment).[7]

3.2 Comparison of identifying moment conditions

Following Ackerberg (2016), consider a model where $g(\cdot)$ is linear and $\varepsilon = 0$. The goal of this section is to compare the moment conditions that identify the parameters from the proposed estimation procedure with the proxy variable and dynamic panel approaches.

Under the given simplifications, the identifying moment condition using the proposed estimation strategy is

$$E\left[(y_{it} - \beta_l l_{it} - \beta_k k_{it} - \rho h(x_{it-1})) \cdot \begin{pmatrix} \Delta\omega_{it-1} \\ \Delta\omega_{it-2} \\ \Delta\omega_{it-3} \end{pmatrix}\right] = \mathbf{0} \tag{4}$$

This moment condition is the same as the moment condition in ACF (Ackerberg's (2016) equation 5) except that the instrument vector in ACF is $z = (h(x_{it-1}), l_{it}, k_{it})$. Since $y_{it} - \beta_l l_{it} - \beta_k k_{it} - \rho h(x_{it-1}) = \xi_{it} + (1-\rho)a_i$ and

$$E\left[a_i \cdot \begin{pmatrix} h(x_{it-1}) \\ l_{it} \\ k_{it} \end{pmatrix}\right] \neq \mathbf{0},$$

the moment condition in ACF no longer holds unless $a_i = 0$. In contrast, the instrument vector in equation 4 has a_i differenced out.

[7] If $g(\cdot)$ is linear, then mean independence between z and v is sufficient. Also, if we observe *all* lags of $\Delta\omega_{it-j}$, i.e. $z = (\Delta\omega_{it-1}, \Delta\omega_{it-2}, ..., \Delta\omega_{i0})$, then we essentially observe ω_{it-1} and $m(z)$ and v are trivially independent.

The moment condition for the dynamic panel approach is given by

$$E\left[(\Delta\Delta_\rho y_{it} - \beta_l \Delta\Delta_\rho l_{it} - \beta_k \Delta\Delta_\rho k_{it}) \cdot \begin{pmatrix} l_{it-3} \\ k_{it-2} \\ y_{it-3} \end{pmatrix}\right] = \mathbf{0} \tag{5}$$

where, for example, $\Delta\Delta_\rho y_{it} = (y_{it} - \rho y_{it-1}) - (y_{it-1} - \rho y_{it-2})$. This moment condition is valid since a_i is differenced away. However, BB98 find that this moment condition leads to poor finite sample properties (i.e. biased and imprecise estimates) due to weak identification. If one is willing to assume the assumption

$$E[a_i | \Delta l_{it}, \Delta k_{it}] = 0,$$

then they suggest adding the following moment condition:

$$E\left[(\Delta_\rho y_{it} - \beta_l \Delta_\rho l_{it} - \beta_k \Delta_\rho k_{it}) \cdot \begin{pmatrix} \Delta l_{it-1} \\ \Delta k_{it} \end{pmatrix}\right] = \mathbf{0}. \tag{6}$$

The estimator based on moment conditions 5 and 6 is sometimes referred to as the "system GMM" estimator.

Observe that

$$\Delta_\rho y_{it} - \beta_l \Delta_\rho l_{it} - \beta_k \Delta_\rho k_{it} = y_{it} - \beta_l l_{it} - \beta_k k_{it} - \rho \left(y_{it-1} - \beta_l l_{it-1} - \beta_k k_{it-1}\right).$$

Since in a model without ε,

$$h(x_{it-1}) = y_{it-1} - \beta_l l_{it-1} - \beta_k k_{it-1},$$

the additional moment used in the system GMM estimator can be rewritten as

$$E\left[(y_{it} - \beta_l l_{it} - \beta_k k_{it} - \rho h(x_{it-1})) \cdot \begin{pmatrix} \Delta l_{it-1} \\ \Delta k_{it} \end{pmatrix}\right] = \mathbf{0}. \tag{7}$$

But this is the same as moment condition 4, except for the choice of instruments. Note though that both sets of instrument essentially differences out the fixed effect a_i. Therefore, one can view the proposed estimation procedure as combining the assumption used in the system GMM estimator, with the proxy variable assumption that allows us to construct the instruments $\Delta h(x_{it-j}) = \Delta\omega_{it-j}$.

3.3 Estimation

I now discuss estimation of the model under the assumption that $g(\cdot)$ is a polynomial of known degree as in Hausman, Newey, Ichimura and Powell (1991). The estimation procedure is iterative and proceeds in several steps.

Step 0 is basically the first stage in OP, LP or ACF. That is, I get an estimate $\widehat{\Phi}_{it}$ of

$$\Phi_{it}(l_{it}, k_{it}, x_{it}) = \beta_l l_{it} + \beta_k k_{it} + h(x_{it})$$

by estimating the partial linear model

$$y_{it} = \Phi_{it}(l_{it}, k_{it}, x_{it}) + \varepsilon_{it}.$$

Step 1 is the start of the iteration. Given a guess $(\widetilde{\beta}_l, \widetilde{\beta}_k)$ and the estimate $\widehat{\Phi}_{it}$ from Step 0, construct instruments:

$$\Delta h(x_{it-j}) = \Delta\widehat{\Phi}_{it-j} - \widetilde{\beta}_l \Delta l_{it-j} - \widetilde{\beta}_k \Delta k_{it-j}$$

for $j \geq 1$. We need to generate at least three instruments (three lags) to be exactly identified.

Once the instruments are constructed, we proceed to **Step 2** which can be thought of as the first stage in two stage least squares. That is, we estimate the conditional expectations

$$\begin{aligned}
\widehat{l}_{it} &= E[l_{it}|z] \\
\widehat{k}_{it} &= E[k_{it}|z] \\
\widehat{\omega}_{it-1} &= E[h(x_{it-1})|z]
\end{aligned}$$

where z is the vector of instruments, e.g. $z = (\Delta h(x_{it-1}), \Delta h(x_{it-2}), \Delta h(x_{it-3}))$.

Step 3 then is just an OLS regression of y_{it} on $\widehat{l}_{it}$, $\widehat{k}_{it}$ and powers of $\widehat{\omega}_{it}$.[8] This will give estimates $\widehat{\beta}_l(\widetilde{\beta}_l, \widetilde{\beta}_k)$ and $\widehat{\beta}_k(\widetilde{\beta}_l, \widetilde{\beta}_k)$. We repeat Steps 1 to 3 until we find the fixed point:

$$\widetilde{\beta}_l = \widehat{\beta}_l(\widetilde{\beta}_l, \widetilde{\beta}_k) \text{ and } \widetilde{\beta}_k = \widehat{\beta}_k(\widetilde{\beta}_l, \widetilde{\beta}_k).$$

For OP or LP, the proxy is not a function of l_{it} (i.e. l_{it} has independent variation), and so $\Phi_{it}(k_{it}, x_{it}) = \beta_k k_{it} + h(x_{it})$ and β_l can be separately identified in Step 0. Moreover, we only need a guess $\widetilde{\beta}_k$ and two lags of the instrument to be exactly identified.

4 Monte Carlo experiments

I perform Monte Carlo experiments to compare the finite sample performance of the proposed estimation procedure. To simplify the simulations, I assume a data generating process where LP is identified and consistent if $a_i = 0$. Specifically, I assume log labor (in reduced form) is given by

$$l_{it} = \omega_{it} + a_i + u_{it}$$

where u_{it} is iid and distributed standard normal. I also assume reduced forms for investment and log materials (i.e. the proxy):

$$I_{it} = \exp(0.1\omega_{it} + a_i + k_{it})$$

and

$$m_{it} = \omega_{it} + a_i + k_{it}$$

respectively. Finally, the capital accumulation equation is $K_{it} = 0.95K_{it-1} + I_{it-1}$ where initial log capital, $\log K_{i0}$, is a random draw from a standard normal distribution.

[8] One can also use $\hat{\Phi}_{it}$ from Step 0 as the dependent variable in the OLS regression instead of y_{it}.

To generate productivity, I assume the fixed component a_i is distributed iid standard normal. For the Markov component, I consider both linear and nonlinear processes:

$$\omega_{it} = \rho\omega_{it-1} + \xi_{it}$$

and

$$\omega_{it} = \rho(\omega_{it-1} - 0.01\omega_{it-1}^3) + \xi_{it}$$

where $\rho \in \{0.2, 0.8\}$, ξ_{it} is iid standard normal, and $\omega_{i0} = 0$.

I run 1000 replications, with $N = 250$ firms and $T = 5$ time periods. I estimate the labor and capital coefficients using OLS, first differencing (FE), the dynamic panel approach with and without the stationarity assumption (DPS and DP respectively), LP, and the approach proposed in the paper (LP + Berkson IV, or LPIV).

Table 2 contains the results for $\rho = 0.2$ and $g(\cdot)$ is linear. Table 3 corresponds to the case where $g(\cdot)$ is nonlinear. The first three columns of numbers are the mean, standard deviation and root mean squared errors over 1000 simulation runs when productivity does not have a fixed effect, while the next three is when productivity is equal to $\omega_{it} + a_i$. Tables 4 and 5 instead contain the results for $\rho = 0.8$. For all of these tables, the true values are $\beta_l = 0.7$ and $\beta_k = 0.3$.

Focusing on the OLS estimates, the bias is larger for β_l compared to β_k. This observation seems to apply to the other estimation procedures as well, except for FE. The capital coefficient under FE is severely downward biased, consistent with what have been observed in the literature (e.g. Griliches and Hausman (1986), and the famous quote from Griliches and Mairesse (1998, p. 178)).

Consistent with BB98, the DP estimates are biased and are terribly imprecise. DPS on the other hand significantly improves the estimates. In fact, the estimate for the capital coefficient seem very reliable across specifications. However, even with DPS, the labor coefficient remains biased and imprecise in almost all specifications except for when there is no fixed effect, $\rho = 0.8$ and $g(\cdot)$ is linear (Table 4).

LP exhibits good finite sample properties in all specification with $a_i = 0$. Since the simulation assumptions allow estimation of β_l in the first stage, the labor coefficient does not suffer from bias even when there is a fixed effect. However, the capital coefficient is biased upwards when $a_i \neq 0$. The bias seems to get worse when ρ is lower, i.e. the persistence in productivity (and other variables) are mostly due to the fixed component rather than the Markov component.

LPIV performs well in all specifications. It is robust to inclusion of a fixed effect, and it does not suffer from a potential weak identification problem as in DP. Therefore LPIV can be seen as taking advantage of both proxy variable and dynamic panel methods.

5 Empirical application

To assess the performance of the proposed estimation procedure on actual data, I estimate value-added production functions for Chilean manufacturing industries.[9] Several papers in the production function estimation literature have used the Chile dataset including LP, a working paper version of ACF (Ackerberg, Caves and Frazer, 2006), and GNR. I focus on the same four industries examined in the literature and use data from 1979-1986. The four industries are Food Products (ISIC 311), Textiles (ISIC 321), Wood Products (ISIC 331) and Metals (ISIC 381).

To compare the proposed estimation procedure, I estimate the model using OLS, FE and LP. I implement LP using the canned Stata command levpet. I estimate two version of the proposed estimation procedure. The first one labeled LPIV estimates the labor coefficient in

[9] I thank Devesh Raval for graciously sharing his version of the Chile dataset.

the first stage similar to LP. The second one labeled ACFIV estimates the labor coefficient in the second stage. To keep as close to LP, I implement LPIV using the same first stage as levpet to estimate the labor coefficient, and then performing a simple grid search over $\{0, 0.001, 0.002, ..., 1\}$ to find the capital coefficient. For ACFIV, I code the procedure in Matlab and use the command fmincon together with a multistart procedure with 30 different initial values.

Table 6 contains the estimates. The first observation is that estimates from FE are highly unreliable as all capital coefficients are negative. Ackerberg, Caves and Frazer (2006) also find a negative coefficient for Wood products while the estimate for the capital coefficient for the other industries are close to zero. Next, the labor coefficient from OLS is higher compared to the other estimators, which may indicate the presence of transmission bias. As for the capital coefficient, the direction of the potential bias goes both ways when comparing OLS with LP. Finally, the estimate for the capital coefficient using the proposed estimation procedure is generally higher compared to the other estimators. Estimates using LPIV tend to be unrealistically high (around 0.9) while estimates using ACFIV looks more reasonable.

It is difficult to ascertain the exact reason for the differences in estimates without prior knowledge about the setting and the underlying data generating process. Moreover, unlike for flexible inputs such as labor were interpretation of the bias may be straightforward, a dynamic input such as capital may interact with both the transitory and permanent components of unobserved productivity in more complex ways, making it difficult to sign the potential bias for such inputs. Nevertheless, the proposed estimator provides a robust and relatively reliable method that can easily be added to the suite of estimators that a wide range of researchers typically use.

6 Conclusion

In this paper, I show how one can allow for a fixed effect component in the widely used proxy variable method. There are two main benefits of the procedure. First, the procedure does not rely on differencing and also allows for a nonlinear Markov component. Second, the procedure is easy to implement as it only entails adding a two stage least squares step. Since the instruments used are constructed from functions of the proxy and its lags, the data requirements are the same as in most applications of proxy variable methods. I illustrate the performance of the estimation procedure using Monte Carlo simulations and an empirical example using data from Chilean manufacturing firms. The procedure performs as well as other existing estimation procedures in settings where assumptions of these estimators hold, and works better for settings where these estimators are expected to fail.

References

Ackerberg, D. A .(2016), "Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context," working paper.

Ackerberg, D. A., L. Benkard, S. Berry and A. Pakes (2007), "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, Amsterdam: North-Holland.

Ackerberg, D. A., K. Caves and G. Frazer (2006), "Structural Identification of Production Functions," working paper.

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

Arellano, M. and S. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations,"*Review of Economic Studies*, 58:2, 277-297.

Arellano, M. and O. Bover (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics* 68:1, 29-51.

Asker, J. A. Collard-Wexler and J. De Loecker (2014), "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 122:5, 1013-1063.

Berkson, J. (1950), "Are There Two Regressions?" *Journal of the American Statistical Association*, 45:250, 164-180..

Blundell, R. and S. Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.

Blundell, R. and S. Bond (2000), "GMM Estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19:3, 321-340.

Chen, X., H. Hong and D. Nekipelov (2011), "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49:4, 901-937.

Collard-Wexler, A. and J. De Loecker (2016), "Production Function Estimation with Measurement Error in Inputs," working paper.

Gandhi, A., S. Navarro and D. Rivers, "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, forthcoming.

Griliches, Z., and J. Hausman (1986), "Errors in Variables in Panel Data," *Journal of Econometrics*, 31, 93-118.

Griliches, Z., and J. Mairesse (1998), "Production Functions: The Search for Identification," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, ed. by S. Strm, Cambridge, U.K.: Cambridge University Press.

Hausman, J., W. Newey, H. Ichimura, and J. Powell (1991), "Measurement Errors in Polynomial Regression Models," *Journal of Econometrics*, 50, 273-295.

Hu, Y., G. Huang and Y. Sasaki (2019), "Estimating production functions with robustness against errors in the proxy variables," *Journal of Econometrics*, forthcoming.

Hu, Y. and S. Schennach (2008), "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76:1, 195-216.

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.

Lee, Stoyanov and Zubanov (2019), "Olley and Pakes-style Production Function Estimators with Firm Fixed Effects," *Oxford Bulletin of Economics and Statistics*, 81:1, 79-97.

Levinsohn, J. and A. Petrin (2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70:2, 317-341.

Marschak, J. and W. Andrews (1944), "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12:3/4, 143-205.

Olley, S. and A. Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64:6, 1263-1297.

Schennach, S. M. (2007), "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 75:1, 201-239.

Syverson, C. (2011), "What Determines Productivity?" *Journal of Economic Literature*, 49:2, 326-365.

Table 1: Abbreviations of key papers cited

ACF	Ackerberg, Caves and Frazer (2015)
BB98	Blundell and Bond (1998)
BB00	Blundell and Bond (2000)
GNR	Gandhi, Rivers and Navarro (2017)
LP	Levinsohn and Petrin (2003)
OP	Olley and Pakes (1996)

Table 2: $\rho = 0.2$, linear

		ω_{it}			$\omega_{it} + a_i$		
		Mean	Std	RMSE	Mean	Std	RMSE
OLS	β_l	1.2078	0.0234	0.5083	1.3400	0.0237	0.6404
	β_k	0.3159	0.0216	0.0268	0.3797	0.0174	0.0816
FE	β_l	1.1222	0.0309	0.4233	1.1338	0.0319	0.4350
	β_k	0.0309	0.0382	0.2718	0.1254	0.0295	0.1771
DP	β_l	1.0602	0.6987	0.7860	1.0827	0.6944	0.7928
	β_k	0.4111	0.4411	0.4549	0.4278	0.4433	0.4614
DPS	β_l	1.2562	0.6749	0.8745	1.2311	0.3501	0.6361
	β_k	0.3026	0.0259	0.0260	0.3015	0.0275	0.0276
LP	β_l	0.7001	0.0274	0.0274	0.7006	0.0286	0.0286
	β_k	0.2866	0.1255	0.1262	0.4513	0.0483	0.1588
LPIV	β_l	0.7001	0.0274	0.0274	0.7006	0.0286	0.0286
	β_k	0.3000	0.0337	0.0337	0.3015	0.0374	0.0374

Note: I estimate the following models 1000 times and compute the mean, standard deviation and root mean squared error of the estimates. For each run, $N = 250$ and $T = 5$. FE refers to estimation via first differencing to remove a_i. DP and DPS are estimates based on the dynamic panel approach. DP only uses the differenced equation with levels as instruments while DPS uses the levels equation with differenced variables as instruments. LP corresponds to Levinsohn and Petrin (2003) where β_l is estimated in the first stage. PIV is the approach proposed in the paper and shares the first stage of LP (hence the estimates for β_l under LP and PIV are identical). The true values for the coefficients of interest are $\beta_l = 0.7$ and $\beta_k = 0.3$.

Table 3: $\rho = 0.2$, nonlinear

		ω_{it}			$\omega_{it} + a_i$		
		Mean	Std	RMSE	Mean	Std	RMSE
OLS	β_l	1.2045	0.0233	0.5050	1.3395	0.0237	0.6400
	β_k	0.3105	0.0217	0.0241	0.3784	0.0174	0.0803
FE	β_l	1.1336	0.0308	0.4347	1.1462	0.0319	0.4474
	β_k	0.0181	0.0388	0.2845	0.1179	0.0300	0.1846
DP	β_l	1.0743	0.6970	0.7911	1.1145	0.6891	0.8041
	β_k	0.4172	0.4591	0.4738	0.4248	0.4427	0.4600
DPS	β_l	1.3049	0.6693	0.9021	1.3290	0.3533	0.7214
	β_k	0.3015	0.0256	0.0257	0.3017	0.0260	0.0261
LP	β_l	0.7001	0.0274	0.0274	0.7006	0.0286	0.0286
	β_k	0.2847	0.1420	0.1428	0.4596	0.0353	0.1635
LPIV	β_l	0.7001	0.0274	0.0274	0.7006	0.0286	0.0286
	β_k	0.3000	0.0330	0.0330	0.3020	0.0371	0.0371

See note in Table 2.

Table 4: $\rho = 0.8$, linear

		ω_{it}			$\omega_{it} + a_i$		
		Mean	Std	RMSE	Mean	Std	RMSE
OLS	β_l	1.3216	0.0228	0.6220	1.3903	0.0231	0.6907
	β_k	0.3904	0.0194	0.0925	0.3978	0.0162	0.0991
FE	β_l	1.0427	0.0318	0.3442	1.0450	0.0327	0.3465
	β_k	0.2045	0.0309	0.1003	0.2307	0.0251	0.0737
DP	β_l	1.0768	0.6870	0.7835	1.0819	0.6781	0.7783
	β_k	0.4020	0.3983	0.4111	0.4162	0.4290	0.4444
DPS	β_l	0.6978	0.1942	0.1942	0.7522	0.1829	0.1902
	β_k	0.2987	0.0482	0.0483	0.2982	0.0480	0.0480
LP	β_l	0.7000	0.0275	0.0275	0.7005	0.0286	0.0286
	β_k	0.2645	0.1363	0.1409	0.3196	0.1152	0.1168
LPIV	β_l	0.7000	0.0275	0.0275	0.7005	0.0286	0.0286
	β_k	0.2966	0.0639	0.0640	0.2977	0.0636	0.0637

See note in Table 2.

Table 5: $\rho = 0.8$, nonlinear

		ω_{it}			$\omega_{it} + a_i$		
		Mean	Std	RMSE	Mean	Std	RMSE
OLS	β_l	1.2630	0.0410	0.5645	1.3634	0.0346	0.6643
	β_k	0.3489	0.0211	0.0532	0.3842	0.0182	0.0862
FE	β_l	1.0988	0.0577	0.4030	1.1055	0.0577	0.4096
	β_k	0.1031	0.0367	0.2003	0.1658	0.0283	0.1371
DP	β_l	1.0334	0.6858	0.7626	1.0550	0.6861	0.7725
	β_k	0.3959	0.3767	0.3887	0.3946	0.3707	0.3826
DPS	β_l	0.9601	0.5256	0.5864	0.9704	0.2848	0.3927
	β_k	0.3040	0.0303	0.0306	0.2998	0.0349	0.0349
LP	β_l	0.7003	0.0281	0.0281	0.7005	0.0286	0.0286
	β_k	0.2937	0.0480	0.0484	0.3929	0.0607	0.1110
LPIV	β_l	0.7003	0.0281	0.0281	0.7005	0.0286	0.0286
	β_k	0.3010	0.0486	0.0486	0.3015	0.0490	0.0490

See note in Table 2.

Table 6: Chile data estimates

	Food products (311)				Textiles (321)				Wood products (331)				Metal (381)			
	Capital	SE	Labor	SE	Capital	SE	Labor	SE	Capital	SE	Labor	SE	Capital	SE	Labor	SE
OLS	0.271	0.010	1.142	0.018	0.208	0.018	1.004	0.029	0.219	0.021	1.062	0.037	0.186	0.015	1.224	0.026
FE	-0.529	0.030	0.936	0.034	-0.215	0.054	0.841	0.059	-0.125	0.065	0.862	0.058	-0.145	0.047	0.802	0.053
LP - M	0.170	0.087	0.672	0.029	0.236	0.134	0.721	0.053	0.214	0.131	0.685	0.051	0.414	0.131	0.835	0.056
LP - E	0.063	0.109	0.779	0.037	0.329	0.725	0.945	0.689	0.232	0.108	0.840	0.063	0.427	0.120	0.995	0.057
LP - F	0.420	0.171	1.030	0.035	0.204	0.133	0.968	0.082	0.423	0.145	1.081	0.065	0.334	0.112	1.066	0.048
LPIV - M	0.907	0.084	0.672	0.018	0.941	0.091	0.721	0.035	0.697	0.180	0.685	0.037	0.948	0.205	0.835	0.032
LPIV - E	0.909	0.042	0.779	0.018	0.812	0.164	0.945	0.036	0.912	0.183	0.840	0.044	0.953	0.130	0.995	0.032
LPIV - F	0.901	0.030	1.030	0.022	0.962	0.199	0.968	0.043	0.905	0.181	1.081	0.053	0.971	0.298	1.066	0.036
ACFIV - M	0.759	0.173	0.733	0.215	0.370	0.263	1.016	0.331	0.375	0.211	1.136	0.206	0.349	0.299	0.622	0.291
ACFIV - E	0.621	0.155	0.883	0.300	0.780	0.306	0.534	0.258	0.593	0.259	0.952	0.209	0.819	0.260	0.509	0.255
ACFIV - F	0.717	0.256	0.601	0.255	0.141	0.304	0.973	0.283	0.409	0.206	0.831	0.197	0.507	0.227	1.170	0.223

Note: This table contains estimates of value-added production functions for four manufacturing industries in Chile over the period 1979-1986. For LP and the proposed estimation procedure (LPIV and ACFIV), I estimate different models using materials (M), electricity (E) and fuel (F) as proxy variables. Construction of labor and capital variables follows LP, but with the labor input as the sum of skilled and unskilled labor, similar to Ackerberg, Caves and Frazer (2006). Unlike in LP and Ackerberg, Caves and Frazer (2006), I do not estimate separate input demand functions for the intermediate good for the three structural breaks mentioned in LP. Standard errors are computed using 100 bootstrapped draws where sampling is over firms.