

# Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context

Daniel A. Akerberg\*

February 24, 2019

## Abstract

Much of the recent empirical work estimating production functions has used methodologies proposed in two distinct lines of literature: 1) the literature started by Olley and Pakes (1996) on "proxy variable" techniques, and 2) what is commonly referred to as the "dynamic panel" literature. We illustrate how timing and firm information set assumptions are key to both methodologies, and how these assumptions can be strengthened or weakened almost continuously. We also discuss other assumptions that have utilized in these literatures to increase the efficiency (or precision) of estimates. Empirically, we then examine how, in a number of plant level production datasets, strengthening or weakening the timing/information set assumptions affects the precision of estimates. We compare these impacts on precision to those achieved by imposing other potential assumptions. This gives the researcher a better idea of the efficiency tradeoffs between different possible assumptions, at least in the production function context.

## 1 Introduction

There is a large and active empirical literature that estimates production functions relating outputs to inputs of firms. It has long been recognized that there are significant econometric hurdles here. In particular there is the issue that observed inputs will likely be endogenously chosen as a function of unobservable components of production, generating endogeneity issues. Classic solutions to these endogeneity problems, e.g. fixed effects approaches or instrumental variables based on observing exogenous shifters of inputs (e.g. observed input prices), have not always been successful at addressing these issues.

---

\*Dept. of Economics, University of Texas at Austin; Comments by seminar participants at Columbia, Northwestern, Boston College, Virginia, UT-Austin, Wisconsin, Western, Rice, MIT/Harvard, and conferences at Penn State, Montreal-HEC, and Pakes-Fest in Vail are much appreciated.

Given this, researchers have turned to other methodologies to estimate production functions accounting for endogeneity. In particular, two methods have seen extensive use in the empirical literature. First, a literature starting with Olley and Pakes (1996) and Levinsohn and Petrin (2003) directly addressing production function estimation has seen extensive use - as of the date of this paper these two papers have more than 10000 citations between them. We call these methodologies "proxy variable" approaches.<sup>1</sup> Second, a broader set of methodologies, building from the panel data literature, has been applied to production function estimation. Blundell and Bond (2000), which deals specifically with production function estimation, is perhaps the most well known paper detailing this approach as applied to production functions - it has more than 2000 citations. We describe these methodologies as "dynamic panel" approaches (though this is somewhat of a misnomer, see Section 2.3).

While obviously quite popular, as might be expected from methodologies that address endogeneity problems *without* directly observing exogenous variation (as in standard instrumental variables methodology), applying these methods can be challenging. First, these methodologies need to make what are arguably strong assumptions on firm behavior. Second, there can be issues with precision of estimates - often, obtaining precise estimates requires enforcing stronger assumptions than one might have liked. On the other hand, there are many things to like about these methods - for example, they are both semi-parametric in the sense that structural production function parameters can be estimated without fully specifying large parts of firms decision making problems. Not only can this help avoid misspecification, but since these decision making problems are often thought to be dynamic, they have computational advantages.

In this paper, we discuss the similarities and differences of the various assumptions used in these two literatures. We pay particular attention to *timing* and *information set* assumptions that we argue are key components of *both* of these methodological approaches. We argue that one nice aspect of these timing/information set assumptions is that they are not 0-1 assumptions - instead, they are quite flexible in that they can be strengthened or weakened almost continuously (up to the discreteness of one's data). Moreover, with data being observed at higher and higher frequency, the flexibility one has in terms of making these assumptions is increasing. For example, if a researchers is willing to make the assumption that there is a year lag between when a decision is made and a variable is observed in a dataset with annual observations, perhaps they would be willing to assume a 12 month lag if they were instead given a dataset with monthly observations.

We first propose a setup that accommodates such flexibility in timing and information set assumptions. Then, as an empirical illustration, we investigate the effects of strengthening or weakening these assumptions on the precision of estimates of returns-to-scale in two commonly used production function datasets. While theory tells us that adding additional (correct) moment conditions based on additional (correct) model assumptions will produce a weakly more efficient estimator, it doesn't tell us *how much more* efficient it will be. Our empirical work, while only

---

<sup>1</sup>These methods are also sometimes described as "control function" approaches.

an example, helps illustrate this, i.e. how much more precision such stronger timing/information set assumptions can generate in a realistic small sample context. As a interesting benchmark to compare this increased precision too, we also consider the effects of adding an alternative auxiliary assumption that has sometimes been made in this literature. In particular, in the dynamic panel literature, an additional stationarity assumption has often been imposed on the model - this additional stationarity assumption also adds extra moment conditions to the model so also tends to increase the precision of the resulting estimators. We compare these precision gains, i.e. we compare, in a real world data setting, the increased precision afforded by one possible assumption, versus the increased precision afforded by an alternative assumption. In some cases, the increased precision one obtains by strengthening ones timing/information set assumption by one period is similar to the increased precision one obtains by adding the stationarity assumption.

While we are only making this comparison for a few production datasets, our hope is that production technology is similar enough across industries/locations that our results can at least provide some guidance on what combinations of assumptions what might be worth investigating in other situations. More generally, we hope this paper illustrates that there are different dimensions in which one can add (or subtract) assumptions in these models, and one should consider which might be the most appropriate or credible in the particular data one has. We also note that these timing and information set assumptions have also seen recent application in other literatures, e.g. demand estimation (e.g. Sweeting (2009), Grennan (2013), Lee (2013), Sullivan (2017)) so our results could be interesting in that context as well.

## **2 Proxy Variable and Dynamic Panel Approaches**

We start by outlining the proxy variable and dynamic panel approaches. We focus on the timing and information set assumptions that we argue are common to the two approaches - illustrating how these assumptions can be weakened or strengthened almost continuously. We also examine other assumptions or choices amongst assumptions in the methods - highlighting other similarities and differences across the approaches.

### **2.1 Proxy Variable Approaches**

We first describe what are often called proxy variable (or control function) approaches, as introduced by Olley and Pakes (1996 - henceforth OP). This has been an active area of recent methodological study - including work by Levinsohn and Petrin (2003 - henceforth LP), Akerberg, Caves, and Frazer (2015 - henceforth ACF), Van Biesebroeck (2007), Wooldridge (2009), De Loecker (2011), Doraszelski and Jaumandreu (2013), Gandhi, Navarro and Rivers (2017 - henceforth GNR), Kim,

Petrin, and Song (2016), Grieco, Li and Zhang (2016) and Collard-Wexler and De Loecker (2016).<sup>2</sup> Consider a simple Cobb-Douglas production function in logs

$$(1) \quad y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \epsilon_{it}$$

where  $y_{it}$  is the log of output,  $k_{it}$  is the log of capital input, and  $l_{it}$  is the log of labor input, all of which are observed by the econometrician. There are two econometric unobservables,  $\omega_{it}$  and  $\epsilon_{it}$ . The  $\epsilon_{it}$  represent shocks to production or productivity that are not observable (or predictable) by firms before making their input decisions at  $t$ . These could be, for example, i.i.d. true shocks to output, or serially correlated measurement error in output. In contrast, the  $\omega_{it}$  represent "productivity" shocks that *are* potentially observed or (partially) predictable by firms when they make input decisions.  $\omega_{it}$  could represent variables such as firm managerial ability of a firm or some other unobserved (exogenously evolving) input of production. Note that given these assumptions,  $\omega_{it}$  (and not  $\epsilon_{it}$ ) is the unobservable that is problematic in terms of econometric endogeneity - given  $\omega_{it}$  are known or partially known when inputs are chosen, those inputs will generally be correlated with  $\omega_{it}$ .

In the canonical proxy variable model,  $\omega_{it}$  is assumed to evolve according to an exogenous first order markov process, i.e.

$$p(\omega_{it} | \{\omega_{i\tau}\}_{\tau=0}^{t-1}) = p(\omega_{it} | \omega_{it-1})$$

where the firm knows the distribution  $p$ . Typically, firms are assumed to observe  $\omega_{it}$  at  $t$ , but have no information on future  $\omega$ 's (other than their conditional distribution  $p$ ). Hence

$$p(\omega_{it} | I_{it-1}) = p(\omega_{it} | \omega_{it-1})$$

where  $I_{it-1}$  is the firms information set at  $t$ . Assumptions are also made on the points in time in which the inputs  $k$  and  $l$  are chosen by the firms. For example, it is often assumed that  $k_{it}$  is chosen by the firm at time  $t-1$ , while  $l_{it}$  is chosen at time  $t$ . This reflects an economic assumption about how far in advance input decisions need to be made, and in this case represent a "time-to-build" assumption on capital of one period. Given this assumption, Akerberg, Benkard, Berry, and Pakes (2007) describe capital as a "fixed" input, and labor as a "variable" input.

Estimation in the proxy variable literature typically proceeds in two stages. As discussed in ACF, the first stage can be thought of as a way to identify  $\epsilon_{it}$ , i.e. one of the two econometric unobservables. Doing this typically relies on an additional assumption that some observed decision variable of the firm at  $t$  (e.g. capital investment in OP, some intermediate input in LP) is strictly monotonic (or weakly monotonic) in only one unobservable,  $\omega_{it}$ . This allows one to invert that decision in the unobservable  $\omega_{it}$  and thus express that unobservable as a function of data and

---

<sup>2</sup>Note that while these different papers and methodologies for the most part share the timing and information set assumptions focused on in this paper, they often differ in terms of other assumptions. The impact of these other assumptions on precision might also be interesting to investigate.

parameters. Intuitively, given the assumption that  $\epsilon_{it}$  is not known to (or irrelevant for) the firm,  $\epsilon_{it}$  should not enter these decisions. Of course, as detailed in prior work and as discussed further later in this paper, the assumption that this decision variable only depends on a single unobservable rules out other unobservables that might affect these observed decision variables, e.g. unobserved input prices.

In the current paper we do not focus on this first stage "inversion". Hence we proceed under the assumption that  $\epsilon_{it}$  has been identified and netted out of the production function (1). In other words, we consider the simpler model *without*  $\epsilon_{it}$ , i.e.

$$(2) \quad y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it}$$

Of course, one could also get to this point by just starting with a model with no  $\epsilon_{it}$ , in which case the assumptions required for the first stage inversion would not be necessary (in fact, Akerberg and Hahn (2015) consider a non-parametric version of this model).<sup>3</sup>

Estimation of (2), i.e. the production function net of  $\epsilon_{it}$ , relies on moment conditions resulting directly from the information set and timing assumptions described above. Specifically, because of the first order markov assumption and information set assumptions on  $\omega_{it}$ , we can decompose  $\omega_{it}$  into its conditional expectation given  $I_{it-1}$  and an innovation term

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

where by construction, the innovation term  $\xi_{it}$  satisfies

$$E[\xi_{it} | I_{it-1}] = 0$$

Substituting this decomposition into the production function, we obtain

$$(3) \quad y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + g(\omega_{it-1}) + \xi_{it}$$

$$(4) \quad = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + g(y_{it-1} - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1}) + \xi_{it}$$

and since we can write the innovation term as a function of data and parameters (where the parameters include the  $g$  function), i.e.

$$\xi_{it} = y_{it} - \beta_0 - \beta_k k_{it} - \beta_l l_{it} - g(y_{it-1} - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1})$$

---

<sup>3</sup>Another possibility would be to make the additional assumption that the process governing  $\omega_{it}$  is linear. This allows one to proceed without eliminating the  $\epsilon_{it}$  and thus without the assumptions necessary for the first stage inversion. This is essentially what the dynamic panel literature described below does.

is straightforward to construct moment conditions that can be used for estimation, i.e.

$$E[y_{it} - \beta_0 - \beta_k k_{it} - \beta_l l_{it} - g(y_{it-1} - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1}) | I_{it-1}] = 0$$

Note that under the specific timing assumptions made earlier,  $k_{it}, l_{it-1}, y_{it-1} \in I_{it-1}$ . Further lags of these variables are also in  $I_{it-1}$ , though leads of these variables (in particular,  $l_{it}$ ) are not. Olley and Pakes (1996) and Pakes and Olley (1995) show that the  $g$  function can be treated non-parametrically, though if one assumes that  $\omega_{it}$  follows an AR(1) process, i.e.  $g(\omega_{it-1}) = \rho\omega_{it-1}$ , one could use the follow simple set of moments that would tend to exactly identify the parameters:

$$(5) \quad E \left[ \begin{array}{c} (y_{it} - \beta_0 - \beta_k k_{it} - \beta_l l_{it} - \rho(y_{it-1} - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1})) \cdot \begin{pmatrix} k_{it} \\ l_{it-1} \\ y_{it-1} - \beta_0 - \beta_k k_{it-1} - \beta_l l_{it-1} \end{pmatrix} \\ y_{it} - \beta_0 - \beta_k k_{it} - \beta_l l_{it} \end{array} \right] = 0$$

where the last moment enforces a normalization  $E[\omega_{it}] = 0$  to identify the constant term  $\beta_0$ . Obviously, this is just one possible set of unconditional moments that are generated by the above conditional moment.

While moments similar to this have been used many times in the literature, exact conditions under which this system identifies the parameters have not been formally derived. This is probably because this procedure implicitly treats large parts of the model non-parametrically (which is a nice attribute of the methodology). In particular, the above model does not fully specify the process by which the firm chooses their inputs - it really only makes timing assumptions about when those inputs are chosen, and what is in firms' information sets when those inputs are chosen (plus the assumptions necessary for the first stage). It is hard to even enumerate the large set of models consistent with these assumptions, let alone show which subset of these models would be identified with this moment condition. That said, there is work that examines various aspects of this issue, e.g. Bond and Söderbom (2005), ACF (2015), and GNR (2017). Some simple intuition on this issue that is apparent from examination of (5) is that identification  $\beta_k$  and  $\beta_l$  will be aided by exogenous variation in  $k_{it}$  and  $l_{it}$  conditional on  $\omega_{it-1} = y_{it-1} - \beta_0 + \beta_k k_{it-1} + \beta_l l_{it-1}$ . In other words, one should think about why two firms with the same lagged productivity shock would have different values of  $k_{it}$  and  $l_{it}$ , and why this variation is exogenous. The exogeneity condition is important, for example, because while in this model  $l_{it}$  will certainly vary conditional on  $\omega_{it-1}$  (since unlike  $k_{it-1}$ ,  $l_{it}$  can be chosen after the firm realizes  $\omega_{it}$ ) the part of that variation due to  $\xi_{it}$ , i.e. the innovation component of  $\omega_{it}$ , is obviously not orthogonal to  $\xi_{it}$ . Some lessons from the aforementioned literature is that aspects of the model such as dynamics (e.g. adjustment costs) in  $k_{it}$  and/or  $l_{it}$  or serially correlated (not necessarily observed by the econometrician) input prices can aid in generating this sort of variation.

## 2.2 Strengthening or Weakening Timing/Information Set Assumptions in Proxy Variable Approaches

The focus of this paper is on the timing and information set assumptions that are fundamental to these proxy variable approaches. In thinking about these assumptions, it is first important to emphasize that the orthogonality conditions arising in these approaches require *both* assumptions on both the time at which inputs are chosen *and* assumptions about what is in firms' information sets at those points in time. To illustrate this, consider a model with just capital, i.e.

$$(6) \quad y_{it} = \beta_0 + \beta_k k_{it} + \omega_{it}$$

where  $\omega_{it}$  again follows a simple AR(1) process  $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$  (with  $E[\xi_{it} | I_{it-1}] = 0$ ). Under 1) the *timing* assumption that  $k_{it}$  is decided at  $t - 1$  and 2) the information set assumption that  $I_{it-1}$  that only contains  $\omega_{it-1}$  (and lags) but not  $\omega_{it}$ ; one obtains that while  $k_{it}$  is *not* orthogonal to the entire  $\omega_{it}$ , it is orthogonal to *part* of  $\omega_{it}$ , i.e. the innovation component of  $\xi_{it}$ . Of course, this orthogonality relies on both assumptions - for example, even if firms decided  $k_{it}$  at  $t - 1$ , if the firm observed  $\omega_{it}$  a period ahead (i.e. observes  $\omega_{it}$  at  $t - 1$ ), this orthogonality condition would not hold. For that matter, even if firms chose capital levels 10 periods in advance, as long as firms also observed  $\omega_{it}$ 's 10 periods ahead, the orthogonality condition would again not hold.

Clearly, what is crucial here is the *difference* between the point in time at which  $k_{it}$  is chosen and the point in time at which  $\omega_{it}$  enters the firms information set. In particular, at least with respect to the moments we are considering to identify the parameters, a model in which  $k_{it}$  is chosen five periods in advance and  $\omega_{it}$  is observed four periods in advance is equivalent to a model in which  $k_{it}$  is chosen one period in advance and  $\omega_{it}$  is observed at  $t$ . We will denote this difference as  $\Delta$ , and given a  $\Delta$  describe our timing/information set assumption as: *At the point in time when the firm chooses  $k_{it}$ , the firms information set includes  $\{\omega_{i\tau}\}_{\tau=0}^{t+\Delta}$  but does not include  $\{\omega_{i\tau}\}_{\tau=t+\Delta+1}^{\infty}$ .* In the model above where  $k_{it}$  is the only input, is chosen at  $t - 1$  and  $\omega_{it}$  is observed at  $t$ ,  $\Delta = -1$ . For a given  $\Delta$ , the model implies the following moment conditions:

$$(7) \quad E[\xi_{it} | \{k_{i\tau}\}_{\tau=0}^{t-\Delta-1}] = 0$$

In words, the innovation component of  $\omega_{it}$  is orthogonal to inputs that were chosen prior to the firm observing  $\omega_{it}$ .

Given this setup, we can easily think about strengthening or weakening the timing/information set assumption. First, we can strengthen the assumption by decreasing  $\Delta$ . This corresponds to either assuming that  $k_{it}$  is chosen at points in time further in the past, or assuming that  $\omega_{it}$  is not observed until later points in time. Examination of (7) indicates why decreasing  $\Delta$  corresponds to a *strengthening* of the timing/information set assumptions - as  $\Delta$  decreases, the period  $t$  innovation is orthogonal to a larger set of  $k$ 's. Econometrically, this strengthened assumption would lead

to a more efficient estimator if all the conditional moments were used<sup>4</sup>. Perhaps a more intuitive way of illustrating this is through recursive substitution of the AR(1) process into the production function (6), i.e.

$$\begin{aligned}
y_{it} &= \beta_0 + \beta_k k_{it} + \omega_{it} \\
&= \beta_0 + \beta_k k_{it} + \rho \omega_{it-1} + \xi_{it} \\
&= \beta_0 + \beta_k k_{it} + \rho^2 \omega_{it-2} + \rho \xi_{it-1} + \xi_{it} \\
&= \beta_0 + \beta_k k_{it} + \rho^3 \omega_{it-3} + \rho^2 \xi_{it-2} + \rho \xi_{it-1} + \xi_{it} \\
&= \beta_0 + \beta_k k_{it} + \rho^4 \omega_{it-4} + \rho^3 \xi_{it-3} + \rho^2 \xi_{it-2} + \rho \xi_{it-1} + \xi_{it} \\
&\cdot \\
&\cdot
\end{aligned}$$

For a given assumed  $\Delta$ , consider the  $-\Delta$ th iterate of the recursive substitution, i.e.

$$y_{it} = \beta_0 + \beta_k k_{it} + \rho^{-\Delta} \omega_{it+\Delta} + \sum_{\tau=1}^{-\Delta} \rho^{\tau-1} \xi_{it-\tau}$$

Under the timing/information set assumption  $\Delta$ ,  $k_{it}$  is correlated with the first component of the residual ( $\rho^{-\Delta} \omega_{it+\Delta}$ ) but orthogonal to the second component of the residual ( $\sum_{\tau=1}^{-\Delta} \rho^{\tau-1} \xi_{it-\tau}$ ). As  $\Delta$  decreases, the first component of the residual becomes smaller in terms of variance, while the second component of the residual becomes bigger. In other words, as  $\Delta$  decreases,  $k_{it}$  is assumed to be orthogonal to "more" of the residual, i.e. in essence a stronger assumption. Note that as  $\Delta \rightarrow -\infty$  (e.g. if  $k_{it}$  is chosen without knowledge of any of the  $\omega$ 's), then we are assuming that  $k_{it}$  is orthogonal to the entire residual, and, e.g., OLS would produce consistent estimates of the production function<sup>5</sup>.

In summary, strengthening the timing/information set assumption from  $\Delta = -1$  to  $-\infty$  corresponds to assuming that  $k_{it}$  is orthogonal to a bigger proportion of the residual, and this will generally generate more efficient estimates. We can also weaken the timing/information set assumption by increasing  $\Delta$ . For example,  $\Delta = 3$  coincides with a model where  $k_{it}$  is chosen at  $t$  and  $\omega_{it}$  is observed three periods ahead at  $t - 3$ . Interestingly, the implications of increasing  $\Delta$  above  $-1$  are fundamentally different than decreasing it. This is because as soon as  $\Delta = 0$ ,  $k_{it}$

---

<sup>4</sup>In practice this comparison is harder to make, because papers using these methodologies vary considerably in terms of 1) the number of conditional moments actually used for estimation (e.g. some enforce  $\xi_{it}$  being orthogonal to only the most recent valid instrument (e.g.  $k_{it-\Delta-1}$  in (7)), and 2) how conditional moments are converted to unconditional moments for estimation.

<sup>5</sup>In relation to the discussion in the prior footnote, note that under some additional assumptions, e.g. heteroskedasticity, OLS provides efficient estimates of a linear model, so to obtain full efficiency one does not need to enforce all the implied conditional moments of the model in (7) - all one needs to use is that that  $\omega_{it}$  is mean zero and uncorrelated with  $k_{it}$ .

is chosen with full knowledge of  $\omega_{it}$  and thus is generally correlated with all components of  $\omega_{it}$ . Hence the proxy variable literature (as well as the dynamic panel literature, see the next section) uses lags of  $k_{it}$  that are orthogonal to components of  $\omega_{it}$ , as shown in (7). For example, when  $\Delta = 0$ ,  $k_{it-1}$  is orthogonal to the innovation component of  $\omega_{it}$ , i.e.  $\xi_{it}$ . This distinction between  $\Delta \leq -1$  versus  $\Delta \geq 0$  is illustrated in Akerberg and Hahn (2015), who using a very simply method of proof based on verifying conditions in Matzkin (2004) and Newey and Imbens (2009), show conditions under which a non-parametric version of (6), i.e.

$$y_{it} = f(k_{it}, \omega_{it})$$

is identified. However, this simple method of proof only applies when  $\Delta \leq -1$  (though this doesn't imply that the model is not identified when  $\Delta \geq 0$ ).

Returning to the canonical two input production function of Section 2.1, one can see that in the proxy variable literature, one can potentially choose a different  $\Delta$  for each input. The moments (5) (as well as the empirical work and monte-carlo experiments in OP, LP, and ACF) reflect the assumption that for capital,  $\Delta = -1$ , while for labor,  $\Delta = 0$ . With more inputs, it is straightforward to set up additional moments given whatever assumption on  $\Delta$  one wants to make for each of the inputs. In the empirical section of this paper, we examine what happens as we vary the strength of the timing assumptions  $\Delta$  - in particular, we compare the precision of estimates given various choices of  $\Delta$ . Thinking precisely about exactly how strong one wants their timing assumptions to be seems particularly relevant given newer high frequency datasets. For example, if one is willing to make the assumption that  $\Delta = -1$  in a dataset with annual observations, presumably one should hypothetically be willing to make the assumption that  $\Delta = -4$  were the same dataset to instead have quarterly observations, or  $\Delta = -12$  in the analogous dataset with monthly observations

## 2.3 Dynamic Panel Approaches

A second approach that has been used in many recent empirical production function studies is based on the "dynamic panel" literature. This approach stems from a long line of research on panel data models, e.g. Chamberlain (1982), Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998). While this is a very general methodology, Blundell and Bond (2000) do a good job of illustrating how it can be applied to production function estimation. An interesting observation is that much of this literature is based on models with fixed effects and state dependence (i.e. the lagged dependent variable has a causal effect on the current dependent variable). While state dependence is typically *not included* in production function models (and we do not consider models with such state dependence in this paper), the methodology can alternatively be applied to models with no state dependence but a serially correlated unobservable (in addition to a fixed effect). This is what we study here, as

the model is relatively comparable to the canonical models described above in the proxy variable literature.

More specifically, Blundell and Bond (2000) consider the following production function

$$(8) \quad y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \alpha_i + \omega_{it} + \epsilon_{it}$$

$\alpha_i$  is a firm fixed effect that for now we will allow to be arbitrarily correlated with values of the inputs in all periods. To keep things simple for now, assume  $\epsilon_{it}$  is measurement error in output that is orthogonal to inputs choices in all periods (we discuss how this assumption can be relaxed later).

The time varying productivity shock  $\omega_{it}$  is assumed to follow an AR(1) process  $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$ . Inputs are classified according to their potential correlation with  $\omega_{it}$  and its components, e.g.  $\xi_{it}$ . A period  $t$  input is described as *exogenous* if it is orthogonal to  $\omega_{it}$ . It is described as *predetermined* if it is orthogonal to  $\xi_{it}$ , the innovation in  $\omega_{it}$ . Lastly, it is described as *endogenous* if it is potentially correlated with  $\xi_{it}$ , but orthogonal to future period innovations in  $\omega$ , e.g.  $\xi_{it}$ . This classification can be interpreted in terms of the timing and information set assumptions discussed earlier, i.e. assumptions regarding which  $\xi_{it}$ 's the firm observes when making their period  $t$  input choices. Specifically, the assumption that an input is *exogenous* corresponds to  $\Delta = -\infty$ , i.e. where the firm observes no  $\xi_{it}$ 's when choosing the input; the assumption that an input is *predetermined* corresponds to  $\Delta = -1$ , i.e. where the firm observes past  $\xi_{it}$ 's but not the current  $\xi_{it}$  when choosing the input; and the assumption that an input is *endogenous* corresponds to  $\Delta = 0$ , where the firm observes current  $\xi_{it}$  but not future  $\xi_{it}$ 's when choosing the input. As we will see momentarily, it is also simple to consider  $\Delta$ 's other than 0, -1, and  $-\infty$  in the dynamic panel framework.

Estimating (8) typically involves double differencing - one difference is to isolate the innovation in  $\omega_{it}$  - the other difference is to eliminate the fixed effect  $\alpha_i$ . The initial difference is a  $\rho$  first difference, i.e.

$$\begin{aligned} y_{it} - \rho y_{it} &= (1 - \rho) \beta_0 + \beta_k (k_{it} - \rho k_{it-1}) + \beta_l (l_{it} - \rho l_{it-1}) + (1 - \rho) \alpha_i + (\omega_{it} - \rho \omega_{it-1}) + (\epsilon_{it} - \rho \epsilon_{it-1}) \\ &= (1 - \rho) \beta_0 + \beta_k (k_{it} - \rho k_{it-1}) + \beta_l (l_{it} - \rho l_{it-1}) + (1 - \rho) \alpha_i + \xi_{it} + (\epsilon_{it} - \rho \epsilon_{it-1}) \end{aligned} \quad (9)$$

and the latter difference is a straight first difference, generating

$$\begin{aligned} (10) \quad & (y_{it} - \rho y_{it-1}) - (y_{it-1} - \rho y_{it-2}) \\ &= \beta_k [(k_{it} - \rho k_{it-1}) - (k_{it-1} - \rho k_{it-2})] + \beta_l [(l_{it} - \rho l_{it-1}) - (l_{it-1} - \rho l_{it-2})] \\ & \quad + \xi_{it} - \xi_{it-1} + (\epsilon_{it} - \rho \epsilon_{it-1}) - (\epsilon_{it-1} - \rho \epsilon_{it-2}) \end{aligned}$$

The residual in (10) only contains  $\xi$ 's and  $\epsilon$ 's, and given the above timing and information set assumptions, will be orthogonal to lags (or leads) of  $k$  and  $l$ . For example, if  $\Delta = -1$  for both  $k$  and  $l$ , the moment conditions

$$(11) \quad E[\xi_{it} - \xi_{it-1} + (\epsilon_{it} - \rho\epsilon_{it-1}) - (\epsilon_{it-1} - \rho\epsilon_{it-2}) \mid k_{it-1}, l_{it-1}] = 0$$

hold, since the timing/information set assumption implies  $k_{it-1}$  and  $l_{it-1}$  are chosen before  $\xi_{it-1}$  (and  $\xi_{it}$  are in the firms information set). For  $\Delta = 0$ , this moment condition would not hold - it would only hold for  $k$  and  $l$  lagged to period  $t - 2$  and prior.

While the dynamic panel literature seems to focus on the cases where  $\Delta = -\infty, -1$ , and  $0$ , it is simple to see how this methodology can also be applied with alternative timing/information set assumptions.<sup>6</sup> As one weakens the timing assumptions by increasing  $\Delta$  from  $0$ , (11) holds for only further lagged  $k$  and  $l$ . As one strengthens the timing assumptions, (11) will hold for more and more  $k$ 's and  $l$ 's into the future.

Following the logic of what was done in the prior section, there is another way to think of strengthening the timing information set assumptions. If, for example,  $\Delta = -2$  (for both  $k$  and  $l$ ), for the initial difference (9), we can do a  $\rho^2$  *two period difference* instead of a  $\rho$  one period difference, i.e.

$$(12) \quad y_{it} - \rho^2 y_{it-2} = (1 - \rho^2) \beta_0 + \beta_k (k_{it} - \rho^2 k_{it-2}) + \beta_l (l_{it} - \rho^2 l_{it-2})$$

$$(13) \quad \quad \quad + (1 - \rho^2) \alpha_i + (\omega_{it} - \rho^2 \omega_{it-2}) + (\epsilon_{it} - \rho^2 \epsilon_{it-2})$$

$$= (1 - \rho^2) \beta_0 + \beta_k (k_{it} - \rho^2 k_{it-2}) + \beta_l (l_{it} - \rho^2 l_{it-2})$$

$$(14) \quad \quad \quad + (1 - \rho^2) \alpha_i + \rho \xi_{it-1} + \xi_{it} + (\epsilon_{it} - \rho^2 \epsilon_{it-2})$$

Then doing a straight first difference to eliminate the  $\alpha_i$  term, we get

$$(15) \quad (y_{it} - \rho^2 y_{it-2}) - (y_{it-1} - \rho^2 y_{it-3}) =$$

$$= \beta_k [(k_{it} - \rho^2 k_{it-2}) - (k_{it-1} - \rho^2 k_{it-3})] + \beta_l [(l_{it} - \rho^2 l_{it-2}) - (l_{it-1} - \rho^2 l_{it-3})]$$

$$+ \rho \xi_{it-1} + \xi_{it} - (\rho \xi_{it-2} + \xi_{it-1}) + (\epsilon_{it} - \rho^2 \epsilon_{it-2}) - (\epsilon_{it-1} - \rho^2 \epsilon_{it-3})$$

---

<sup>6</sup>The popular Stata command *xtabond2* can accomodate  $\Delta = -\infty, -1$ , and  $0$ , as well as weaker timing assumptions where  $\Delta > 0$ . It does not appear to be able to be used for  $-\infty < \Delta < -1$  (of course, these can be straightforwardly programmed in Mata or other programming languages). It is also somewhat challenging to enforce the restrictions necessary to consider a model with an AR(1) error term instead of a lagged dependent variable. (see Soderbom (2015)).

$$(16) \quad = \beta_k [(k_{it} - \rho^2 k_{it-2}) - (k_{it-1} - \rho^2 k_{it-3})] + \beta_l [(l_{it} - \rho^2 l_{it-2}) - (l_{it-1} - \rho^2 l_{it-3})] \\ + \xi_{it} + (\rho - 1) \xi_{it-1} - \rho \xi_{it-2} + (\epsilon_{it} - \rho^2 \epsilon_{it-2}) - (\epsilon_{it-1} - \rho^2 \epsilon_{it-3})$$

Given the timing/information set assumption of  $\Delta = -2$ , the error term in (16) will be orthogonal to  $k_{it-1}$ ,  $l_{it-1}$  and further lags. This can straightforwardly be extended to stronger timing/information set assumptions - the first step being to do a  $\rho^{-\Delta}$ ,  $-\Delta$ -period difference, the second step being to do a straight first difference to eliminate the  $\alpha_i$  term.

### 2.3.1 Additional Stationarity Assumptions

As noted by Blundell and Bond (1998, 2000), the "double differences" typically done when the dynamic panel literature is applied to production functions appear to be fairly demanding on the data - specifically, many empirical researchers have found the procedures generate large standard errors, even, for example when  $\Delta = -1$  (we do not know of an empirical study that has used this methodology with  $\Delta < -1$ ). Of course, this is not a criticism of the procedures - it is a result of the fact that they are making relatively weak assumptions - i.e. allowing extensive correlation between the unobservable and the explanatory variables. Blundell and Bond (2000) and Bond and Soderbom (2005) do a good job illustrating what aspects of the data generating process tend to help or hinder this imprecision. To mitigate this issue, Blundell and Bond suggest adding some additional assumptions to the model to sharpen predictions - these additional assumptions are based on stationarity restrictions that were also proposed in Arellano and Bover (1995).

In the production function context, these additional assumptions regard how the fixed effects  $\alpha_i$  relate to the inputs. In the prior section, the  $\alpha_i$  were permitted to be arbitrarily correlated with the inputs. Blundell and Bond (1998, 2000) suggest the additional assumption that the  $\alpha_i$  are orthogonal to *specific functions* of the inputs. In particular, they assume that

$$(17) \quad E[\alpha_i | k_{it} - k_{it-1}, l_{it} - l_{it-1}] = 0$$

i.e. that the fixed effects are mean independent of *changes* in the inputs, or relatedly, that they are uncorrelated with changes in the inputs. In other words, while  $\alpha_i$  is permitted to be correlated with levels of inputs, it is assumed to be orthogonal to *changes* in these inputs. Because in these techniques (like the proxy variable techniques), much of the underlying economic model is not specified (e.g. the precise way in which firms make input choices), this is a bit of a high level assumption that can be challenging to interpret. That said, in the production context, this assumption essentially says that inherently more productive firms (i.e. those with high  $\alpha_i$ 's) are not growing faster (or slower) than those who are less productive. One can think of situations where this assumption might not hold - a primary example would be in a relatively new and

growing market, where one might expect better firms to become bigger quicker than worse firms. Relatedly it could be a problem if, in a relatively short panel, industry wide variables are trending a market upwards or downwards. Blundell and Bond (2000) illustrate how in the production function context, (17) can essentially be interpreted as a stationarity assumption on the firm operating environment, i.e. that firms have reached their "long-run" optimal input levels given their  $\alpha_i$ 's and that the variation in the data is around those optimal levels.

Given this stationarity assumption, one can form additional moments based only on the single  $\rho$ -differenced production function, i.e. prior to doing the latter difference to eliminate the  $\alpha_i$ . Specifically, under (17), the residual in

$$(18) \quad y_{it} - \rho y_{it-1} = (1 - \rho) \beta_0 + \beta_k (k_{it} - \rho k_{it-1}) + \beta_l (l_{it} - \rho l_{it-1}) + (1 - \rho) \alpha_i + \xi_{it} + (\epsilon_{it} - \rho \epsilon_{it-1})$$

will be orthogonal to appropriately lagged differences in inputs, e.g. if  $\Delta = -1$  then this residual is orthogonal to differences in inputs  $k_{it} - k_{it-1}, l_{it} - l_{it-1}$  and prior, while if  $\Delta = 0$  then this is orthogonal to differences in inputs  $k_{it-1} - k_{it-2}, l_{it-1} - l_{it-2}$  and prior. Note that if one were to strengthen the timing assumption to  $\Delta < -1$ , then the same logic would apply to form valid moments based on a  $\rho^{-\Delta}$ ,  $-\Delta$ -period differenced equation as discussed above. In Blundell and Bond (2000) and following empirical work, these additional moments appear to have been relatively successful at increasing precision of estimates - many empirical papers now use estimators utilizing these additional moments. Estimators that utilize these additional assumptions and moments are often referred to as a "System GMM" (or SYS-GMM) estimator, in contrast to the "Differenced GMM" (or DIFF-GMM) estimator that uses only moments based on, e.g. (10). While the stationarity assumption required to generate these additional moments might be considered strong, it is of course weaker than the implicit assumption that  $\alpha_i = 0$  in the proxy variable techniques for estimating production functions, though as we discuss in the next section, there are other dimensions in which the proxy variable literature makes weaker assumptions than the dynamic panel literature.

## 2.4 Comparison of Assumptions in Proxy Variable and Dynamic Panel Approaches

The discussion above illustrates how timing/information set assumptions are common to both the proxy variable and dynamic panel literatures, and these assumptions can be strengthened or weakened quite easily, essentially in a continuous fashion (up to the discreteness inherent in the data). We now summarize differences in other assumptions across the two methodologies, starting with assumptions on  $\omega_{it}$ , then proceeding to assumptions on  $\alpha_i$  and  $\epsilon_{it}$  (more discussion on this is in ACF).

Regarding assumptions on  $\omega_{it}$ , the discussion above illustrates that, generally speaking, the dynamic panel literature makes stronger assumptions than the proxy variable literature. While

the proxy variable methodology can straightforwardly allow  $\omega_{it}$  to follow a general (and potentially non-parametrically estimated) first order markov process  $\omega_{it} = g(\omega_{it-1}) + \xi_{it}$ , the dynamic panel literature assumes that  $\omega_{it}$  follows an AR(1) process  $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$ . The issue is that without that linear structure of the markov process, simple linear differencing as in (9) to isolate the innovation term  $\xi_{it}$  is not possible. In contrast, the non-linear structure can be accommodated in the proxy variable literature because of the lack of  $\alpha_i$  and the fact that  $\epsilon_{it}$  is netted out in the first stage. Note that both methodologies can be extended to allow for higher order markov processes - though in the proxy variable approach this requires more data and assumptions regarding the first stage inversion (as discussed in Akerberg, Berry, Benkard, and Pakes (2007)), and for the dynamic panel approach, simple differencing requires the assumption of a linear higher order markov process, e.g.  $\omega_{it} = \rho_1\omega_{it-1} + \rho_2\omega_{it-2} + \xi_{it}$ .

However, while the proxy variable approaches generally makes weaker assumptions on the  $\omega_{it}$  process, they typically make stronger assumptions regarding  $\alpha_i$ . As detailed in the prior section, the dynamic panel literature does allow heterogeneous  $\alpha_i$  across firms. In some cases it makes literally no assumptions on the  $\alpha_i$ 's (when the additional SYS-GMM moments are not used), while in others (when the SYS-GMM moments are used), it makes the stationarity assumptions discussed above.

On the other hand, for the proxy variable literature (for the most part) assumes that  $\alpha_i = 0$ . It is interesting to note that the primary reason for requiring this assumption is the first stage inversion, which as noted above serves to "identify"  $\epsilon_{it}$  and net it out of the production function as in (2). This first stage inversion requires that an observed decision variable of the firm (e.g. investment, choice of an input) depend on only a scalar unobservable, i.e.  $\omega_{it}$ . This scalar unobservable assumption allows one to write  $\omega_{it}$  as a function of observables and parameters, and this is what allows one to identify  $\epsilon_{it}$  in the first stage. The problem with  $\alpha_i$ 's that vary across firms<sup>7</sup> is that generally speaking, this means that any input decision  $i$  will depend on *both*  $\alpha_i$  and  $\omega_{it}$ , hence violating the scalar unobservable assumption and not permitting the first stage inversion. Interestingly, GNR show that in some special cases, one can still do the first stage inversion with  $\alpha_i$ 's in the proxy variable methodology. Specifically, if the observed decision variable that is being used as the proxy variable depends only on the *sum*  $\alpha_i + \omega_{it}$  (and not  $\alpha_i$  and  $\omega_{it}$  individually), then one can still do the first stage inversion. While this condition is not likely to hold for dynamic decision variables like investment (since investment is a long-run decision, it is likely to respond differently to an increase in  $\alpha_i$  vs an increase in  $\omega_{it}$ ), it may often hold for static decision variables, e.g. intermediate inputs.

Lastly, consider assumptions in the two literatures regarding the  $\epsilon_{it}$ 's. As noted above, simply allowing any  $\epsilon_{it}$ 's in the proxy variable approach typically requires a host of auxiliary assumptions - in particular the key scalar unobservable and monotonicity assumptions required to do the

---

<sup>7</sup>Assuming fixed  $T$  - if one assumes  $T \rightarrow \infty$ , allowing  $\alpha_i$ 's that vary across firms is trivial given appropriate assumptions. One can simply estimate the production function separately for each firm.

first stage inversion and net out the  $\epsilon_{it}$ 's. As noted in ACF, for example, this can require strong assumptions limiting unobserved heterogeneity in demand functions. They describe, for example, how in a value added production function with inputs capital ( $\Delta = -1$ ), labor ( $\Delta = 0$ ), and materials ( $\Delta = 0$ ), that 1) using investment demand (e.g. OP) to proxy for unobserved productivity requires assuming away serially correlated unobserved input price shocks to either capital, labor, or materials, that 2) using unconditional materials demand (e.g. LP) to proxy for unobserved productivity requires assuming away serially correlated unobserved input price shocks to labor and materials, and 3) using conditional materials demand (e.g. ACF - conditional on labor input choice) to proxy for unobserved productivity requires assuming away serially correlated unobserved input prices to materials. The proxy variable approaches also place restrictions on the  $\epsilon_{it}$  process itself - essentially it requires  $\epsilon_{it}$  to either be measurement error (potentially serially correlated) in output, or for  $\epsilon_{it}$  to be real shocks to output but that do not impact choice of the proxy input in any period. For the latter to hold will typically require  $\epsilon_{it}$  to be independent over time (otherwise optimal input choices will generally depend on past  $\epsilon_{it}$ ) and also rules out past  $\epsilon_{it}$ 's from impacting future choices of the proxy input, which might, e.g., occur in a context with credit constraints (i.e. current inputs might depend on past shocks to output - see Shenoy (2016) for similar points).

The dynamic panel literature arguably makes weaker assumptions regarding the  $\epsilon_{it}$ 's. Perhaps most importantly, the dynamic panel literature does not require any first stage inversion to net out the  $\epsilon_{it}$ 's. Hence it does not need to make the scalar unobservable and monotonicity assumptions that the proxy variable approach does - this means that dynamic panel methods do not place significant restrictions on unobserved, serially correlated, input prices. This can be important, since as observed by, e.g. Bond and Söderbom (2005) and ACF, serially correlated input prices can be helpful in generating exogenous variation. The precise assumptions that the dynamic panel literature makes on the  $\epsilon_{it}$ 's depends on exactly which moments one uses (e.g. DIFF-GMM vs SYS-GMM) and values of  $\Delta$ . For example, examining (10), the  $\epsilon_{it}$ 's in the unobserved term will need to be orthogonal to the various contemporaneous or lagged inputs used as instruments. Interestingly, note that, depending on  $\Delta$ ,  $\epsilon_{it}$ 's could be permitted to be correlated with future (or sufficiently in the future) choices of inputs, which again might occur in a situation with credit constraints. In summary, at a general level, the dynamic panel literature appears to typically make stronger restrictions on the  $\omega_{it}$  process, while the proxy variable literature tends to make stronger restrictions on the  $\alpha_i$  and  $\epsilon_{it}$ 's.

### 3 Empirical Work

In the prior sections, we argued that timing and information set assumptions play a key role in both the proxy variable and dynamic panel approaches to production function estimation. Moreover, these timing and information set assumptions can be strengthened or weakened in a

very natural, and almost continuous way. The empirical goal of this paper is to examine the relationship between the strengthening or weakening of these assumptions and precision of ones' estimates. Obviously, stronger assumptions will lead to more precise estimates, so for this to be useful, we want to quantify this increased precision in some helpful way. The way we do this is by comparing the increased precision from strengthened timing/information set assumptions to the increased precision from strengthening other assumptions in the production function context. Since, in the dynamic panel literature, many papers utilize the additional SYS-GMM moments (and corresponding stationarity assumptions) proposed by Blundell and Bond (2000), we use this as our benchmark. In other words, we compare the additional precision obtained from strengthening ones timing/information set assumptions vs. the additional precision from moving from the DIFF-GMM moments to the augmented SYS-GMM moments. As noted earlier, one apparent reason for substantial use of the SYS-GMM moments is the lack of precision of estimates generated by DIFF-GMM moments. Our comparisons illustrate how there are alternative ways to increase precision, e.g. strengthening timing assumptions, and our results show, at least in some commonly studied datasets, what the precision tradeoff is between the additional SYS-GMM moments and stronger timing assumptions. We feel that, particularly in higher frequency datasets where, e.g.  $\Delta < -1$  can still be a short lag, strengthening timing assumptions is something that might be considered.

We investigate this issue in two frequently used production datasets from Chile and Mexico. The Chilean dataset, covering plants with at least 10 workers over 1979-1986, was the subject of LP, for example, and the Mexican plant level data, covering 1984-1990, has also been used by many studies, e.g. Asker, Collard-Wexler, and DeLoecker (2015). For each dataset we focus on 3 industry classifications - food products, clothing, and wood products. These were 3 of the industry classifications considered by LP. In both datasets the period of time is one year, so the distinction between, e.g.  $\Delta = -1$  and  $\Delta = 0$  is that the former is a stronger timing assumption by a full year.

To compare apples to apples, our empirical results only utilize dynamic panel methodology.<sup>8</sup> Specifically, in the context of the dynamic panel methodology, we investigate and compare the effects of 1) increasing or decreasing  $\Delta$  and 2) including or not including the additional stationarity moments (i.e. SYS-GMM vs DIFF-GMM). We use the STATA command *xtabond2* to compute our estimates. This has a couple of caveats. First, while it is straightforward to allow the inputs to be incrementally more endogenous using the command (i.e.  $\Delta = -1$  (predetermined), 0 (endogenous), 1 (more endogenous), 2, 3, .....), it does not appear possible to strengthen the timing assumption, i.e. decreasing  $\Delta$  below  $-1$ . Second, it is not straightforward to use *xtabond2* to estimate a dynamic panel model with an AR(1) process with parameter  $\rho$  in the error term

---

<sup>8</sup>One could also assess the impact of stronger or weaker timing assumptions in the context of proxy variable approaches. But in those approaches there is no  $\alpha_i$ , so there is no additional stationarity assumption to add (or remove) and compare the changes in precision to that of changing timing assumptions to. Since we thought this comparison would be interesting, our simple examples utilize dynamic panel approaches.

(i.e.  $\omega_{it}$  from above) - see, e.g. Söderbom (2016). Hence we focus attention on a version of the model where  $\rho = 0$ , and consider  $\Delta = -1, 0, 1$ , and  $2$ .<sup>9</sup> Obviously for all the specifications to produce consistent estimates (of parameters and standard errors), one needs that the strongest assumptions hold in all the datasets, i.e. in the data,  $\Delta \leq -1$  and the additional SYS-GMM stationarity assumptions to hold.

As our goal is to compare the effect of stronger or weaker assumptions on the precision of production function estimators, so our main focus is on the estimated standard errors of our parameter estimates. Being cognizant of the existence of estimation error in the estimates of standard errors (which presumably could be assessed by bootstrapping) and wanting to limit this estimation error, we consider a simple production function with just one input, total labor, and thus one input elasticity coefficient. Given just one input, this input elasticity coefficient can be interpreted as a measure of returns-to-scale.<sup>10</sup> Returns-to-scale is typically an important aspect of a production function because of, e.g. its impacts on market structure.

Table 1 contains our estimates of the returns to scale parameters and their standard errors for the three industries in the two countries. Again, our main focus is on the precision of these estimates, and how they change with the various assumptions. The extent to which a standard error decreases when one adds an additional assumption to the model measures the "information" that imposing that assumption provides on the returns to scale parameter (similar to how a Bayesian posterior on a parameter could change with imposition of an additional assumption). The general patterns of these standard errors across the different estimators are exactly what one would expect. Specifically, as one decreases the strength of the timing assumption from  $\Delta = -1$  to  $\Delta = 2$ , standard errors tend to increase. Similarly, when one moves from the DIFF-GMM specifications to the SYS-GMM specifications that make the additional stationarity assumption, standard errors decrease. Under the weakest sets of assumptions, i.e. high  $\Delta$  and DIFF-GMM, standard errors can be quite high (and estimates quite varied) relative to reasonable ranges of the returns to scale parameter.

To us what is interesting is to compare the magnitude of these standard error changes. In particular, it is interesting to note that starting from a DIFF-GMM specification with  $\Delta = 0$ , one gets quite similar precision gains from a one period stronger timing assumption (i.e.  $\Delta = -1$ ) as one gets from additionally imposing the SYS-GMM stationarity assumptions. For example, in Chilean Industry 312, the production function coefficient with DIFF-GMM and  $\Delta = 0$  has a standard error of 0.157 - moving to  $\Delta = -1$  reduces this to 0.049, while adding the stationarity assumptions reduces it to 0.059 (imposing both the additional assumptions together reduces it to

---

<sup>9</sup>An alternative would be to program the moment conditions ourselves. This would give considerably more flexibility, but also add ambiguity since the many papers in the literature operationalize the various moment conditions implied by the model in different ways. Hence for this simple comparison we thought it best to stick with a comparison using a well-known package that has been used in practice. Note that given  $\rho = 0$ , the inability to decrease  $\Delta$  below -1 is not particularly restrictive, since when  $\Delta = -1$ , current input choice is already orthogonal to the entirety of  $\omega_{it}$ .

<sup>10</sup>Formally, this allows for other inputs to the extent that those other inputs enter in a Leontief way with labor.

0.023). One finds a similar tradeoff in the other production function datasets.

One observation looking at the parameter estimates themselves is that for the strongest timing assumption  $\Delta = -1$ , the SYS-GMM and DIFF-GMM parameter estimates are sometimes very different, with confidence intervals not even close to overlapping. This implies that some assumption must be incorrect in these specifications - for example, the stationarity assumption could be incorrect, or the timing assumption  $\Delta = -1$  could be wrong (or some other assumption could be incorrect). Given this, the standard error estimates for the  $\Delta = -1$  specification might be suspect. But even when one ignores the first column (i.e. starts with the base assumption that  $\Delta = 0$  and the stationarity assumption hold), the more general pattern of precision differences across different timing and stationarity assumptions is similar. Lastly, we again note that these are only examples and of course these quantitative results could differ in different datasets (e.g. other production datasets, non production function datasets, or datasets with different time period lengths). However, we think they are helpful in illustrating the more general point that there are different dimensions on which one can strengthen or weaken assumptions in these methodologies, and one should consider all these in trying to find reasonable and credible specifications.

## 4 Conclusion

We highlight the similarity of two recent literatures on production function estimation in relying on timing and information set assumptions to identify parameters. We show how these assumptions can be weakened or strengthened, and investigate using two production datasets how this changes the precision of the estimates of production function parameters. We compare these precision gains and losses to those from alternative assumptions that have been imposed in the literature, in particular a stationarity assumption often utilized in dynamic panel approaches. Future work on this topic could go in at least a couple of directions. First, it would be interesting to test these various assumptions. In particular, if one is willing to assume that  $\Delta = a$ , then one can in theory use overidentifying restrictions to test the stronger assumption that  $\Delta = a - k$  for  $k > 0$ . More generally, one might use this idea to try to estimate  $\Delta$ , or at least put a lower bound on it. In some cases, one might expect  $\Delta$  to change over time. For example, in a production function context, advances in technology might allow firms to make decisions on inputs closer to production time, or have more information when they make those choices. In other words,  $\Delta$  could be a source of "productivity" improvements distinct from actual production function coefficients. This is related to, e.g. recent work by Asker, Collard-Wexler, and DeLoecker (2015). Second, it might be interesting to investigate these assumptions from a mean squared error perspective, in which case one might be willing to accept some bias in return for lower variance. So for example, even if one knows that  $\Delta = a$ , one might still make the stronger (incorrect) assumption that  $\Delta = a - k$  for  $k > 0$  for the mean/variance tradeoff.

## References

- Akerberg, D., Benkard, L., Berry, S., and Pakes, A. (2007) "Econometric Tools for Analyzing Market Outcomes", *Handbook of Econometrics*. Amsterdam: North-Holland.
- Akerberg, D., Caves, K., and Frazer, G. (2015) "Identification Properties of Recent Production Function Estimators" *Econometrica* 83: 2411-2451
- Akerberg, D, and Hahn J (2015) "Some Non-Parametric Identification Results using Timing and Information Set Assumptions" mimeo
- Anderson T.W. and C. Hsiao (1982) "Formulation and Estimation of Dynamic Models using Panel Data" *Journal of Econometrics* 18: 47-82.
- Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations" *The Review of Economic Studies* 58: 277-297.
- Arellano, M. and Bover, O. (1995) "Another Look at the Instrumental Variable Estimation of Error Components Models" *Journal of Econometrics*, 68: 29-51
- Asker, J., Collard-Wexler, A. and De Loecker, J. (2014) "Dynamic Inputs and Resource (Mis)Allocation" *Journal of Political Economy* 122: 1013-1063
- Blundell, R. and Bond, S. (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models" *Journal of Econometrics* 87: 115-143.
- Blundell, R. and Bond, S. (2000) "GMM estimation with persistent panel data: an application to production functions", *Econometric Reviews* 19: 321-340
- Bond, S. and Söderbom, M. (2005) "Adjustment Costs and the Identification of Cobb-Douglas Production Functions," The Institute for Fiscal Studies, Working Paper Series No. 05/04.
- Chamberlain, G. (1982) "Multivariate Regression Models for Panel Data", *Journal of Econometrics* 18: 5-46.
- Collard-Wexler, A. and De Loecker, J. (2016) "Production Function Estimation with Measurement Error in Inputs" mimeo
- De Loecker, J. (2011) "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 79: 1407-1451.
- De Loecker, J. and Warzynski, F. (2012) "Markups and Firm-level Export Status," *American Economic Review*, 102: 2437-2471.
- Doraszelski, U. and Jaumandreu, J. (2013) "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, 80: 1338-1383.
- Gandhi, A. Navarro, S., and Rivers, D. (2016) "On the Identification of Production Functions: How Heterogeneous is Productivity," forthcoming, *Journal of Political Economy*.

- Grennan, M. (2013), "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices", *American Economic Review*, 103: 145-177
- Grieco, P., Li, S., and H. Zhang (2016) "Production Function Estimation with Unobserved Input Price Variation" *International Economic Review*, 57: 665-690
- 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. S. Strøm. Cambridge, UK: Cambridge University Press.
- Huang, G. and Hu, Y. (2011). "Estimating Production Functions with Robustness Against Errors in the Proxy Variables," mimeo, Johns Hopkins University.
- Kim, K, Petrin, A. and Song, S. (2016) "Estimating Production Functions with Control Functions when Capital is Measured with Error," *Journal of Econometrics* 190: 267-279
- Lee, R. (2013) "Vertical Integration and Exclusivity in Platform and Two-Sided Markets" *American Economic Review* 103: 2960-3000
- Levinsohn, J. and Petrin, A. (2003) "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies* 70: 317-342
- Olley, S. and Pakes, A.(1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry" *Econometrica* 64:1263-1295
- Pagan, A. (1986) "Two Stage and Related Estimators and their Applications" *Review of Economic Studies* 53: 517-538
- Pakes, A. and Olley, S. (1995). "A Limit Theorem for a Smooth Class of Semiparametric Estimators," *Journal of Econometrics* 65: 295-332.
- Shenoy, A. (2016) "Estimating the Production Function when Firms are Constrained", mimeo, UCSC
- Söderbom, M (2016) Personal Website <http://www.soderbom.net/Resources.htm>
- Sullivan, C. (2017) "The Ice-Cream Split: Empirically Distinguishing Price and Product Space Collusion", mimeo, Wisconsin
- Sweeting, A. (2013) "Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry" *Econometrica* 81:1763-1803
- Van Biesebroeck, J. (2007) "Robustness of productivity estimates," *The Journal of Industrial Economics* , 55: 529-569
- Wooldridge (2009) "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 104: 112-114.

Table 1 - Returns to Scale Estimates with Different Timing ( $\Delta$ )  
and Stationarity (DIFF-GMM vs SYS-GMM) Assumptions

Chile						
			$\Delta = -1$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$
Industry 312 Food Products n=10790	DIFF-GMM	Estimate	0.674	1.010	2.241	0.803
		SE	(0.049)	(0.157)	(0.356)	(0.541)
	SYS-GMM	Estimate	1.522	1.151	1.439	1.285
		SE	(0.023)	(0.059)	(0.114)	(0.123)
Industry 321 Textiles n=2689	DIFF-GMM	Estimate	0.555	0.666	0.868	0.640
		SE	(0.077)	(0.144)	(0.164)	(0.197)
	SYS-GMM	Estimate	1.231	1.121	1.093	1.170
		SE	(0.023)	(0.075)	(0.109)	(0.149)
Industry 331 Wood Products n=2539	DIFF-GMM	Estimate	0.887	0.921	1.446	1.630
		SE	(0.092)	(0.154)	(0.292)	(0.370)
	SYS-GMM	Estimate	1.323	1.205	1.393	1.245
		SE	(0.036)	(0.109)	(0.172)	(0.281)
Mexico						
			$\Delta = -1$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$
Food Products n=1849	DIFF-GMM	Estimate	0.557	0.336	1.891	2.804
		SE	(0.083)	(0.365)	(0.775)	(1.258)
	SYS-GMM	Estimate	0.962	0.368	0.185	0.509
		SE	(0.042)	(0.188)	(0.372)	(0.290)
Textiles n=919	DIFF-GMM	Estimate	0.580	1.213	1.230	2.289
		SE	(0.078)	(0.339)	(0.395)	(0.839)
	SYS-GMM	Estimate	0.977	0.875	0.519	0.304
		SE	(0.052)	(0.172)	(0.267)	(0.360)
Wood Products n=378	DIFF-GMM	Estimate	0.464	0.849	2.684	2.868
		SE	(0.224)	(0.499)	(1.060)	(1.270)
	SYS-GMM	Estimate	1.163	0.976	0.879	1.247
		SE	(0.106)	(0.311)	(0.220)	(0.489)