Structural identification of production functions

Ackerberg, Daniel and Caves, Kevin and Frazer, Garth

UCLA

26 December 2006

Online at https://mpra.ub.uni-muenchen.de/38349/
MPRA Paper No. 38349, posted 25 Apr 2012 00:44 UTC
Abstract

This paper examines some of the recent literature on the identification of production functions. We focus on structural techniques suggested in two recent papers, Olley and Pakes (1996), and Levinsohn and Petrin (2003). While there are some solid and intuitive identification ideas in these papers, we argue that the techniques, particularly those of Levinsohn and Petrin, suffer from collinearity problems which we believe cast doubt on the methodology. We then suggest alternative methodologies which make use of the ideas in these papers, but do not suffer from these collinearity problems.

1 Introduction

Production functions are a fundamental component of all economics. As such, estimation of production functions has a long history in applied economics, starting in the early 1800’s. Unfortunately, this history cannot be deemed an unqualified success, as many of the econometric problems that hampered early estimation are still an issue today.

Production functions relate productive inputs (e.g. capital, labor) to outputs. Perhaps the major econometric issue confronting estimation of production functions is the possibility that there are determinants of production that are unobserved to the econometrician but observed by the firm. If this is the case, and if the observed inputs are chosen as a function of these determinants (as will typically be the case for a profit-maximizing or cost-minimizing firm), then there is an endogeneity problem and OLS estimates of the coefficients on the observed inputs will be biased.

Much of the literature in the past half century has been devoted to solving this endogeneity problem. Two of the earliest solutions to the problem are instrumental variables (IV) and fixed-effects estimation (Mundlak (1961)). IV estimation requires finding variables that are correlated...
with observed input choices, but uncorrelated with the unobservables determining production. Fixed-effects estimation requires the assumption that the unobservables are constant across time. Unfortunately, for a variety of reasons, these methodologies have not been particularly successful at solving these endogeneity problems. As such the search has continued for reliable methods for identifying production function parameters.

The past fifteen years has seen the introduction of a couple of new techniques for identification of production functions. One set of techniques follows the dynamic panel data literature, e.g. Chamberlain (1982), Arellano and Bover (1995), Blundell and Bond (2000). A second set of techniques, advocated by Olley and Pakes (1996) and Levinsohn and Petrin (2003), are somewhat more structural in nature - using observed input decisions to "control" for unobserved productivity shocks. This second set of techniques has been applied in a large number of recent empirical papers, including Pavcnik (2002), Sokoloff (2003), Sivadasan (2004), Fernandes (2003), Ozler and Yilmaz (2001), Criscuola and Martin (2003), Topalova (2003), Blalock and Gertler (2004), and Alvarez and Lopez (2005).¹

This paper starts by analyzing this second set of techniques. We first argue that there are potentially serious collinearity problems with these estimation methodologies.² We show that, particularly for the Levinsohn and Petrin approach, one needs to make what we feel are very strong and unintuitive assumptions for the model to remain correctly identified in the wake of this collinearity problem. To address this problem, we then suggest an alternative estimation approach. This approach builds upon the ideas in Olley and Pakes and Levinsohn and Petrin, e.g. using investment or intermediate inputs to "proxy" for productivity shocks, but does not suffer from these collinearity problems. As well as solving the above collinearity problem, another important benefit of our estimator is that it makes comparison to the aforementioned dynamic panel literature, e.g. Blundell and Bond, quite easy. This is important, as up to now, the two literatures have evolved separately. In particular, our estimator makes it quite easy to see the tradeoffs in assumptions needed by the two distinct literatures. We feel that this should help guide empirical researchers in choosing between the approaches. Lastly, using the same dataset as Levinsohn and Petrin, we examine how our estimator works in practice. Estimates using our methodology appear more stable across different potential proxy variables than the Levinsohn-Petrin methodology, consistent with our theoretical arguments.

¹This list is far from exhaustive. A recent search using Google Scholar shows 598 cites of Olley and Pakes (1996) and 219 cites of Levinsohn and Petrin (2003).

²Susanto Basu made a less formal argument regarding this possible collinearity problem in 1999 as a discussant of an earlier version of the Levinsohn-Petrin paper.
2 Review of Olley/Pakes and Levinsohn/Petrin

We start with a brief review of the techniques of Olley/Pakes (henceforth OP) and Levinsohn/Petrin (henceforth LP). Consider the following Cobb-Douglas production function in logs:

\[ y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \epsilon_{it} \]

\( y_{it} \) is the log of output, \( k_{it} \) is the log of capital input, and \( l_{it} \) is the log of labor input.\(^3\) There are two terms in this equation that are unobservable to the econometrician, \( \omega_{it} \) and \( \epsilon_{it} \). The distinction between the two is important. The \( \epsilon_{it} \) are intended to represent shocks to production or productivity that are not observable (or predictable) by firms before making their input decisions at \( t \). In contrast, the \( \omega_{it} \) represent shocks that are potentially observed or predictable by firms when they make input decisions. Intuitively, \( \omega_{it} \) might represent variables such as the managerial ability of a firm, expected down-time due to machine breakdown, expected defect rates in a manufacturing process, or the expected rainfall at a farm’s location. On the other hand, \( \epsilon_{it} \) might represent deviations from expected breakdown, defect, or rainfall amounts in a given year. \( \epsilon_{it} \) can also represent measurement error in the output variable. We will often refer to \( \omega_{it} \) as the "productivity shock" of firm \( i \) in period \( t \). Note that we have subsumed the constant term in the production function into the productivity term \( \omega_{it} \).

The classic endogeneity problem estimating equation (1) is that the firm’s optimal choice of inputs \( k_{it} \) and \( l_{it} \) will generally be correlated with the observed or predictable productivity shock \( \omega_{it} \). This renders OLS estimates of the \( \beta \)'s biased and inconsistent. As mentioned in the introduction, perhaps the two most commonly used solutions to this endogeneity problem are fixed effects (Mundlak (1961), Hoch (1962)) and instrumental variables estimation techniques. In our context, fixed-effects estimation requires the additional assumption that \( \omega_{it} = \omega_{it-1} \forall t \). This is a strong assumption and, perhaps as a result, the technique has not worked well in practice - often generating unrealistically low estimates of \( \beta_k \). IV estimation requires instruments that are correlated with input choices \( k_{it} \) and \( l_{it} \) and uncorrelated with \( \omega_{it} \). On one hand, there do exist natural instrumental variables in this situation - input prices, as long as one is willing to assume firms operate in competitive input markets. On the other hand, this again has not worked well in practice. Too often these input prices are not observed, do not vary or vary enough across firms, or are suspected to pick up variables, e.g. input quality, that would invalidate their use as instruments. The review of this literature in Ackerberg, Benkard, Berry, and Pakes (2005) (ABBP) contains more discussion of the limitations of the fixed effects and IV approaches.

\(^3\)These inputs and outputs are measured in various ways across studies depending on data availability. For example, labor inputs could be measured in man-hours, or in money spent on labor. Output could also be measured in either physical or monetary units, and in some cases is replaced with a value added measure.
2.1 Olley and Pakes

The OP and LP methodologies take a more structural approach to identification of production functions. OP address the endogeneity problem as follows. They consider a firm operating through discrete time, making decisions to maximize the present discounted value of current and future profits. First, they assume that the productivity term \( \omega_{it} \) evolves exogenously following a first-order Markov process, i.e.

\[
p(\omega_{it+1}|I_{it}) = p(\omega_{it+1}|\omega_{it})
\]

where \( I_{it} \) is firm \( i \)'s information set at \( t \). Current and past realizations of \( \omega \), i.e. \( (\omega_{it}, ..., \omega_{i0}) \) are assumed to be part of \( I_{it} \). Importantly, this is not just an econometric assumption on the statistical properties of unobservables. It is also an economic assumption regarding what determines a firm’s expectations about future productivity \( \omega_{it+1} \), i.e. that these expectations depend only on \( \omega_{it} \).

OP assume that labor is a non-dynamic input. More specifically, a firm’s choice of labor for period \( t \) has no impact on the future profits of the firm. In contrast, capital is assumed to be a dynamic input subject to an investment process. Specifically, in every period, the firm decides on an investment level \( i_{it} \). This investment adds to future capital stock deterministically, i.e.

\[
k_{it} = \kappa(k_{it-1}, i_{it-1})
\]

Importantly, this formulation implies that the period \( t \) capital stock of the firm was actually determined at period \( t - 1 \). The economics behind this is that it may take a full period for new capital to be ordered, delivered, and installed. Intuitively, one can see how this assumption regarding timing helps solve the endogeneity problem with respect to capital. Since \( k_{it} \) is actually decided upon at \( t - 1 \) (and thus is in \( I_{it-1} \)), the above informational assumptions imply that it must be uncorrelated with the unexpected innovation in \( \omega_{it} \) between \( t - 1 \) and \( t \), i.e. \( \omega_{it} - E[\omega_{it}|I_{it-1}] = \omega_{it} - E[\omega_{it} | \omega_{it-1}, I_{it-1}] \). This orthogonality will be used to form a moment to identify \( \beta_k \).

More challenging is solving the endogeneity problem with respect to the assumed variable input, \( l_{it} \). This is because unlike capital, \( l_{it} \) is decided at \( t \) and thus potentially correlated with even the innovation component of \( \omega_{it} \). To accomplish this, OP make use of the investment variable \( i_{it} \). Considering the firm’s dynamic decision of investment level \( i_{it} \), OP state conditions under which a firm’s optimal investment level is a strictly increasing function of their current productivity \( \omega_{it} \), i.e.

\[
i_{it} = f_i(\omega_{it}, k_{it})
\]

\(^4\)In the special case where \( \omega_{it} \) is a random walk, i.e. \( \omega_{it} = \omega_{it-1} + \eta_{it} \), one can easily see how this can be done - if we first-difference the production function, \( (k_{it} - k_{it-1}) \) is uncorrelated with the resulting unobserved term.
Note that this investment function will in general contain all current state variables for the optimizing firm, e.g. its current level of capital and the current $\omega_{it}$. Labor does not enter the state because it is a non-dynamic input, and values of $\omega_{it}$ prior to $t$ do not enter because of the first order Markov assumption on the $\omega_{it}$ process. The reason $f$ is indexed by $t$ is that variables such as input prices, demand, etc. also may be part of the state space. OP simply treat these as part of $f_t$. The assumption here is that these variables are allowed to vary across time, but not across firms (i.e. firms operate in the same input markets).

Given that this investment function is strictly monotonic in $\omega_{it}$, it can be inverted to obtain

$$\omega_{it} = f_t^{-1}(i_{it}, k_{it}) \quad (4)$$

The essence of OP is to use this inverse function to control for $\omega_{it}$ in the production function. Substituting this into the production function, we get:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + f_t^{-1}(i_{it}, k_{it}) + \epsilon_{it} \quad (5)$$

$$= \beta_l l_{it} + \Phi_t(i_{it}, k_{it}) + \epsilon_{it} \quad (6)$$

The first stage of OP involves estimating this equation. Recall that $f$ is the solution to a complicated dynamic programming problem. As such, solving for $f$ (and thus $f^{-1}$) would not only require assuming all the primitives of the firm (e.g. demand conditions, evolution of environmental state variables), but also be computationally demanding. To avoid these extra assumptions and computations, OP simply treat $f_t^{-1}$ non-parametrically. Given this non-parametric treatment, direct estimation of (5) does not identify $\beta_k$, as $k_{it}$ is collinear with the non-parametric function. However, one does obtain an estimate of the labor coefficient $\beta_l$, $\widehat{\beta}_l$. One also obtains an estimate of the composite term $\Phi_t(i_{it}, k_{it}) = \beta_1 k_{it} + f_t^{-1}(i_{it}, k_{it})$, which we denote $\widehat{\Phi}_t$.

The second stage of OP proceeds given these estimates of $\widehat{\beta}_l$ and $\widehat{\Phi}_t$. Given (2), we can write

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

This simply decomposes $\omega_{it}$ into its conditional expectation at time $t - 1$, $E[\omega_{it} | I_{it-1}]$, and a deviation from that expectation, $\xi_{it}$. The second equality follows from the first order markov assumption on $\omega_{it}$. We will often refer to $\xi_{it}$ as the "innovation" component of $\omega_{it}$. By the properties of a conditional expectation, this innovation component satisfies:

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

Thus, since the timing assumption regarding capital implies that $k_{it} \in I_{it-1}$ (since $k_{it}$ was decided
at \( t - 1 \), this implies that \( \xi_{it} \) is orthogonal to \( k_{it} \), i.e.

\[
E[\xi_{it}|k_{it}] = 0
\]

This mean independence in turn implies that \( \xi_{it} \) and \( k_{it} \) are uncorrelated, i.e.

\[
E[\xi_{it}k_{it}] = 0
\]

This is the moment which OP use to identify the capital coefficient. Loosely speaking, variation in \( k_{it} \) conditional on \( \omega_{it-1} \) is the exogenous variation being used for identification here. To operationalize this procedure in a GMM context, note that given a guess at the capital coefficient \( \beta_k \), one can "invert" out the \( \omega_{it} \)'s in all periods, i.e.

\[
\omega_{it}(\beta_k) = \bar{\Psi}_{it} - \beta_k k_{it}
\]

Given these \( \omega_{it}(\beta_k) \)'s, one can compute \( \xi_{it} \)'s in all periods by non-parametrically regressing \( \omega_{it}(\beta_k) \)'s on \( \omega_{it-1}(\beta_k) \)'s (and a constant term) and forming the residual

\[
\xi_{it}(\beta_k) = \omega_{it}(\beta_k) - \bar{\Psi}(\omega_{it-1}(\beta_k))
\]

where \( \bar{\Psi}(\omega_{it-1}(\beta_k)) \) are predicted values from the non-parametric regression.\(^5\) This non-parametric treatment of the regression of \( \omega_{it} \) on \( \omega_{it-1} \) allows for \( \omega_{it} \) to follow an arbitrary first-order Markov process. These \( \xi_{it}(\beta_k) \)'s can then be used to form a sample analogue to the above moment, i.e.

\[
\frac{1}{T} \sum_{t=1}^{T} \sum_{i} \xi_{it}(\beta_k) \cdot k_{it}
\]

In a GMM procedure, \( \beta_k \) is estimated by setting this empirical analogue as close as possible to zero.\(^6\) Quickly recapping the intuition behind identification in OP, \( \beta_i \) is identified by using the information in firms' investment decisions \( i_{it} \) to control for the productivity shock \( \omega_{it} \) that is correlated with \( l_{it} \). \( \beta_k \) is identified by the timing assumption that \( k_{it} \) is decided before the full realization of \( \omega_{it} \).\(^7\)

---

5. Note that both OP and LP use a slightly different moment condition than this. Instead of regressing implied \( \omega_{it} \) on implied \( \omega_{it-1} \), they regress \( y_{it} - \beta_k k_{it} - \beta_i l_{it} \) on implied \( \omega_{it-1} \). Their procedure corresponds to a moment in the residual \( \xi_{it} + \epsilon_{it} \) rather than our procedure, which corresponds to a moment in the residual \( \xi_{it} \). In our experience, the moment in \( \xi_{it} \) tends to produce lower variance and more stable estimates. This is probably because the extra \( \epsilon_{it} \) term adds variance to the moment, thus increasing the variance of the estimates.

6. Wooldridge (2005) shows how one can perform both the first and second stages of OP (or LP) simultaneously. Not only is this more efficient, but it also makes it easier to compute standard errors. We discuss the Wooldridge moments in more detail when we describe our suggested procedure. For details on standard errors for the OP 2-step process, see Pakes and Olley (1995).

7. Recall that the constant term in the production function is subsumed into the \( \omega_{it} \)'s. Hence the above procedure does not produce a direct estimate of the constant term. To form an estimate of the constant term ex-post, one
2.2 Levinsohn and Petrin

LP take a related approach to solving the production function endogeneity problem. The key difference is that rather than using the investment demand equation, they use an intermediate input demand function to "invert" out $\omega_{it}$. Their motivation for this alternative inversion equation is very reasonable. For the straightforward OP procedure to work, recall one needs the investment function to be strictly monotonic in $\omega_{it}$. However, in actual data, investment is often very lumpy, and one often sees zeros. In the Chilean data studied by LP, for example, more than 50% of firm-year observations have zero investment. This casts doubt on this strict monotonicity assumption regarding investment. While the OP procedure can actually work in this situation, it requires discarding the data with zero investment (see ABBP for discussion), an obvious efficiency loss.

LP avoid this efficiency loss by considering the following production function:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \omega_{it} + \epsilon_{it}$$

where $m_{it}$ is an intermediate input such as electricity, fuel, or materials. LP’s basic idea is that since intermediate input demands are typically much less lumpy (and prone to zeros) than investment, the strict monotonicity condition is more likely to hold and these may be superior "proxies" to invert out the unobserved $\omega_{it}$. LP assume the following intermediate input demand function:

$$m_{it} = f_t(\omega_{it}, k_{it})$$

Again, $f$ is indexed by $t$, implicitly allowing input prices (and/or market conditions) to vary across time (but not across firms). Note the timing assumptions implicit in this formulation. First, the intermediate input at $t$ is chosen as a function of $\omega_{it}$. This implies that the intermediate input is essentially chosen at the time production takes place. We describe this as a "perfectly variable" input. Secondly, note that $l_{it}$ does not enter (7). This implies that labor is also a "perfectly variable" input, i.e. chosen simultaneously with $m_{it}$.\(^8\) If $l_{it}$ was chosen at some point in time before $m_{it}$, then $l_{it}$ would impact the firm’s optimal choice of $m_{it}$.

Given this specification, LP proceed similarly to OP. Under the assumption that intermediate input demand (7) is monotonic in $\omega_{it}$, we can invert:

$$\omega_{it} = f_t^{-1}(m_{it}, k_{it})$$

can simply compute the average of the implied $\omega_{it}(\beta_k)$ evaluated at the estimate $\hat{\beta}_k$ (or, to allow the constant term to vary across time, one would just use the average of $\omega_{it}(\hat{\beta}_k)$ at each time period).

\(^8\)Note the difference between 1) the distinction of whether an input is variable or fixed, and 2) the distinction of whether an input is dynamic or non-dynamic. 1) refers to the point in time in which the input is chosen. 2) refers to whether the choice of the input currently has any implications on future profits.

\(^9\)LP provide conditions on primitives such that this is the case.
Substituting this into the production function gives

\begin{equation}
    y_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + f_{it}^{-1}(m_{it}, k_{it}) + \epsilon_{it}
\end{equation}

The first step of the LP estimation procedure estimates \( \beta_l \) using the above equation, treating \( f_{it}^{-1} \) non-parametrically. Again, \( \beta_k \) and \( \beta_m \) are not identified as \( k_{it} \) and \( m_{it} \) are collinear with the non-parametric term. One also obtains an estimate of the composite term, in this case \( \beta_k k_{it} + \beta_m m_{it} + f_{it}^{-1}(m_{it}, k_{it}) \), which we again denote \( \hat{\Phi}_{it} \).

The second stage of the LP procedure proceeds as OP, the only difference being that there is one more parameter to estimate, \( \beta_m \). LP use the same moment condition as OP to identify the capital coefficient, i.e. that the innovation component of \( \xi_{it}, \omega_{it} \), is orthogonal to \( k_{it} \). \( \xi_{it}(\beta_k, \beta_m) \) can again be constructed as the residual from a non-parametric regression of \((\omega_{it}(\beta_k, \beta_m) = \hat{\Phi}_{it} - \beta_k k_{it} - \beta_m m_{it}) \) on \((\omega_{it-1}(\beta_k, \beta_m) = \hat{\Phi}_{it-1} - \beta_k k_{it-1} - \beta_m m_{it-1}) \). They also add an additional moment to identify \( \beta_m \), the condition that that \( \xi_{it}(\beta_k, \beta_m) \) is orthogonal to \( m_{it-1} \). This results in the following moment condition on which to base estimation:

\[
    E[\xi_{it}(\beta_k, \beta_m) | k_{it-1}, m_{it-1}] = 0
\]

Note that the innovation \( \xi_{it} \) is clearly not orthogonal to \( m_{it} \). This is because \( \omega_{it} \) is observed at the time that \( m_{it} \) is chosen. On the other hand, according to the model, \( \xi_{it} \) should be uncorrelated with \( m_{it-1} \), as \( m_{it-1} \) was decided at \( t-1 \) and hence part of \( I_{it-1} \).

2.3 Key Assumptions of OP and LP

Note that both the OP and LP procedures rely on a number of key structural assumptions in addition to the first order markov assumption on the \( \omega_{it} \) process. While these assumptions are described in these papers (see also Griliches and Mairesse (1998) and ABBP), we summarize them here. A first key assumption is the strict monotonicity assumption - for OP investment must be strictly monotonic in \( \omega_{it} \) (at least when it is non-zero), while for LP intermediate input demand must be strictly monotonic in \( \omega_{it} \). Monotonicity is required for the non-parametric inversion because otherwise, one cannot perfectly invert out \( \omega_{it} \) and completely remove the endogeneity problem in (5).

A second key assumption is that \( \omega_{it} \) is the only unobservable entering the functions for investment (OP) or the intermediate input (LP). We refer to this as a "scalar unobservable" assumption. This rules out, e.g. measurement error or optimization error in these variables, or a model in which exogenous productivity is more than single dimensional. Again, the reason for this assumption is that if either of these were the case, one would not be able to perfectly invert out \( \omega_{it} \).\(^{10}\)

\(^{10}\) ABBP discuss how this assumption can be relaxed in some very specific dimensions (e.g. allowing \( \omega_{it} \) to
A third key set of assumptions of the models regards the timing and dynamic implications of input choices. By timing, we refer to the point in the \( \omega_{it} \) process at which inputs are chosen. First, \( k_{it} \) is assumed to have been decided exactly at (OP) or exactly at/prior to (LP) time period \( t - 1 \). Any later than this would violate the moment condition, as \( k_{it} \) would likely no longer be orthogonal to the innovation term \( \xi_{it} \). For OP, were \( i_{it-1} \) (and thus \( k_{it} \)) to be decided any earlier than \( t - 1 \), then one could not use \( i_{it-1} \) to invert out \( \omega_{it-1} \), making first-stage estimation problematic.

Regarding the labor input, there are a couple of important assumptions. First, in OP, \( l_{it} \) must have no dynamic implications. Otherwise, \( l_{it} \) would enter the investment demand function and prevent identification of the labor coefficient in the first stage. In LP, labor can have dynamic implications, but one would need to adjust the procedure suggested by LP by allowing \( l_{it-1} \) into the intermediate input demand function. Note that in principle, this still allows one to identify the coefficient on labor in the first stage. Second, for LP it is important that \( l_{it} \) and \( m_{it} \) are assumed to be perfectly variable inputs. By this we mean that they are decided \textit{when} \( \omega_{it} \) is observed by the firm. If \( m_{it} \) were decided before learning \( \omega_{it} \), then \( m_{it} \) could not be used to invert out \( \omega_{it} \) and control for it in the first stage. If \( l_{it} \) were chosen before learning \( \omega_{it} \), then \( l_{it} \) would also be chosen before \( m_{it} \). In this case, a firm’s choice of materials \( m_{it} \) would directly depend on \( l_{it} \) and \( l_{it} \) would enter the LP non-parametric function, preventing identification of the labor coefficient in the first stage.

### 3 Collinearity Issues

This paper argues that even if the above assumptions hold, there are potentially serious identification issues with these methodologies, particularly the LP approach. The problem is one of collinearity arising in the first stage of the respective estimation procedures, respectively:

\[
y_{it} = \beta_l l_{it} + f_t^{-1}(i_{it}, k_{it}) + \epsilon_{it}
\]

and

\[
y_{it} = \beta_l l_{it} + f_t^{-1}(m_{it}, k_{it}) + \epsilon_{it}
\]

where the obviously non-identified terms (\( \beta_k k_{it} \) in OP, \( \beta_k k_{it} \) and \( \beta_m m_{it} \) in LP) have been subsumed into the non-parametric functions. Recall that in the first stage, the main goal in both methods is to identify \( \beta_l \), the coefficient on the labor input. What we now focus on is the question of whether even \( \beta_l \) can be identified from these regressions under the above assumptions. There is clearly no endogeneity problem - \( \epsilon_{it} \) are either unanticipated shocks to production not known at \( t \) (follow a higher than first order Markov process), but all these cases require the econometrician to observe and use additional control variables in the first stage.
or purely measurement error in output, so they are by assumption are uncorrelated with all the right hand variables. Thus, the only real identification question here is whether $l_{it}$ is "collinear" with the non-parametric terms in the respective regressions, i.e. whether $l_{it}$ varies independently of the non-parametric function that is being estimated.

### 3.1 Levinsohn and Petrin

First, consider the LP technique. To think about whether $l_{it}$ varies independently of $f_t^{-1}(m_{it}, k_{it})$, we need to think about the data generating process for $l_{it}$, i.e. how the firm chooses $l_{it}$. Given that we have already assumed that $l_{it}$ and $m_{it}$ are chosen simultaneously and are both perfectly-variable, non-dynamic inputs, a natural assumption might be that they are decided in similar ways. Since $m_{it}$ has been assumed to be chosen according to

\[(12) \quad m_{it} = f_t(\omega_{it}, k_{it})\]

this suggests that $l_{it}$ might be chosen according to

\[(13) \quad l_{it} = g_t(\omega_{it}, k_{it})\]

While $g_t$ will typically be a different function than $f_t$ (e.g. because of different prices of the inputs), they both will generally depend on the same state variables, $\omega_{it}$ and $k_{it}$. Intuitively, this is just saying that the choice of both variable inputs at $t$ depends on the predetermined value of the fixed input and the current productivity shock.

Substituting (8) into (13) results in

\[
l_{it} = g_t(f_t^{-1}(m_{it}, k_{it}), k_{it}) = h_t(m_{it}, k_{it})
\]

which states that $l_{it}$ is some time-varying function of $m_{it}$ and $k_{it}$. While this is a very simple result, it has some very strong implications on the LP first stage estimating equation (11). In particular, it says that the coefficient $\beta_t$ is not identified. One simply cannot simultaneously estimate a fully non-parametric (time-varying) function of $(\omega_{it}, k_{it})$ along with a coefficient on a variable that is only a (time-varying) function of those same variables $(\omega_{it}, k_{it})$. Given this perfect collinearity between $l_{it}$ and the non-parametric function, $\beta_t$ should not be identified.

That said, while (13) might be the most natural specification for the data generation process (DGP) for $l_{it}$, it is not the only possibility. Our goal now is to search for an alternative DGP for $l_{it}$ (and possibly for $m_{it}$) that will allow the LP first stage procedure to work. Not only must this alternative DGP move $l_{it}$ around independently of the non-parametric function $f_t^{-1}(m_{it}, k_{it})$, but it must simultaneously be consistent with the basic assumptions of the LP procedure detailed in the last section.
First, consider adding firm-specific input prices to the above model of input choice, e.g. prices of labor \((p_{il})\) and materials \((p_{im})\). Obviously these firm-specific input prices will generally affect a firm’s choices of \(l_{it}\) and \(m_{it}\). A first note is that these input prices would have to be observed by the econometrician. Unobserved firm-specific input prices would enter \((7)\) and violate the scalar unobservable assumption necessary for the first stage LP inversion. In other words, with unobserved firm-specific input prices, one can no longer invert out the firm’s productivity shock as a function of the observables \(m_{it}\) and \(k_{it}\) and perform the first stage.

If the firm-specific input prices are observed, the inversion is not a problem - one simply can include the observed input prices in the non-parametric function. However, for the same reason, observed firm-specific input prices also do not solve the collinearity problem. Given that \(l_{it}\) and \(m_{it}\) are set at the same points in time, they will generally both be a function of both \(p_{il}\) and \(p_{im}\). As such, we have the same problem as before - there are no variables that affect \(l_{it}\) but that do not affect \(m_{it}\) (and thus enter the non-parametric function). Our conclusion is that firm-specific input prices do not generally help matters. A related possibility is to allow labor to have dynamic effects. As discussed above, this is consistent with the LP assumptions as long as one adds \(l_{it-1}\) to the first stage non-parametric term. However, for the same reason, dynamic labor does not break the collinearity problem. As both \(l_{it}\) and \(m_{it}\) will generally depend on \(l_{it-1}\), the term will not move \(l_{it}\) around independently of \(f_{it}^{-1}\).

In the basic model described above, \(l_{it}\) and \(m_{it}\) are chosen simultaneously at period \(t\), i.e. after observing \(\omega_{it}\). A second alternative to try to break the collinearity problem is to perturb the model by changing these points in time at which \(l_{it}\) and \(m_{it}\) are set, i.e. allow \(l_{it}\) to be set before or after \(m_{it}\). To formally analyze these situations, consider a point in time, \(t - b\), between period \(t - 1\) and \(t\) (i.e. \(0 < b < 1\)). Assume that \(\omega\) evolves through these "subperiods" \(t - 1\), \(t - b\), and \(t\) according to a first order Markov process, i.e.

\[
p(\omega_{it-b}|I_{it-1}) = p(\omega_{it-b}|\omega_{it-1})
\]

and

\[
p(\omega_{it}|I_{it-b}) = p(\omega_{it}|\omega_{it-b})
\]

Note that we continue to assume that production occurs "on the period", i.e. at periods \(t - 1\) and \(t\). The main point of introducing the subperiod \(t - b\) is to allow the firm to have a different information set when choosing \(l_{it}\) than when choosing \(m_{it}\). The hope is that these different information sets might generate independent variation in the two variables that could break the collinearity problem.

Given this setup, we can now consider perturbing the points in time at which \(l_{it}\) and \(m_{it}\) are set. First consider the situation where \(m_{it}\) is chosen at \(t - b\) and \(l_{it}\) is chosen at \(t\). Now a firm’s optimal choice of \(m_{it}\) will depend on \(\omega_{it-b}\), while the choice of \(l_{it}\) will depend on \(\omega_{it}\). In this setup,
$l_{it}$ does have variance that is independent of $m_{it}$, because of the innovation in $\omega_{it}$ between $\omega_{it-b}$ and $\omega_{it}$. However, this setup is also problematic for the first stage of the LP procedure. Since $m_{it}$ is a function of $\omega_{it-b}$, not $\omega_{it}$, it cannot completely inform us regarding $\omega_{it}$. In other words, the first stage non-parametric function will not be able to capture the entire productivity shock $\omega_{it}$. Unfortunately, the part of $\omega_{it}$ that is not captured and left in the residual (which amounts to the unexpected innovation in $\omega_{it}$ given $\omega_{it-b}$, i.e. $\omega_{it} - E[\omega_{it}|\omega_{it-b}]$) will be highly correlated with any independent variation in $l_{it}$. This creates an endogeneity problem whereby first stage estimates of $\beta_i$ will be biased.$^{11}$

Alternatively, consider the situation where where $l_{it}$ is chosen at $t-b$ and $m_{it}$ is chosen at $t$. Again, in this case, the fact that $m_{it}$ and $l_{it}$ are chosen with different information sets generates independent variation. However, in this case, there is another problem. Since $l_{it}$ is chosen before $m_{it}$, a profit maximizing (or cost-minimizing) firm’s optimal choice of $m_{it}$ will generally directly depend on $l_{it}$, i.e.

$$m_{it} = f_t(l_{it}, \omega_{it}, k_{it})$$

Given this, $l_{it}$ should directly enter the first-stage non-parametric function and an LP first stage estimate of $\beta_i$ is obviously not identified. In summary, neither of these timing stories appears to be able to justify the LP first stage procedure.

We next consider stories based on measurement error or optimization error on the part of firms. The difference between the two can be illustrated in the following model

$$m_{it} = m^*_it + \lambda^m_{it} = f_t(\omega_{it}, k_{it}) + \lambda^m_{it}$$

where, as above, $m_{it}$ is the value of the material input choice observed by the econometrician. When $\lambda^m_{it}$ represents measurement error, $m^*_it$ is the variable that actually enters the production function. When $\lambda^m_{it}$ represents optimization error, $m_{it}$ is the variable entering the production function. With optimization error, a firm should optimally be choosing input level $f_t(\omega_{it}, k_{it})$, but for some reason chooses $f_t(\omega_{it}, k_{it}) + \lambda^m_{it}$ instead.

A first observation is that neither measurement error or optimization error in $m_{it}$ is a workable solution to the collinearity problem. Either measurement error or optimization error in $m_{it}$ adds another unobservable to the $m_{it}$ equation, which violates the scalar unobservable assumption. In either case, we can not write $\omega_{it}$ as a function of observables, making the first stage inversion impossible.

What if there is measurement error in $l_{it}$? In this case, $l_{it}$ will vary independently of the non-
parametric function, as the measurement error moves \( l_{it} \) around independently of \( m_{it} \). However, while there is independent variation in \( l_{it} \), this independent variation is just noise that does not affect output. All the meaningful variation in \( l_{it} \) is still collinear with the non-parametric function. Because the only independent variation in \( l_{it} \) is noise, the LP first stage estimate of \( \beta_i \) will converge to zero - certainly not a consistent estimate of the labor coefficient.

Lastly, consider optimization error in \( l_{it} \). Like measurement error, this optimization error will move \( l_{it} \) around independently of the non-parametric function. However, unlike the measurement error situation, this independent variance does end up affecting output through \( \beta_i \). Hence, LP first stage estimates should correctly identify the coefficient. While this does finally give us a DGP that validates the LP first stage procedure, we feel that it is not an identification argument that empirical researchers will generally feel comfortable applying. First, in this situation, the extent of identification is completely tied to the extent of optimization error. In many situations one might feel uncomfortable basing identification entirely on the existence of optimization error. Second, note that while one needs to assume that there is enough optimization error in \( l_{it} \) to identify \( \beta_i \), one simultaneously needs to assume exactly no optimization error in \( m_{it} \). Recall, that if there were optimization error in \( m_{it} \), the inversion would not be valid. This sort of DGP assumption, i.e. that there is simultaneously lots of optimization error in one variable input yet no optimization error in the other variable input, strikes us as one that would be very hard to motivate or maintain in practice.\(^\text{12}\)

In addition to this optimization error story, there is one other DGP that can at least in theory rationalize the LP first stage procedure. Let us go back to moving around the points in time when inputs are chosen. Specifically, suppose that at time \( t - b \), intermediate input \( m_{it} \) is chosen by the firm. Subsequently, at time \( t \), labor input \( l_{it} \) is chosen. Recall from the above that this is problematic if \( \omega \) varies between these two points in time. Therefore, consider a DGP where \( \omega \) does not evolve between the points \( t - b \) and \( t \). What we do want to happen between the choice of \( m_{it} \) at \( t - b \) and the choice of \( l_{it} \) at \( t \) is some other unanticipated shock that affects a firm’s choice of \( l_{it} \). Consider, e.g., an unobserved and unanticipated shock to the price of labor that occurs between these two points in time. Call this shock \( \zeta_{it} \). Since \( \zeta_{it} \) is unanticipated and realized after the firm’s choice of \( m_{it} \), the firm’s choice of \( m_{it} \) will not depend on the shock. Hence, the first stage inversion is still valid. Because the shock occurs before the choice of \( l_{it} \), it does influence the firm’s choice of \( l_{it} \) and hence moves \( l_{it} \) around independently from the non-parametric function. As such, the existence of \( \zeta_{it} \) will break the collinearity problem and in theory will allow first stage identification of the labor coefficient. However, we again believe that this DGP is one that would be very hard to motivate in real world examples. One needs to assume that 1) firms choose \( m_{it} \) before choosing \( l_{it} \), 2) in the period of time between these choices, \( \omega_{it} \) does not evolve, and 3) in the period of time between these choices, \( \zeta_{it} \) does evolve (i.e. is realized). One additionally needs

\(^{12}\)Note that it is hard to motivate such an assumption by appealing to unions restricting the hiring and firing of labor. Such restrictions will generally affect choice of \( m_{it} \) as well as \( l_{it} \), invalidating the first stage inversion.
to assume that 4) $x_{it}$ is i.i.d. over time - otherwise, a firm’s optimal $m_{it}$ would depend on the unobserved $x_{it-1}$, violating the first stage inversion, and 5) that the unobserved $x_{it}$ varies across firms - because the non-parametric function is indexed by $t$, variation in $x_{it}$ across time is not helpful at moving around $l_{it}$ independently. This strikes us as a very particular and unintuitive set of assumptions. Not only are they unintuitive, but the assumptions also seem asymmetric in somewhat arbitrary ways - one unobservable is allowed to be correlated across time while the other is not, there must exist a period of time during which one unobservable evolves but the other doesn’t, and some input prices must be constant across firms, while others must not be. It is hard for us to imagine a dataset or situation where this set of identification assumptions would hold, even to an approximation.

To summarize, there appears to be only two potential DGPs that save the LP procedure from collinearity problems. One requires a significant amount of optimization error in $l_{it}$, yet no optimization error in $m_{it}$. The second requires a seemingly unintuitive set of assumptions on timing and unobservables. Neither of these DGPs appear to us like particularly reasonable arguments on which to base identification. An important note is that, in practice, one probably would not observe this collinearity problem. It is very likely that estimation of (9) would produce an actual numerical estimate. Our point is that unless one believes that one or both of the above two DGPs holds (and additionally that these are the only reasons why the first stage equation is not collinear), this is simply not a consistent estimator of $\beta_t$. Another way to describe this result is that unless one believes in one of the above two DGPs, the extent to which the LP first stage is identified is also the extent to which is misspecified.

3.2 Olley and Pakes

Now consider the OP model. Given the above results regarding the LP procedure, a reasonable question is whether the OP model also suffers from a similar collinearity problem. While we show there are similar collinearity issues with the OP model, we argue that this collinearity can be "broken" under what may be more reasonable assumptions than in LP.

In OP, the question is whether $l_{it}$ is collinear with the non-parametric function $f_t^{-1}(i_{it}, k_{it})$. Again, the most obvious formulation of labor input demand is that $l_{it}$ is just a function of $\omega_{it}$ and $k_{it}$, i.e. $l_{it} = g_t(\omega_{it}, k_{it})$. If this is the case, it is easy to show that we again have a collinearity problem. To obtain identification, one again needs a DGP in which something moves $l_{it}$ around independently of $f_t^{-1}(i_{it}, k_{it})$. Two possibilities are analogous to the two DGPs we just described in the LP model - i.e. either optimization error in $l_{it}$ (with no optimization error in $i_{it}$), or i.i.d., firm-specific, shocks to the price of labor (or other relevant variables) that are realized between the points in time at which $i_{it-1}$ is chosen and $l_{it}$ is chosen. However, as we have just argued, these

$^{13}$Analogously, one might regress $l_{it}$ on $k_{it}$ and $m_{it}$ and not find a perfect fit. In the context, our point would be that according to the LP assumptions, there is no really believable DGP why one wouldn’t get a perfect fit in such a regression.
DGPs seem to rely on very strong and unintuitive assumptions.

Fortunately, in the OP context, there is an alternative DGP which breaks the collinearity problem and is simultaneously consistent with the assumptions of the model. In contrast to the prior two DGPs, we feel that this DGP might be a reasonable approximation to the true underlying process in many datasets. Consider the case where $l_{it}$ is actually not a perfectly variable input, and is chosen at some point in time between periods $t - 1$ and $t$. Similar to above, denote this point in time as $t - b$, where $0 < b < 1$. Suppose that $\omega$ evolves between the subperiods $t - 1$, $t - b$, and $t$ according to a first-order markov process, as in eqs (14) and (15).

In this case, a firm’s optimal labor input will not be a function of $\omega_{it}$, but of $\omega_{it-b}$, i.e.

$$l_{it} = f_t(\omega_{it-b}, k_{it})$$

Since $\omega_{it-b}$ cannot generally be written as a function of $k_{it}$, and $i_{it}, l_{it}$ will not generally be collinear with the non-parametric term in (5), allowing the equation to be identified. Note the intuition behind this - the fact that labor is set before production means that labor is determined by $\omega_{it-b}$ rather than $\omega_{it}$. The movement of $\omega$ between $t - b$ and $t$ is what breaks the collinearity problem between $l_{it}$ and the non-parametric function. Put another way, the idea here is that labor is chosen without perfect information about what $\omega_{it}$ is, and this incomplete information is what moves $l_{it}$ independently of the non-parametric function.

To us, this DGP seems like something that could be motivated in some empirical situations. One would need to argue that labor is not a perfectly variable input, and hence is set as function of a different information set than is $i_{it}$. However, note that this DGP does need to rule out a firm’s choice of $l_{it}$ having dynamic implications. If labor did have dynamic effects, then $l_{it}$ would directly impact a firm’s choice of $i_{it}$. As a result, $l_{it}$ would directly enter the first stage non-parametric function and prevent identification of $\beta_t$.

Lastly, note why this DGP does not solve the collinearity problem in the context of the LP model. In the LP model, if $l_{it}$ is chosen before $m_{it}$, then $m_{it}$ will directly depend on $l_{it}$, making $\beta_t$ unidentified in the first stage. In OP, even if $l_{it}$ is chosen before $i_{it}$, $i_{it}$ does not depend on $l_{it}$ (as long as one maintains the assumption that labor has no dynamic implications). This is because $i_{it}$, unlike $m_{it}$, is not directly linked to period $t$ outcomes, and thus $l_{it}$ will not affect a firm’s optimal choice of $i_{it}$. The fact that this type of DGP does work in the OP context but does not work in the LP context is the reason that we describe the collinearity problem as being worse for the LP methodology.
4 Parametric Versions of LP?

The collinearity problem in LP is that in the first stage equation,

\[ y_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + f_t^{-1}(m_{it}, k_{it}) + \epsilon_{it} \]  

the non-parametric function \( f_t^{-1}(m_{it}, k_{it}) \) will generally be collinear with \( l_{it} \) under the maintained assumptions of the model. One approach to solving this collinearity problem might be to treat \( f_t^{-1}(m_{it}, k_{it}) \) parametrically. Note that even though \( l_{it} \) might again just be a function of \( m_{it} \) and \( k_{it} \), if it is a different function of \( m_{it} \) and \( k_{it} \) than \( f_t^{-1} \) is, this parametric version is potentially identified. While using a parametric version makes more assumptions than the non-parametric approach, one might be willing to make such assumptions with relatively uncomplicated input choices such as materials.

Unfortunately, this parametric approach does not work, at least for some popular production functions. In the case of Cobb-Douglas, the first order condition for \( m_{it} \) (conditional on \( k_{it} \), \( l_{it} \), and \( \omega_{it} \)) is:

\[ \beta_m K_{it}^{\beta_k} L_{it}^{\beta_l} M_{it}^{\beta_m-1} \epsilon_{\omega_{it}} = \frac{p_m}{p_y} \]

assuming firms are price takers in both input and output markets. Recall that capital letters represent levels (rather than logs) of the inputs. Inverting this out for \( \omega_{it} \) gives:

\[ \epsilon_{\omega_{it}} = \frac{1}{\beta_m} \frac{p_m}{p_y} K_{it}^{-\beta_k} L_{it}^{-\beta_l} M_{it}^{1-\beta_m} \]

\[ \omega_{it} = \ln\left(\frac{1}{\beta_m}\right) + \ln\left(\frac{p_m}{p_y}\right) - \beta_1 k_{it} - \beta_2 l_{it} + (1 - \beta_m)m_{it} \]

and plugging this inversion into the production function results in:

\[ y_{it} = \ln\left(\frac{1}{\beta_m}\right) + \ln\left(\frac{p_m}{p_y}\right) + m_{it} + \epsilon_{it} \]  

The key point here is that \( \beta_1 \) has dropped out of the estimating equation, making a moment condition in \( \epsilon_{it} \) unhelpful in identifying \( \beta_1 \). As such, with a Cobb-Douglas production function, a parametric approach cannot generally be used as a first stage to identify \( \beta_1 \).\(^{14}\)

One gets a similar result with a production function that is Leontief in the material inputs. Consider, for example:

\[ Y_{it} = \min \left[ \gamma_0 + \gamma_1 M_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} \epsilon_{\omega_{it}} \right] + \epsilon_{it} \]

\(^{14}\)The above analysis uses the choice of \( m_{it} \) conditional on levels of \( k_{it} \), \( l_{it} \), and \( \omega_{it} \). This is most naturally interpreted in the case where \( l_{it} \) is chosen before \( m_{it} \). One obtains the same result if one solves for simultaneous \( m_{it} \) and \( l_{it} \) choices conditional on levels of \( k_{it} \) and \( \omega_{it} \).
With this production function, the first order condition for $M_{it}$ satisfies

$$\gamma_0 + \gamma_1 M_{it} = K_{it}^{\beta_k} L_{it}^{\beta_l} e^{\omega_{it}}$$

as long as $\gamma_1 p_y > p_m$. At this optimum, note that:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \epsilon_{it}$$

which could form an estimating equation if not for endogeneity problems. Inverting out $\omega_{it}$ results in:

$$e^{\omega_{it}} = \frac{\gamma_0 + \gamma_1 M_{it}}{K_{it}^{\beta_k} L_{it}^{\beta_l}}$$

$$\omega_{it} = \ln(\gamma_0 + \gamma_1 M_{it}) - \beta_k k_{it} - \beta_l l_{it}$$

and substituting into (18) results in

$$y_{it} = \ln(\gamma_0 + \gamma_1 M_{it}) + \epsilon_{it}$$

so again, this procedure is not helpful for identifying $\beta_l$.

In summary, even with parametric assumptions, there may be an identification problem in the first stage of the LP technique using intermediate inputs to control for unobserved factors of production. However, it is possible that as one moves away from Cobb-Douglas production functions (or Hicks neutral unobservables or perfectly competitive output and input markets), a parametric approach might be identified (see Van Biesebroeck (2003) for a related example).

5 Our Alternative Procedure

We now suggest an alternative estimation procedure that avoids the collinearity problems discussed above. This procedure draws on aspects of both the OP and LP procedures and is able to use either the ‘intermediate input as proxy’ idea of LP, or the ‘investment as proxy’ idea of OP. The main difference between this new approach and OP and LP is that in the new approach, no coefficients will be estimated in the first stage of estimation. Instead, the input coefficients are all estimated in the second stage. However, as we shall see, the first stage will still be important to net out the untransmitted error $\epsilon_{it}$ from the production function. We exhibit our approaches using value added production functions. They could also be used in the case of gross output production functions, although in this case one might need to consider issues brought up by Bond.
and Söderbom (2005). We start by showing how our method works with the LP intermediate input proxy. We then show how our method is consistent if labor has dynamic implications and illustrate how our procedure works using the OP investment proxy. Lastly, we compare our procedure to methods used in the dynamic panel data literature, e.g. Arellano and Bover (1995), and Blundell and Bond (1998, 2000). We feel that this is important because up to now, these two literatures (OP, LP vs. dynamic panel methods) have evolved somewhat separately. Our estimation procedure makes it quite easy to see the tradeoffs and different assumptions behind the two approaches.

5.1 The Basic Procedure

Consider the following value added production function,

\[ y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \epsilon_{it} \]  

Our basic idea is quite simple - to give up on trying to estimate \( \beta_l \) in the first stage. However, we will still estimate a first stage - the goal of this first stage will be to separate \( \omega_{it} \) from \( \epsilon_{it} \). As we will see momentarily, this will be a key step in allowing us to treat the \( \omega_{it} \) process non-parametrically.

Perhaps the most intuitive way to "give up" estimation of \( \beta_l \) in the first stage is to allow for labor inputs to be chosen before material inputs. More precisely, suppose that \( l_{it} \) is chosen by firms at time \( t - b \) (\( 0 < b < 1 \)), after \( k_{it} \) was chosen at (or before) \( t - 1 \) but prior to \( m_{it} \) being chosen at \( t \). Suppose that \( \omega_{it} \) evolves according to a first order markov process between these subperiods \( t - 1, t - b, \) and \( t \), i.e.

\[ p(\omega_{it} | I_{it-b}) = p(\omega_{it} | I_{it-b}) \]

and

\[ p(\omega_{it-b} | I_{it-1}) = p(\omega_{it-b} | I_{it-1}) \]

Our feeling is that this assumption that labor is "less variable" than materials may make sense in many industries. For example, it is consistent with firms needing time to train new workers, or needing to give workers some period of notice before firing. Given these timing assumptions, a firm's material input demand at \( t \) will now directly depend on the \( l_{it} \) chosen prior to it, i.e.

\[ m_{it} = f_t (\omega_{it}, k_{it}, l_{it}) \]

\[ ^{15} \text{Bond and Söderbom (2005) argue that it may be hard (if not impossible) to identify coefficients on perfectly variable (and non-dynamic) inputs in a Cobb-Douglas framework. Note that this is also an critique of the original LP procedure's identification of the materials coefficient. On the other hand, value-added production functions have their own issues, see, e.g. Basu and Fernald (1997).} \]
Inverting this function for $\omega_{it}$ and substituting into the production function results in a first stage equation of the form:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + f_t^{-1}(m_{it}, k_{it}, l_{it}) + \epsilon_{it}$$

(22)

$\beta_i$ is clearly not identified in this first stage. However, one does obtain an estimate, $\hat{\Phi}_{it}$, of the composite term,

$$\Phi_t(m_{it}, k_{it}, l_{it}) = \beta_k k_{it} + \beta_l l_{it} + f_t^{-1}(m_{it}, k_{it}, l_{it})$$

which represents output net of the untransmitted shock $\epsilon_{it}$. Intuitively, by conditioning on a firm’s choice of material inputs (or analogously in this case conditioning on the information set at $t$), this procedure allows us to isolate and eliminate the portion of output determined by either shocks unanticipated at $t$ (e.g. unanticipated weather shocks, defect rates, or machine breakdown) or by measurement error.

However, with no coefficients obtained in the first stage, we still need to identify $\beta_k$ and $\beta_l$. This now requires two independent moment conditions for identification in the second stage. Given the first-order Markov assumption on $\omega_{it}$, we have

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

where $\xi_{it}$ is mean independent of all information known at $t - 1$. Given the OP/LP timing assumption that $k_{it}$ was decided at $t - 1$ (and hence $k_{it} \in I_{it-1}$), this leads to the second stage moment condition used by both OP and LP, namely that:

$$E[\xi_{it}|k_{it}] = 0$$

(23)

Of course, this moment would not hold if one replaced $k_{it}$ with $l_{it}$. Since $l_{it}$ is chosen after $t$, at time $t - b$, $l_{it}$ will generally be correlated with at least part of $\xi_{it}$. On the other hand, lagged labor, $l_{it-1}$, was chosen at time $t - b - 1$. Hence, it is in the information set $I_{it-1}$ and will be uncorrelated with $\xi_{it}$. This implies

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

(24)

which in turn implies that

$$E[\xi_{it} \cdot \begin{pmatrix} k_{it} \\ l_{it-1} \end{pmatrix}] = 0$$

These are the two moments we suggest using in estimation to identify $\beta_k$ and $\beta_l$.

Operationalizing this moment is analogous to the second stage of the OP and LP procedures.
We can recover the implied $\xi_{it}$’s for any value of the parameters $(\beta_k, \beta_l)$ as follows. First, given a candidate value of $(\beta_k, \beta_l)$, compute the implied $\omega_{it}(\beta_k, \beta_l)$’s $\forall t$ using the formula:

$$\omega_{it}(\beta_k, \beta_l) = \tilde{\Phi}_{it} - \beta_k k_{it} - \beta_l l_{it}$$

Second, non-parametrically regress $\omega_{it}(\beta_k, \beta_l)$ on $\omega_{i,t-1}(\beta_k, \beta_l)$ (and a constant term) - the residuals from this regression are the implied $\xi_{it}(\beta_k, \beta_l)$’s. Given these implied $\xi_{it}(\beta_k, \beta_l)$’s, one can form a sample analogue to the above moment, i.e.

$$(25) \quad \frac{1}{T} \frac{1}{N} \sum_t \sum_i \xi_{it}(\beta_k, \beta_l) \cdot \begin{pmatrix} k_{it} \\ l_{it-1} \end{pmatrix}$$

and estimate $(\beta_k, \beta_l)$ by minimizing this sample analogue.

We end this section with several important observations regarding our suggested procedure. First, the moment condition we use to identify the labor coefficient, i.e. $E[\xi_{it} \cdot l_{it-1}] = 0$ is actually used by LP (and OP in a more informal way) as an overidentifying restriction on the model in their second stage procedure. However, there is a fundamental difference between what we are doing and what OP/LP do - in OP/LP, the labor coefficient is estimated in the first stage without using any of the information from the second element of (25). In our procedure, the information in (25) is crucial in identifying the labor coefficient. Given the problems with the LP first stage identification of $\beta_l$ described above, we prefer our method of identification.

Second, it is important to note that our procedure is completely consistent with labor choices having dynamic implications. This would probably be the case if, e.g. there were firing, hiring or training costs of labor. Note that in this case, firms’ optimal choices of $l_{it}$ and $k_{it}$ will depend on $l_{it-1}$, but the intermediate input demand function $m_{it}$ will not. This is because choice of $m_{it}$ already depends on $l_{it}$ (since $l_{it}$ was chosen before $m_{it}$), and because $m_{it}$ is only relevant for period $t$ production. We feel that this is important not only for robustness reasons, but also because the additional variation in $l_{it}$ generated by dynamic issues will likely improve identification (see Bond and Söderbom (2005)).

Third, our procedure is also consistent with other unobservables, e.g. input price shocks or dynamic adjustment costs affecting firm’s choices of $l_{it}$ and $k_{it}$. Importantly, these other unobservables can be correlated across time - this is because 1) $m_{it}$ depends directly on $l_{it}$ and $k_{it}$ and

\footnote{It is possible that such an overidentifying restriction test might alert one to a spuriously identified first stage labor coefficient. However, it seems presumptuous to rely on such a test given that it is not clear how much power it has. It may in fact be a very weak test.}

\footnote{If one wanted to assume that $b = 0$, i.e. that $l_{it}$ is chosen at the same time as $m_{it}$, then one would want to replace $l_{it}$ with $l_{it-1}$ in the first stage non-parametric function.}

\footnote{While it may not be as empirically relevant, our procedure can also be extended to allow dynamics in $m_{it}$ - this could be accomplished by adding $m_{i,t-1}$ into the first stage non-parametric function. However, in this case, one would probably need to rule out the possible additional unobservables discussed in the next paragraph.}
2) because \( m_{it} \) only affects current production. Point 2) implies that even serially correlated such unobservables will not influence a firm’s optimal choice of \( m_{it} \). Again, such unobservables would likely actually be helpful for identification by generating extra exogenous variation in \( k_{it} \) and \( l_{it} \). Note that one cannot allow other unobservables to directly affect a firm’s optimal choice of \( m_{it} \) - this would violate the scalar unobservable assumption necessary for the inversion.

Fourth, in some situations one might feel comfortable assuming that \( l_{it} \) was chosen at or prior to \( t - 1 \). This might be the case if the time period in a particular dataset is short, or if, e.g. there is a significant amount of training required before workers can enter production. If this is the case, one can alternatively use the moment conditions

\[
E[\xi_{it} | k_{it} l_{it}] = 0
\]

This is likely to generate more efficient estimates than the moment condition using \( l_{it-1} \), as \( l_{it} \) is more directly linked to current output. Note that one could add additional lags of capital and labor to either set of moments (24) or (26) to generate overidentifying restrictions, although it is unclear how much extra identifying power these additional moments add.

Fifth, as is the case with OP and LP, the above can be generalized to production functions other than Cobb-Douglas. What is necessary is that it can be written as \( y_{it} = h(k_{it}, l_{it}, \omega_{it} + \epsilon_{it}; \beta) \) where \( h \) is strictly monotonic in the combined unobservable term \( \omega_{it} + \epsilon_{it} \). In this case, the first stage involves a moment in the term \( \epsilon_{it} = h^{-1}(k_{it}, l_{it}, y_{it}; \beta) - f_t^{-1}(m_{it}, k_{it}, l_{it}) \). As above, the non-parametric treatment of \( f_t^{-1} \) will tend to make the production function parameters \( \beta \) not identified by this moment. However, one will get estimates of the unobservables \( \epsilon_{it} \) - denote them \( \hat{\epsilon}_{it} \). Then the second stage can proceed using the inversion \( \omega_{it} = h^{-1}(k_{it}, l_{it}, y_{it}; \beta) - \hat{\epsilon}_{it} \). This allows, conditional on parameters \( \beta \), one to regress \( \omega_{it}(\beta) \) on \( \omega_{it-1}(\beta) \) and form a moment in the residual \( \xi_{it} \) analagous to (26). This permits one to be as flexible as one wishes in terms of the production function or value added production function \( h \), although one also needs to be sure that, given an \( h \), the strict monotonicity condition holds on \( f_t \), the input demand function that is being inverted.

Lastly, note that with the above two-stage procedure, it is probably most straightforward

\[ \text{An interesting case occurs when there are no dynamic effects of labor and no other unobservables affecting labor and/or capital. Suppose also that firms are price takers, are risk-neutral and choose labor optimally given a Cobb-Douglas production function (given the risk neutrality, firms choose labor as a function of the expectation of } \omega_{it} \text{ at } t - b, \text{i.e. } E[\omega_{it}|\omega_{it-b}]}. \] This result is related to, but distinctly different from, the complete non-identification result in Bond and Söderbom (2005), which assumes that \( b = 0 \). Monte-carlo results when \( b > 0 \) are at least suggestive that the model is identified away from the above boundary point. However, identification based on dynamic effects of labor or other unobservables seems preferable.

\[ \text{If one had multiple intermediate inputs and conditioned on all these inputs along with making an appropriate multivariate invertibility assumption, one might be able to allow a limited number of such unobservables. The key is whether one can still recover } \omega_{it} \text{ as a function of the intermediate inputs, } l_{it}, \text{ and } k_{it}.} \]
to derive asymptotic standard errors as done in LP, by bootstrapping. As mentioned above, Wooldridge (2005) suggests an alternative implementation of OP/LP that involves estimating the first and second stages simultaneously. This can easily be extended to our methodology by simply adding \( l_{it} \) to the first stage non-parametric function. This leads to the following two moments:

\[
(27) 
E \left[ \begin{array}{c}
\xi_{it} \\
\varepsilon_{it} \\
I_{it} \\
I_{it-1}
\end{array} \right] = E \left[ \begin{array}{c}
y_{it} - \beta_k k_{it} - \beta_l l_{it} - f^{-1} \left( m_{it}, k_{it}, l_{it} \beta_{f,t} \right) \\
f^{-1} \left( m_{it-1}, k_{it-1}, l_{it-1} \beta_{f,t-1} \right) - g \left( f^{-1} \left( m_{it-1}, k_{it-1}, l_{it-1} \beta_{f,t-1} \right); \beta_g \right) | I_{it-1}
\end{array} \right] = 0
\]

where \( f \) and \( g \) are, e.g., polynomial functions with parameters \( \beta_{f,t} \) and \( \beta_g \). The two moments correspond, respectively, to our first and second stages. Note the different conditioning sets, as \( \xi_{it} \) will generally be correlated with \( I_{it} \). In practice, one would likely want to use \( k_{it}, l_{it} \) and a set of appropriate (e.g. polynomial) basis functions of the arguments of \( f^{-1} \) interacted with time dummies as instruments for the first moment \( (\varepsilon_{it}) \), and \( k_{it}, l_{it-1} \) (or \( l_{it} \) depending on one’s timing assumptions) and basis functions of the scalar \( f^{-1} \left( m_{it-1}, k_{it-1}, l_{it-1} \beta_{f,t} \right) \) as instruments for the second moment \( (\xi_{it}) \). Note that the polynomial basis functions of \( f^{-1} \left( m_{it-1}, k_{it-1}, l_{it-1} \beta_{f,t-1} \right) \) will depend on the parameters \( \beta_{f,t-1} \). The sample analogue of this set of moments can be minimized w.r.t. \( (\beta_k, \beta_l, \beta_{f,t}, \beta_g) \) to generate consistent estimates of these parameters.²¹

An important advantage of applying the Wooldridge one-step approach to our estimating equations is that standard errors can be computed using standard GMM formulas. Another potential advantage is efficiency. One limitation is that it requires a non-analytic search over a much larger set of parameters, \( (\beta_k, \beta_l, \beta_{f,t}, \beta_g) \). The dimension of \( \beta_g \) is the dimension of the polynomial used to represent \( g \), and the dimension of \( \beta_{f,t} \) is the dimension of the polynomial used to represent \( f \) times the number of time periods. In the two-stage approach, one only has to search over the two production function parameters \( \beta_k \) and \( \beta_l \) - the parameters of the polynomials are all analytically computable. To avoid optimization problems, we suggest using parameters from the two-step procedure as starting values if using the one-step approach. An even more reliable alternative might be to take one Newton-Raphson step with the one-step objective function using two-step estimates as starting parameters. This requires no additional optimization, and based on a result by Pagan (1986) produces estimates asymptotically equivalent to maximizing the one-stage objective function. As such, one can use the simpler method to compute asymptotic standard errors from the one stage approach. That said, bootstrapping standard errors is fairly straightforward if one prefers the simpler 2-step approach.

²¹Given our arguments questioning the LP first stage identification, Wooldridge (2005) (pp. 12) suggests the alternative possibility of dropping the first moment condition (i.e. the moment \( E[\varepsilon_{it} | I_{it}] \)) and only using a moment in \( \xi_{it} + \varepsilon_{it} \), i.e. \( E \left[ y_{it} - \beta_k k_{it} - \beta_l l_{it} - g \left( f^{-1} \left( m_{it-1}, k_{it-1}, l_{it-1} \beta_{f,t-1} \right); \beta_g \right) | I_{it-1} \right] \) to identify the parameters. However, it is hard to see how just using this one equation could well identify the non-parametric function \( f^{-1} \) and \( \beta_l \) simultaneously (since if one uses all functions of \( l_{it-1} \) as instruments to identify the non-parametric \( f^{-1} \), one cannot use \( l_{it-1} \) as an instrument for \( l_{it} \)). It also would not separately identify \( g \) and \( f^{-1} \) (nor the productivity shocks \( \omega_{it} \), which are often objects of interest. In our opinion, the \( E[\varepsilon_{it} | I_{it}] = 0 \) moment should definitely be used, as it should provide a great deal of information on the \( f^{-1} \) function.
5.2 Investment Proxy

One can also use our methodology with the investment proxy variable of OP. Interestingly, moving all identification to the second stage simultaneously makes the procedure robust to dynamic effects of labor.\textsuperscript{22} Suppose that, as in OP, $i_{it-1}$ (and thus $k_{it}$) is chosen exactly at $t-1$ (unlike when using an intermediate input proxy, where $k_{it}$ can be chosen at or before $t$, this assumption is necessary for $i_{it}$ to "invert out" the correct $\omega_{it}$). As above, suppose that $l_{it}$ is chosen at time $t-b$, and allow there to be possible dynamic effects of labor. In this case, a firm’s optimal investment decision will generally take the form

$$i_{it} = f_{t}(\omega_{it}, k_{it}, l_{it})$$

since $l_{it}$ is chosen before $i_{it}$ and because $l_{it}$ has possible dynamic implications. Inverting this function and substituting into the production function results in:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + f_{t}^{-1}(i_{it}, k_{it}, l_{it}) + \epsilon_{it}$$

which again clearly does not identify any coefficients in the first stage. However, one can again use the first stage to estimate the composite term

$$\Phi_{t}(i_{it}, k_{it}, l_{it}) = \beta_k k_{it} + \beta_l l_{it} + f_{t}^{-1}(i_{it}, k_{it}, l_{it})$$

and proceed exactly as above, using the estimated $\Phi_{t}$’s to infer $\omega_{it}(\beta_k, \beta_l)$’s, $\xi_{it}(\beta_k, \beta_l)$’s, and form the moment (25).

Like our procedure using the intermediate input proxy, this procedure is consistent with labor having dynamic effects. However, unlike the above, it is not generally consistent with other, serially correlated unobservables entering either the $i_{it}$ or $l_{it}$ decisions. Another unobservable affecting the $i_{it}$ equation is clearly problematic for the inversion. Less obviously, another serially correlated unobservable that affects the $l_{it}$ decision will generally also affect the $i_{it}$ decision directly since $i_{it}$ is a dynamic decision variable. As a result, the inversion is problematic. The reason the intermediate input proxy is more robust to these additional serially correlated unobservables is because intermediate inputs are only relevant for current output.

5.3 Relation to Dynamic Panel Models

Interestingly, the form of our suggested estimators make them fairly easy to compare to estimators used in an alternative literature, the dynamic panel literature. This is important because up to now, researchers interested in estimating production functions have essentially been choosing between the OP/LP general approach versus the dynamic panel approach without a clear description

\textsuperscript{22}Buettner (2005) also makes this suggestion for extending OP to allow dynamic effects of labor.
of the similarities and differences of the identifying assumptions used in the two methods. We start with a brief discussion of dynamic panel methods before comparing them to our estimator. To briefly summarize, there are distinct advantages and disadvantages of both approaches.

As developed by work such as Chamberlain (1982), Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998, 2000), the dynamic panel literature essentially extends the fixed effects literature to allow for more sophisticated error structures. Consider the following production function model:

\begin{equation}
    y_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \alpha_{it} + \epsilon_{it}
\end{equation}

Whereas the standard fixed effects estimator necessarily assumes that \( \alpha_{it} \) is constant over time, the dynamic panel literature can allow more complex error structures. For example, suppose that \( \alpha_{it} \) is composed of both a fixed effect (\( \alpha_i \)) and a serially correlated unobservable (\( \omega_{it} \)), i.e.

\begin{align}
    y_{it} &= \beta_1 k_{it} + \beta_2 l_{it} + \alpha_i + \omega_{it} + \epsilon_{it} \\
    &= \beta_1 k_{it} + \beta_2 l_{it} + \psi_{it}
\end{align}

Notationally, the composite error term \( \psi_{it} \) represents the sum of all three error term components.

The dynamic panel literature proceeds by first making assumptions on 1) the evolution of the error components \( \alpha_i, \omega_{it}, \) and \( \epsilon_{it} \), and 2) possible correlations between these error components and the explanatory variables \( k_{it} \) and \( l_{it} \). Given these assumptions, the key is then to find functions of the aggregate error terms \( \psi_{it} \) (often these functions involve differencing the \( \psi_{it} \)'s) that are uncorrelated with past, present, or future values of the explanatory variables. Since the \( \psi_{it} \)'s are "observable" given particular values of the parameters (unlike the individual components of \( \psi_{it} \)), one can easily set up sample analogues of these moment conditions.

Continuing with the above production function example, a reasonable set of assumptions on the error components might be as follows. First, one might allow for potential correlation between the time-invariant error component \( \alpha_i \) and \( k_{it} \) and \( l_{it} \). Second, one could assume that \( \epsilon_{it} \) is i.i.d. over time and uncorrelated with \( k_{it} \) and \( l_{it} \) for all \( t \) (e.g. \( \epsilon_{it} \) might represent measurement error or unanticipated shocks to \( y_{it} \)). Lastly, one could assume that \( \omega_{it} \) follows an AR(1) process, i.e. \( \omega_{it} = \rho \omega_{it-1} + \xi_{it} \). Regarding correlation between \( \omega_{it} \) and the inputs, one might allow that \( \omega_{it} \) is correlated with \( k_{it} \) and \( l_{it} \) \( \forall t \) but assume that the innovation in \( \omega_{it} \) between \( t - 1 \) and \( t \), i.e. \( \xi_{it} \), is uncorrelated with all input choices prior to \( t \). Note that the intuition behind this assumption is similar to that behind the second stage moments in our procedure (and OP/LP). This idea is that since the innovation in \( \omega_{it}, \xi_{it} \) occurs after time \( t - 1 \), it may not be correlated with inputs dated \( t - 1 \) and earlier.\(^{23}\)

\(^{23}\) As with the analogous assumption in the OP/LP/ACF models, this assumption is not just an assumption on the time series properties of \( \nu_{it} \) - it is also an assumption on the information sets of firms (i.e. that firms do not observe \( \xi_{it} \)'s until they occur).
Given these particular assumptions, estimation can proceed as follows. Consider the following function of $\psi_{it}$,

$$(\psi_{it} - \rho \psi_{i(t-1)}) - (\psi_{i(t-1)} - \rho \psi_{i(t-2)}) = \xi_{it} - \xi_{i(t-1)} + (\epsilon_{it} - \rho \epsilon_{i(t-1)}) - (\epsilon_{i(t-1)} - \rho \epsilon_{i(t-2)})$$

The equality follows from the definitions of $\psi_{it}$ and $\omega_{it}$. Note that only $\epsilon_{it}$’s and innovations in the AR(1) process enter this expression - all terms containing $\alpha_t$ have been differenced out. Now, since the innovations $\xi_{it}$ and $\xi_{i(t-1)}$ have been assumed uncorrelated with all input choices prior to $t-1$ (and $\epsilon_{it}$ have been assumed uncorrelated with all input choices), we can easily form a method of moments estimator for $\beta$ and $\rho$. By assumption the moment

$$(33) \quad E \left[ (\psi_{it} - \rho \psi_{i(t-1)}) - (\psi_{i(t-1)} - \rho \psi_{i(t-2)}) \mid \left\{ \frac{k_{it}}{l_{it}} \right\}_{t=1}^{t-2} \right]$$

is equal to zero. A sample analogue of this moment is trivial to construct, since given values of the parameters, all $\psi_{it}$’s (and thus all $(\psi_{it} - \rho \psi_{i(t-1)}) - (\psi_{i(t-1)} - \rho \psi_{i(t-2)})$’s) are "observed".

Before continuing, note that this estimation procedure can be adapted in various dimensions. For example, suppose one is unwilling to assume that the $\epsilon_{it}$ are uncorrelated with all inputs in all time periods, but prefers making the weaker assumption that $\epsilon_{it}$ is sequentially exogenous, i.e. uncorrelated with all input choices dated prior to $t$. In this case, the above moment is not equal to zero, as there is potentially correlation between $\epsilon_{it-2}$ and $(k_{i(t-2)}, l_{i(t-2)})$. However, the moment still holds for lagged inputs prior to $t-2$, so the alternative moment

$$E \left[ (\psi_{it} - \rho \psi_{i(t-1)}) - (\psi_{i(t-1)} - \rho \psi_{i(t-2)}) \mid \left\{ \frac{k_{it}}{l_{it}} \right\}_{t=1}^{t-3} \right]$$

could be used for estimation.

As another example, suppose we remove the fixed effect from the model, i.e.

$$(34) \quad y_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \psi_{it}$$

but keep the same assumptions on $\omega_{it}$ and $\epsilon_{it}$. In this case, one only needs to difference once to form a usable moment. More specifically, since

$$\psi_{it} - \rho \psi_{i(t-1)} = \xi_{it} + (\epsilon_{it} - \rho \epsilon_{i(t-1)})$$
we can use

\[
E \left[ \psi_{it} - \rho \psi_{it-1} \mid \left\{ k_{it} \right\}_{\tau=1}^{t-1} \right]
\]  

(35)

as a moment for estimation if the \( \epsilon_{it} \) are strictly exogenous (\( k_{it-1} \) and \( l_{it-1} \) could not be used as instruments if the \( \epsilon_{it} \) were assumed sequentially exogenous).

This last example is particularly relevant for our current goals since the model (34) is very similar to our OP/LP style model and makes comparison quite easy. Note the differences in the construction of the second stage moments in the dynamic panel model versus those used in the second stage of our suggested procedure. In our procedure, the first stage serves to net out the \( \epsilon_{it} \). After this is done, we can compute \( \omega_{it} \forall t \) (conditional on parameters) and form moments in the innovations in \( \omega_{it} \). This contrasts with the dynamic panel approach, where conditional on the parameters, one cannot compute the individual \( \omega_{it} \)'s, but can instead only compute the sums \( \psi_{it} = \omega_{it} + \epsilon_{it} \forall t \). While in both these cases one can form moments for estimation to consistently estimate the parameters, the difference between being able to "observe" \( \omega_{it} \) versus being able to only "observe" the sum \( \omega_{it} + \epsilon_{it} \) (conditional on parameters) has a number of important implications.

First, recall that in our model, \( \omega_{it} \) can follow an arbitrary first order Markov process. This is not the case in the dynamic panel model. Not only must the Markov process generating \( \omega_{it} \) be parametric, but it must also have a linear form. In the above example, it is the linearity of the AR(1) process that allows us to construct a useable moment using the sums \( \omega_{it} + \epsilon_{it} \). To see this, suppose that instead of the Markov process being \( \omega_{it} = \rho \omega_{it-1} + \xi_{it} \), it is \( \omega_{it} = \rho \omega_{it-1}^3 + \xi_{it} \). The problem here is that it is not clear how one can manipulate the sums \( \psi_{it} \) to form a useable moment. One can construct the difference \( \psi_{it} - \rho \psi_{it-1}^3 \), i.e.

\[
\psi_{it} - \rho \psi_{it-1}^3 = \omega_{it} + \epsilon_{it} - \rho (\omega_{it-1} + \epsilon_{it-1})^3
= \omega_{it} + \epsilon_{it} - \rho (\omega_{it-1}^3 + \epsilon_{it-1}^3 + 2 \omega_{it-1} \epsilon_{it-1}^2 + 2 \omega_{it-1}^2 \epsilon_{it-1})
= \xi_{it} + \epsilon_{it} - \rho (\epsilon_{it-1}^3 + 2 \omega_{it-1} \epsilon_{it-1}^2 + 2 \omega_{it-1}^2 \epsilon_{it-1})
\]

but while one term in this expression is the innovation term \( \xi_{it} \), the expression also contains numerous other terms that are very likely correlated with the inputs.\(^{24}\) More generally, it appears that with a non-linear Markov process, it will not be possible to cleanly construct a valid moment in the innovation term using the sums \( \psi_{it} \). In contrast, in our procedure, because we are able to recover the individual \( \omega_{it} \)'s, it is trivial to deal with non-linear first order Markov processes (in this example, just regress \( \omega_{it} \) on \( \omega_{it-1}^3 \) and form moments with the residual). Not only can our

\(^{24}\)In particular, given that \( \omega_{it} \) is correlated with input choices (in all periods), it is highly likely that \( 2 \omega_{it-1} \epsilon_{it}^2 \) will be correlated with the inputs as well.
procedure deal with such non-linearities, but it also easily permits non-parametric estimation of these processes. Again, the crucial step here is the first stage estimation, which nets out the $\epsilon_{it}$ and allows us to "observe" $\omega_{it}$ conditional on the parameters. This flexibility in modelling of the $\omega_{it}$ process is a clear advantage of our procedure over dynamic panel methods.\(^{25}\)

A second difference concerns the relative efficiency of the two estimators. The variance of a GMM estimator is proportional to the variance of the moment condition being used. Suppose, for example, that we know that $\omega_{it}$ follows an AR(1) process. In this case, our second stage would involve regressing $\omega_{it}$ on just $\omega_{it-1}$ (conditional on parameters) and setting the residual orthogonal to appropriately lagged instruments. This residual is equal to the innovation in $\omega_{it}$, i.e. $\xi_{it}$, at the true parameters. In contrast, the dynamic panel approach sets the residual $\psi_{it} - p\psi_{it-1}$ orthogonal to instruments. This residual is equal to the innovation in $\omega_{it}$ plus some additional terms, i.e. $\xi_{it} + (\epsilon_{it} - p\epsilon_{it-1})$. Since these additional terms add variance to the moment condition (for a given set of instruments), this difference will tend to make our estimator asymptotically more efficient than the dynamic panel estimator.\(^{26}\) That said, our estimator requires estimation of two distinct non-parametric functions that the dynamic panel estimator does not. This difference could detrimentally impact the small sample distribution of our estimator relative to the dynamic panel estimator.

There are also significant advantages of the dynamic panel estimator over our estimator. We feel that the most important one concerns possible fixed effects. For example, the start of this section showed how dynamic panel methods can allow for a fixed effect $\alpha_i$ in addition to the serially correlated process $\omega_{it}$. The resulting estimator is consistent even for fixed $T$. This, to our knowledge, cannot be done with our estimator.\(^{27}\) On the other hand, allowing for fixed effects in the dynamic panel literature requires an additional differencing and further lagging of instruments (compare the moment in (33) to that in (35)) - this likely puts considerably greater demands on the data.\(^{28}\) Perhaps this is one reason why these estimators have sometimes not worked particularly well in practice.\(^{29}\) Regardless, the ability to allow for $\alpha_i$'s is definitely an advantage of the dynamic panel approach.

Another advantage of the dynamic panel literature is that it requires fewer assumptions regarding input demand equations. Recall that our procedure requires both a strict monotonicity and a scalar unobservable assumption on one of the input demand equations, e.g. on either investment or materials.\(^{30}\) The dynamic panel literature does not require such assumptions. Of course, it

\(^{25}\)Note that in the special case where $\epsilon_{it}$ is assumed zero for all $i$ and $t$, the dynamic panel methodology can allow a non-linear (or non-parametric) Markov process. This is because in this case (i.e. $\epsilon_{it} = 0$), the $\omega_{it}$'s are recoverable given parameters.

\(^{26}\)Formally proving this would need to account for the first stage estimation error in netting out the $\epsilon_{it}$'s.

\(^{27}\)If $T \rightarrow \infty$, we could simply estimate the fixed effects in our model, but this is a much weaker result.

\(^{28}\)Both the additional differencing and the further lagging of the instruments are likely to reduce the information in the moment condition (see, e.g., Griliches and Hausman (1981)).

\(^{29}\)Blundell and Bond (1998) suggest additional moments based on initial conditions to address this problem.

\(^{30}\)It is possible to relax the scalar unobservable assumption in some cases, but this requires multiple proxy variables (e.g. investment choice and advertising choice) and a multidimensional strict monotonicity assumption.
is these assumptions that allow us to form the first stage equation, net out the $\epsilon_{it}$’s, observe the $\omega_{it}$’s conditional on the parameters, and thus treat the $\omega_{it}$ process non-parametrically.

The dynamic panel literature also permits one to make slightly weaker assumptions on the $\epsilon_{it}$’s. As described above, dynamic panel procedures can proceed either under a strict exogeneity assumption ($\epsilon_{it}$ uncorrelated with input choices at all $t$) or a weaker sequential exogeneity assumption ($\epsilon_{it}$ uncorrelated with input choices prior to $t$). For all practical purposes, our procedure depends on the strict exogeneity assumption. The problem here does not regard the second stage moments ($\epsilon_{it}$ does not even enter the second stage moments) - it is with the first stage. The problem is that sequential exogeneity permits $\epsilon_{it}$ to affect future input choices. This will tend to violate the scalar unobservable assumption necessary for the first stage of our procedure. Note that both types of procedures can allow $\epsilon_{it}$ to be correlated over time, at least in some cases. The key assumption in both is that $\epsilon_{it}$ is not in any way predictable by firms. For example, $\epsilon_{it}$ could contain measurement error in $y_{it}$ that is serially correlated over time. Another seeming advantage of the dynamic panel literature is that it can allow for a higher than first order Markov process for $\omega_{it}$, as long as this process is linear (e.g. an AR(2) process - note that this would require further differencing to construct a valid moment). However, ABBP show that our methods can be extended to non-parametric higher order Markov process if one observes a set of control variables equal to the order of the Markov process.

Lastly, there are some differences in how these two types of estimators have been used in practice, but that are less fundamental than the differences above. For example, a frequent assumption in the literature applying OP/LP methods has been that $k_{it}$ is part of $I_{it-1}$. This generates orthogonality between $\xi_{it}$ and $k_{it}$, which is likely a more informative moment than orthogonality between $\xi_{it}$ and $k_{it-1}$. This assumption has typically not been made in the dynamic panel literature, but it easily could be. One can simply add $k_{it-1}$ to the conditioning set in (33) or $k_{it}$ to the conditioning set in (35) (under strict exogeneity). Presumably, this would increase the efficiency of dynamic panel estimates. The same idea could be applied to other "fixed" inputs as well. Another difference between how these estimators have been applied in practice is that while the dynamic panel literature has typically utilized orthogonality between differenced residuals and all inputs suitably lagged (i.e. from $\tau = 1$ to $\tau = t - 2$ or $t - 3$), applications using OP/LP methodology have often only used the latest dated valid observation for each input as instruments (the application in LP is a notable exception). Of course, all further lagged inputs are also valid instruments in our methodology (or OP/LP) and could also be used, analogous to the dynamic panel methodology. The tradeoff is as often the case - more moments generate more efficiency and result in overidentification (which can be useful for testing purposes), but they often can also generate significant small sample biases.

(see ABBP).

31Formally, our procedure can allow $\epsilon_{it}$ to affect future choices of inputs not used for the first stage inversion (e.g. labor), but allowing $\epsilon_{it}$ to affect future labor choices but not, e.g. future material or investment choices, seems somewhat arbitrary.
In summary, while our procedure and dynamic panel methods for estimating production functions are related, there are fundamental differences between the two. While our procedure has more flexibility regarding the serially correlated transmitted error $\omega_{it}$, it is less flexible regarding the non-transmitted error $\epsilon_{it}$ and in allowing fixed effects $\alpha_i$. Our procedure also requires the additional assumptions necessary for the first stage inversion. In some cases, data considerations and/or a-priori beliefs about a particular production process may guide choices between the two approaches. In other cases, one may want to try both techniques. Finding that production function parameters are consistent across multiple techniques with different assumptions is surely more convincing than only using one.

## 6 Empirical Example

We now briefly compare our estimator to existing estimators with a commonly used dataset. Generally, we feel that in practice one should take the key timing, scalar unobservable, and strict monotonicity assumptions behind these methods quite seriously. For example, one should be relatively sure that the variable being used to "invert" out unobserved productivity, whether it be investment or an intermediate input, is well measured. In addition, one will hopefully be able to use industry sources to motivate whichever timing assumptions one chooses to make, e.g. that capital (and/or labor) is decided a full period before production. We are much more cursory in motivating these assumptions here. This is both for brevity and because our interest is not in the empirical results per-se, but in simply exhibiting that our estimator can generate reasonable results. The exact empirical results should be interpreted with this caveat.

For our example, we utilize the same Chilean plant level data as do LP. One can consult LP and the references therein for details on the dataset. We also examine the same four industries as LP - food products (ISIC code 311), Textiles (321), Wood Products (331), and Metals (381). These were chosen by LP because they contain a large number of plant-year observations. ISIC 311 has the most, with more than 5000 plant-year observations over the period 1979-1986. One key difference between our results and those exhibited in LP is that we estimate value-added production functions rather than gross-revenue production functions. There are two reasons for this. First, as noted previously, the aforementioned work by Bond and Söderbom (2005) casts some doubt on being able to reliably identify coefficients on perfectly variable inputs in Cobb-Douglas production functions without input price variation across firms. Second, estimating a gross-revenue production function requires estimating coefficients on all intermediate inputs. In this dataset this includes materials, electricity, and fuels. These variables are highly collinear with each other (and with capital and labor), and we have found it hard using any of the available techniques to generate particularly stable estimates for parameters on all these inputs.

In addition to standard OLS and fixed effects estimators, we examine the LP estimator, our
ACF estimator, and a version of the dynamic panel methodology described above. With the LP method, we use \( k_{it} \) as the second stage instrument. For ACF, our main results use \( k_{it} \) and \( l_{it} \) as second stage instruments, i.e. we use the moment (26). As such we make the timing assumption that \( l_{it} \) was decided before (or without knowledge of) the realization of \( \xi_{it} \). We have also tried our procedure under weaker timing assumptions where we use \( k_{it} \) and \( l_{it-1} \) as second stage instruments. This allows labor to be chosen with knowledge of the full \( \omega_{it} \). While the results are qualitatively similar to the main results, standard errors were generally higher. Table 1A in the appendix contains these alternative estimates. There are three intermediate inputs in the dataset - materials, electricity, and fuel. Following LP, we try using each separately as the proxy variable in the LP and ACF methods. Also following LP, we allow the inverse intermediate input demand function to vary (non-parametrically) across three macroeconomic cycles in the data (1979-81, 1982-83, 1984-86). As described above, one can generate overidentification restrictions with the LP, DP, and ACF estimators by adding further lags of the inputs to the conditioning set of the moment conditions. However, because 1) numerically it is easier to estimate exactly identified systems (particularly given that we bootstrap the standard errors), and 2) because we sometimes reject these overidentifying restrictions (LP also find this), we simply work with the exactly identified set of moments.

As discussed above, there are various sets of identifying assumptions one can make in applying the dynamic panel (DP) methodology. To make our estimates as comparable as possible, we choose these assumptions to be as similar as possible to those we are making in the ACF and LP procedures. Specifically, we assume that the composite error \( \psi_{it} \) is composed of only an AR(1) process \( (\omega_{it}) \) and an iid process \( (\epsilon_{it}) \) (as in model (34)). Although the DP literature could potentially also allow for a fixed effect \( \alpha_i \), this 1) would not be as similar to our ACF/LP assumptions, and 2) it is also considerably more demanding on the data (because it requires double differencing). We also assume for the DP estimator that \( k_{it} \) and \( l_{it} \) (\( k_{it} \) and \( l_{it-1} \) in the table in the appendix) are orthogonal to the innovation in the AR(1) process \( \xi_{it} \). Again, this is analogous to what we are assuming in the ACF/LP procedures and can be motivated by the same timing/informational assumptions. Thus, the basic moment used for DP estimation is

\[
E \left[ (\psi_{it} - \rho\psi_{it-1}) \cdot \left( \begin{array}{c} k_{it} \\ l_{it} \end{array} \right) \right] = 0
\]

where \( \psi_{it} = \omega_{it} + \epsilon_{it} \) and \( \omega_{it} = \rho \omega_{it-1} + \xi_{it} \). \(^{34}\)

\(^{32}\) Probably because it requires dropping more than 50% of the observations (there are a large number of observations with 0 investment in this developing country dataset), the OP estimator gives considerably different estimates than the ACF, LP, and DP estimators. Hence, we do not report these results.

\(^{33}\) We presume that the reason LP had problems with the DP methods (and ended up not reporting results) is because they tried estimating the more demanding versions that allow for a fixed effect \( \alpha_i \).

\(^{34}\) In the DP procedure, one needs an additional moment for estimation to identify \( \rho \). We use the moment \( E [\xi_{it} \cdot (\omega_{it-1} + \epsilon_{it-1})] = 0 \), i.e. we assume that the innovation \( \xi_{it} \) is uncorrelated with \( \omega_{it-1} \) and \( \epsilon_{it-1} \). This
Table 1 presents our main results. In the LP and ACF procedures, we use kernel estimators for the non-parametric first stages. For all estimators we block-bootstrap (at the plant level) the standard errors - this allows for correlations between the moment conditions of the same plant in different years. It also appropriately computes the LP and ACF standard errors given that two stage procedures are used in estimation. The first two rows for each industry exhibit OLS and fixed effects estimators. As typical in production datasets, the fixed effects approach generates what seem to be unrealistically low estimates of the capital coefficient and returns to scale. In industry 331, the fixed effects estimate of the capital coefficient is actually negative.

The ACF estimates seem reasonable, regardless of which intermediate input is used as the proxy. With each of the 3 proxies across all 4 industries, the estimated returns to scale using ACF are lower than the returns to scale estimated by OLS. This makes sense, as one would generally expect input choices to be positively correlated with \( \omega_{it} \), biasing the OLS estimates of returns to scale upwards. Table 2 tests whether these differences are significant. The values in the cells of the table are the proportion of bootstrap repetition in which the ACF estimate is lower than the corresponding OLS (or LP or DP) coefficient. As such, a value either higher than 0.95 or lower than 0.05 indicates that the coefficients are significantly different from each other (at 90% confidence level). For example, in Industry 321, ACF with the material proxy produces a lower returns to scale coefficient than OLS in 98.2% of the bootstrap replications - this is a significant difference. In fact, the ACF returns to scale estimate is significantly lower than the OLS estimate in all 12 specifications (4 industries with each of 3 proxy variables). Most of the differences in the estimates of returns to scale appear to be coming from the respective labor coefficients, as the ACF labor coefficient estimates are also significantly lower than their OLS analogues in all 12 specifications. On the other hand, the capital coefficients go in various directions - in some cases the ACF estimate is higher than the corresponding OLS estimate. While this movement in the capital coefficient is not necessarily an intuitive result, it is possible if labor is more "variable" than capital and as a result \( l_{it} \) is more correlated with \( \omega_{it} \) than is \( k_{it} \).

Comparing the ACF results to the LP results, a few interesting patterns arise. First, there are many significant differences in the coefficients. Of the 12 sets of estimates, 7 of the capital coefficients, 8 of the labor coefficients, and 7 of the returns to scale coefficients are significantly different between the ACF and LP specifications. In terms of the directions of the differences, they can go either way, but the LP estimates of the labor coefficients are more often smaller

35 is analogous to what is being done in the LP and ACF procedures when one regresses implied \( \omega_{it} \) on implied \( \omega_{it-1} \) to construct \( \xi_{it} \). It is also the same as using a specific function of the lagged data \( f(y_{it-1}, l_{it-1}, k_{it-1}) = y_{it-1} - \beta_l l_{it-1} - \beta_k k_{it-1} \) as an "instrument".

35 We use the "rule-of-thumb" bandwidth for the multivariate case proposed in Hardle, Muller, Sperlich, and Werwatz (2004). We suggest some care here, as in our experience these estimators can be somewhat sensitive to choice of non-parametric technique and degree of smoothing. For the second stage non-parametric regressions of \( \omega_{it} \) on \( \omega_{it-1} \) we use a 5th order polynomial instead of a kernel. This is done because the regressor is one dimensional and to save computational time since these regressions need to be run many times (for each candidate value of the 2nd stage parameters).
than their ACF counterparts. This is suggestive that the LP first stage labor coefficient estimates may be biased downward. An interesting difference between the estimators is their sensitivity to which proxy is used. The ACF estimates are fairly stable across the materials, electricity, and fuel proxies. In ISIC 311, for example, the ACF estimates of the labor coefficient only varies between 0.842 and 0.884 depending on which proxy is used. In contrast, the LP estimates vary much more across the 3 different proxies - in ISIC 311 the estimated LP labor coefficient is 0.676 when the materials proxy is used, but 0.942 when the fuel proxy is used. Again in ISIC 311, the ACF estimates of returns to scale vary from 1.212 to 1.279, while the LP estimates vary from 1.131 to 1.352. The other ISICs exhibit a similar pattern - the LP estimates generally seem much more sensitive to the particular proxy used. This instability of the LP labor coefficients seems consistent with our arguments questioning the source of identification of the LP first stage labor coefficient.\[36\]

The last row for each ISIC contains estimates using the DP methodology. The DP estimates also generally look reasonable - for example, the estimates of returns to scale are generally lower than OLS. While the DP estimates generally seem closer to the ACF estimates than do the LP estimates, there are still a number of significant differences. Of the 12 comparisons, 4 of the capital coefficients, 5 of the labor coefficients, and 9 of the returns to scale estimates are significantly different. These significant differences suggest that one of the assumptions behind the estimators may be incorrect. For example, it is possible that $\omega_{it}$ follows a 1st order Markov process that is more complicated than a AR(1) process - this would invalidate the DP estimates (but not the ACF estimates). Alternatively, perhaps the scalar unobservable and strict monotonicity assumptions behind the ACF first stage inversion are incorrect - this would invalidate the ACF estimates (but not the DP estimates). That said, while the estimates are statistically different, they are somewhat close economically, so it is possible that any economic predictions might be insensitive to which estimates are used.

Lastly, it is interesting to examine the standard errors of the various estimators. As expected, the OLS estimates have the lowest standard errors while the fixed effects estimates have the highest standard errors. Regarding the LP, DP, and ACF standard errors, it is interesting that none seem to dominate - they are all generally in the same range. That said, when using $l_{it-1}$

\[36\]There is still the question of why the LP procedure seems to consistently generate positive (and significant) labor coefficients in practice. Recall that at least in the "simplest" possible DGP process for the labor variable, the labor coefficient should not be identifiable in the first stage. In our opinion, there are a number of possible explanations for this. First, it is possible that one (or a combination) of the alternative DGP’s described in section 3.1 is occurring. For example, the combination of labor being decided at $t - b$ (as a function of $\omega_{it-1}$ rather than $\omega_{it}$) plus some optimization error in labor could generate this finding. Of course, in this case, the LP estimate is not a consistent estimate of $\beta_1$. Another possible story is that the non-parametric approximations are not working well. In general, this will generate a positive, but again spurious estimate of $\beta_1$ in the LP first stage. Lastly, one might want to consider the possibility that maybe some of the more fundamental assumptions behind both LP and ACF are wrong. For example, there could be optimization or measurement error in the proxy variables. This would almost surely generate a positive (but spurious) coefficient on labor in the LP first stage procedure. Of course, it is also likely to generate spurious coefficients in the ACF procedure.
as the instrument (Table 1A), the standard errors of the DP and ACF estimates increase as expected. It is also interesting that the LP, DP, and ACF standard errors seem closer to OLS standard errors than they do to the fixed effects standard errors. This seems to be a positive result for these methods.

7 Conclusions

This paper has examined some of the recent literature on identification of production functions (Olley and Pakes (1996) and Levinsohn and Petrin (2003)) and argues that there may be significant collinearity problems in the first stages of these methods. Given these potential collinearity problems, we search for possible data generating processes that simultaneously 1) break this collinearity problem, and 2) are consistent with the LP/OP assumptions. For LP, we conclude that there are only two such DGP’s, and that both rely on very strong and unintuitive assumptions - one involves a story where one variable input choice has a large amount of optimization error, while another variable input choice has exactly no optimization error. The second DGP involves a story where 1) intermediate inputs are chosen prior to labor, 2) that between the points in time when intermediate inputs are chosen and when labor is chosen, the firm’s productivity level does not change, 3) that between these points in time, the firm is exposed to a price or demand shock that influences its choice of labor, and 4) that this price or demand shock varies across firms and is not correlated across time. Neither of these DGP assumptions seem realistic enough (even to an approximation) to generally rely on in practice. For OP, there is an additional DGP that breaks the collinearity and is consistent with the model - this involves labor being chosen prior to production and relies on the evolution of productivity between the time when labor is chosen and when production takes place to break the collinearity. This DGP seems more realistic to us than those needed validate the LP procedure.

We then suggest a new approach for estimating production functions. This approach builds upon the ideas in OP and LP, e.g. using investment or intermediate inputs to "proxy" for productivity shocks, but does not suffer from the above collinearity problems. The key difference is that unlike the OP and LP procedures, which estimate the labor coefficient in the first stage (where the collinearity issue arises), our estimator involves estimating the labor coefficient in the second stage. Even though no parameters are identified in our first stage, we still use the first stage to net out the non-transmitted production function error \( \epsilon_{it} \). This is what allows us to treat the evolution of the transmitted error \( \omega_{it} \) non-parametrically. We show that our estimator is robust to a number of alternative (and seemingly reasonable) DGPs. As well as addressing the above collinearity problem, another important benefit of our estimator is that it makes comparison to the dynamic panel literature, e.g. Arellano and Bond (1991), quite easy. We are able to highlight the advantages and disadvantages of our estimator in relation to this dynamic panel literature. Lastly, using the same dataset as Levinsohn and Petrin, we examine how our estimator...
works in practice. Estimates using our methodology appear more stable across different potential proxy variables than do estimates using the Levinsohn-Petrin methodology, consistent with our theoretical arguments.

References


Buettner, T. (2005) "Productivity Dynamics when Labor is a State Variable", mimeo, LSE


34


Kasahara, H. and Rodrique, J. (2004) "Does the Use of Imported Intermediates Increase Productivity? Plant Level Evidence", mimeo, Queen’s University


### 8 Appendix 1 - Alternative Procedure

This section examines an alternative procedure to estimate production function coefficients. While it also breaks the potential collinearity problems of OP/LP, it does rely on some additional assumptions, specifically an additional monotonicity assumption and independence assumptions on innovations in the \( \omega_{it} \) process. It is also a bit more complicated than the procedure we suggest above. On the other hand, this procedure does allow one to learn something about when inputs are chosen, e.g. how "variable" an input labor is.

The intuition behind identification in this second approach follows directly from the intuition of identification of the coefficient on capital in OP (and LP). We make heavy use of the fact that if an input is determined prior to production, the innovation in productivity between the time of the input choice and the time of production should be orthogonal to that input choice. Again, this is not only an econometric assumption, but an assumption on the information set of the firm at various points in time. More formally, if \( \omega_i \) is the productivity level of the firm at the time input level \( i \) is chosen, and \( \omega_p \) is the productivity level at the time of production, then:

\[
(\omega_p - E[\omega_p|\omega_i]) \perp i
\]

This type of moment identifies the capital coefficient in OP and LP. Our approach simply extends this intuition to identification of parameters on labor inputs, combining this with non-parametrics to "invert out" values of the productivity shock at various decision times.

Consider a production model with 3 inputs, capital, labor, and an intermediate input, e.g. materials. We make the following timing assumptions regarding when \( k, l, \) and \( m \) are chosen. Suppose between periods \( t - 1 \) and \( t \), the following occurs, where \( 0 < b < 1 \):

<table>
<thead>
<tr>
<th>Time</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t - 1 )</td>
<td>( \omega_{it-1} ) is observed, ( m_{it-1} ) is chosen, ( k_{it} ) is chosen, period ( t - 1 ) production occurs</td>
</tr>
<tr>
<td>( t - b )</td>
<td>( \omega_{it-b} ) is observed, ( l_{it} ) is chosen</td>
</tr>
<tr>
<td>( t )</td>
<td>( \omega_{it} ) is observed, ( m_{it} ) is chosen, ( k_{it+1} ) is chosen, period ( t ) production occurs</td>
</tr>
</tbody>
</table>

Like OP/LP, we assume that \( k_{it} \) is determined at time \( t - 1 \). Actually, like LP (but not OP), we only really need to assume that \( k_{it} \) is determined at either \( t - 1 \) or earlier. For the more variable inputs, we assume that \( l_{it} \) is chosen at some time \( t - b \) (between \( t - 1 \) and \( t \)), and that \( m_{it} \) is perfectly flexible and chosen at time \( t \).

Note that we assume \( \omega \) evolves between \( t - 1, t - t_b \), and \( t \). As in our "story" behind OP, this movement is needed to alleviate possible collinearity problems between labor and other inputs. We assume that \( \omega \) evolves as a first-order markov process between these stages, i.e.:

\[
\begin{align*}
\omega_{it-b} &= g_1(\omega_{it-1}, \eta_{it}^b) \\
\omega_{it} &= g_2(\omega_{it-1}, \eta_{it})
\end{align*}
\]

(37)
where the $\eta$'s are independent of the $\omega$'s (as well as all other variables that are chosen before their realizations). Note that this is a stronger assumption than that of OP, LP, and the estimator proposed in the main section of this paper. Those assume only a first-order markov process on $\omega$. On the other hand, the fact that the $g$'s are arbitrary functions does allow some forms of heteroskedasticity. While our "staggered" input choice process might initially seem somewhat ad-hoc, we feel that it does capture some interesting aspects of reality.\footnote{Though this is clearly a stylized model of what is likely a more continuous decision process.}

Given the above timing assumptions and assuming that labor is a static input, a firm’s choice of labor will be a function of $\omega_{it-b}$, i.e.

\begin{equation}
    l_{it} = f_{1t}(\omega_{it-b}, k_{it})
\end{equation}

Since the firm’s choice of labor in a given period is made before its choice of materials, the labor term will be taken into account when choosing the level of materials, i.e.

\begin{equation}
    m_{it} = f_{2t}(\omega_{it}, k_{it}, l_{it})
\end{equation}

Once again, we will assume monotonicity of this equation in $\omega_{it}$, allowing us to invert this function and obtain:

\[ \omega_{it} = f_{2t}^{-1}(m_{it}, k_{it}, l_{it}) \]

This term can be substituted into the production function from (1) to get:

\[ y_{it} = \beta_k k_{it} + \beta_l l_{it} + f_{2t}^{-1}(m_{it}, k_{it}, l_{it}) + \epsilon_{it} \]

and collecting terms results in the first stage equation:

\begin{equation}
    y_{it} = \Phi_t(m_{it}, k_{it}, l_{it}) + \epsilon_{it}
\end{equation}

This is exactly the same first stage as section 5.1, and the $\Phi$ function can be estimated in the same way. Similarly, we can construct the same moment condition for capital:

\begin{equation}
    E[\xi_{it}(\beta_k, \beta_l)|k_{it}] = 0
\end{equation}

where $\xi_{it} = \omega_{it} - E[\omega_{it}|\omega_{it-1}]$, and $\xi_{it}(\beta_k, \beta_l)$ can be constructed in the usual way, i.e. by non-parametrically regressing $(\omega_{it}(\beta_k, \beta_l)) = \Phi_t(m_{it}, k_{it}, l_{it}) - \beta_k k_{it} - \beta_l l_{it})$ on $(\omega_{it-1}(\beta_k, \beta_l)) = \Phi_t(m_{it-1}, k_{it-1}, l_{it-1}) - \beta_k k_{it-1} - \beta_l l_{it-1})$.

What differs between this and the above procedures is the moment condition intended to identify the labor coefficient. Define $\xi_{it}^b$ as the unexpected innovation in $\omega$ between time $t-b$ and $t$, i.e.

\[ \xi_{it}^b = \omega_{it} - E[\omega_{it}|\omega_{it-b}] \]

Given that labor is chosen at $t-b$, it should be orthogonal to this innovation

\begin{equation}
    E[\xi_{it}^b l_{it}] = 0
\end{equation}

This is the moment condition we will use - what remains to be shown is how we can construct a sample analog to this moment given a value of the parameter vector. To do this, first note that
the first stage estimates of (40) allow us to compute, conditional on the parameters, $\omega_{it}$ for all $t$. Call these terms $\omega_{it}(\beta_k, \beta_l)$. Now consider the firm’s labor demand function (38). Substituting in (37) results in

$$
l_{it} = f_{it}(g_{1}(\omega_{it-1}, \eta_{it}^{b}), k_{it})
= \tilde{f}_{it}(\omega_{it-1}(\beta_k, \beta_l), \eta_{it}^{b}, k_{it})
$$

Note that conditional on $(\beta_k, \beta_l)$, the only unobservable in this equation is $\eta_{it}^{b}$. Thus, assuming that the equation is strictly monotonic in $\eta_{it}^{b}$, one can use the methods of Matzkin (2003) to non-parametrically invert out $\eta_{it}^{b}$ up to a normalization. Call this function $\tau(\eta_{it}^{b}; \beta_k, \beta_l)$. Again, the dependence on $\beta_k$ and $\beta_l$ comes from the fact that the $\omega_{it}$ are inferred conditional on $\beta_k$ and $\beta_l$. This non-parametric inversion relies on the assumption that $\eta_{it}^{b}$ is independent of $\omega_{it-1}$ and $k_{it}$. The basic intuition is that for a given $\omega_{it-1}$ and $k_{it}$, one can form a distribution of $l_{it}$. $\tau(\eta_{it}^{b}; \beta_k, \beta_l)$ for a given $i$ is simply the quantile of $l_{it}$ in that distribution.

Next, note that since $\omega_{it-b}$ is a function of $\eta_{it}^{b}$ and $\omega_{it-1}$, we can also write it as a function of $\tau(\eta_{it}^{b}; \beta_k, \beta_l)$ and $\omega_{it-1}$, i.e.

$$
\omega_{it-b}(\beta_k, \beta_l) = g_{1}(\omega_{it-1}(\beta_k, \beta_l), \eta_{it}^{b})
= \tilde{g}_{1}(\omega_{it-1}(\beta_k, \beta_l), \tau(\eta_{it}^{b}; \beta_k, \beta_l))
$$

As a result, to construct $\xi_{it}^{b} = \omega_{it} - E[\omega_{it} | \omega_{it-b}]$, we can form the necessary conditional expectation by non-parametrically regressing $\omega_{it}(\beta_k, \beta_l)$ on $\omega_{it-1}(\beta_k, \beta_l)$ and $\tau(\eta_{it}^{b}; \beta_k, \beta_l)$ (as an alternative to non-parametrically regressing $\omega_{it}(\beta_k, \beta_l)$ on $\omega_{it-b}(\beta_k, \beta_l)$). Denoting the residual from this regression by $\xi_{it}^{b}(\beta_k, \beta_l)$, we can form the moment

$$
E[\xi_{it}^{b}(\beta_k, \beta_l)|l_{it}] = 0
$$

to be used for estimation. Note that this procedure can easily be adjusted to allow for labor to have dynamic implications. One simply needs to include $l_{it-1}$ in both the material and labor demand functions.

One nice aspect of this procedure is that it allows us to infer something about when inputs are chosen. The basic idea here is to compare how well $\omega_{it-1}(\beta_k, \beta_l)$ non-parametrically predicts $\omega_{it}(\beta_k, \beta_l)$ to how well $\omega_{it-1}(\beta_k, \beta_l)$ and $\tau(\eta_{it}^{b}; \beta_k, \beta_l)$ (i.e. $\omega_{it-b}(\beta_k, \beta_l)$) non-parametrically predicts $\omega_{it}(\beta_k, \beta_l)$. Intuitively, if adding $\tau(\eta_{it}^{b}; \beta_k, \beta_l)$ sharpens the prediction by alot, it suggests that $\omega_{it-b}(\beta_k, \beta_l)$ is "close" to $\omega_{it}(\beta_k, \beta_l)$, i.e. that $t - b$ is close to $t$, and that labor is a fairly variable input. In contrast, if adding $\tau(\eta_{it}^{b}; \beta_k, \beta_l)$ does not help explain $\omega_{it}(\beta_k, \beta_l)$ much, it suggests that $\omega_{it-b}(\beta_k, \beta_l)$ is close to $\omega_{it-1}(\beta_k, \beta_l)$ (and $t - b$ is close to $t - 1$) and that labor is more of a fixed input.
<table>
<thead>
<tr>
<th>Industry 311</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.336</td>
<td>1.080</td>
<td>1.416</td>
</tr>
<tr>
<td>FE</td>
<td>0.081</td>
<td>0.719</td>
<td>0.800</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.371</td>
<td>0.842</td>
<td>1.212</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.379</td>
<td>0.865</td>
<td>1.244</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.395</td>
<td>0.884</td>
<td>1.279</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.455</td>
<td>0.676</td>
<td>1.131</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.446</td>
<td>0.764</td>
<td>1.210</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.410</td>
<td>0.942</td>
<td>1.352</td>
</tr>
<tr>
<td>DP</td>
<td>0.391</td>
<td>0.987</td>
<td>1.378</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry 321</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.256</td>
<td>0.953</td>
<td>1.210</td>
</tr>
<tr>
<td>FE</td>
<td>0.204</td>
<td>0.724</td>
<td>0.927</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.242</td>
<td>0.893</td>
<td>1.135</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.272</td>
<td>0.832</td>
<td>1.104</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.272</td>
<td>0.873</td>
<td>1.145</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.320</td>
<td>0.775</td>
<td>1.094</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.241</td>
<td>0.978</td>
<td>1.219</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.254</td>
<td>1.008</td>
<td>1.262</td>
</tr>
<tr>
<td>DP</td>
<td>0.320</td>
<td>0.837</td>
<td>1.157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry 331</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.236</td>
<td>1.038</td>
<td>1.274</td>
</tr>
<tr>
<td>FE</td>
<td>-0.028</td>
<td>0.897</td>
<td>0.869</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.196</td>
<td>0.923</td>
<td>1.119</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.195</td>
<td>0.897</td>
<td>1.092</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.212</td>
<td>0.915</td>
<td>1.127</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.352</td>
<td>0.678</td>
<td>1.030</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.305</td>
<td>0.786</td>
<td>1.090</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.241</td>
<td>0.993</td>
<td>1.234</td>
</tr>
<tr>
<td>DP</td>
<td>0.252</td>
<td>0.998</td>
<td>1.249</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry 381</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.223</td>
<td>1.160</td>
<td>1.383</td>
</tr>
<tr>
<td>FE</td>
<td>0.036</td>
<td>0.783</td>
<td>0.819</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.262</td>
<td>1.010</td>
<td>1.273</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.250</td>
<td>1.002</td>
<td>1.252</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.259</td>
<td>1.022</td>
<td>1.280</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.342</td>
<td>0.803</td>
<td>1.145</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.306</td>
<td>0.944</td>
<td>1.251</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.265</td>
<td>1.090</td>
<td>1.355</td>
</tr>
<tr>
<td>DP</td>
<td>0.275</td>
<td>1.056</td>
<td>1.331</td>
</tr>
<tr>
<td></td>
<td>Industry 311</td>
<td>Industry 311</td>
<td>Industry 321</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>ACF vs OLS</td>
<td>K</td>
<td>0.111</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>RTS</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>ACF vs LP</td>
<td>K</td>
<td>0.737</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>RTS</td>
<td>0.000</td>
<td>0.061</td>
</tr>
<tr>
<td>ACF vs DP</td>
<td>K</td>
<td>0.962</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>0.940</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>RTS</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Value is the % of bootstrap reps where ACF coeff is less than OLS, LP, or DP coef. A value either above 0.95 or below 0.05 indicates that coefficients are significantly different from each other.
<table>
<thead>
<tr>
<th>Industry</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>0.336</td>
<td>1.080</td>
<td>1.416</td>
</tr>
<tr>
<td><strong>FE</strong></td>
<td>0.081</td>
<td>0.719</td>
<td>0.800</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.304</td>
<td>0.993</td>
<td>1.297</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.344</td>
<td>0.957</td>
<td>1.301</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.371</td>
<td>0.940</td>
<td>1.311</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.455</td>
<td>0.676</td>
<td>1.131</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.446</td>
<td>0.764</td>
<td>1.210</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.410</td>
<td>0.942</td>
<td>1.352</td>
</tr>
<tr>
<td>DP</td>
<td>0.335</td>
<td>1.128</td>
<td>1.463</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>0.256</td>
<td>0.953</td>
<td>1.210</td>
</tr>
<tr>
<td><strong>FE</strong></td>
<td>0.204</td>
<td>0.724</td>
<td>0.927</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.211</td>
<td>0.959</td>
<td>1.170</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.261</td>
<td>0.856</td>
<td>1.117</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.264</td>
<td>0.892</td>
<td>1.155</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.320</td>
<td>0.775</td>
<td>1.094</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.241</td>
<td>0.978</td>
<td>1.219</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.254</td>
<td>1.008</td>
<td>1.262</td>
</tr>
<tr>
<td>DP</td>
<td>0.271</td>
<td>0.940</td>
<td>1.211</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>0.236</td>
<td>1.038</td>
<td>1.274</td>
</tr>
<tr>
<td><strong>FE</strong></td>
<td>-0.028</td>
<td>0.897</td>
<td>0.869</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.178</td>
<td>0.966</td>
<td>1.144</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.199</td>
<td>0.888</td>
<td>1.087</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.231</td>
<td>0.868</td>
<td>1.099</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.352</td>
<td>0.678</td>
<td>1.030</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.305</td>
<td>0.786</td>
<td>1.090</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.241</td>
<td>0.993</td>
<td>1.234</td>
</tr>
<tr>
<td>DP</td>
<td>0.206</td>
<td>1.119</td>
<td>1.324</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry</th>
<th>Capital</th>
<th>Labor</th>
<th>Returns to Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>0.223</td>
<td>1.160</td>
<td>1.383</td>
</tr>
<tr>
<td><strong>FE</strong></td>
<td>0.036</td>
<td>0.783</td>
<td>0.819</td>
</tr>
<tr>
<td>ACF – M</td>
<td>0.238</td>
<td>1.088</td>
<td>1.326</td>
</tr>
<tr>
<td>ACF – E</td>
<td>0.230</td>
<td>1.069</td>
<td>1.300</td>
</tr>
<tr>
<td>ACF – F</td>
<td>0.250</td>
<td>1.049</td>
<td>1.300</td>
</tr>
<tr>
<td>LP – M</td>
<td>0.342</td>
<td>0.803</td>
<td>1.145</td>
</tr>
<tr>
<td>LP – E</td>
<td>0.306</td>
<td>0.944</td>
<td>1.251</td>
</tr>
<tr>
<td>LP – F</td>
<td>0.265</td>
<td>1.090</td>
<td>1.355</td>
</tr>
<tr>
<td>DP</td>
<td>0.244</td>
<td>1.141</td>
<td>1.385</td>
</tr>
</tbody>
</table>
### TABLE 2A

<table>
<thead>
<tr>
<th>Industry 311</th>
<th>Industry 321</th>
<th>Industry 331</th>
<th>Industry 381</th>
</tr>
</thead>
<tbody>
<tr>
<td>M E F M E F</td>
<td>M E F M E F</td>
<td>M E F M E F</td>
<td>M E F M E F</td>
</tr>
<tr>
<td><strong>ACF vs OLS</strong></td>
<td><strong>ACF vs OLS</strong></td>
<td><strong>ACF vs OLS</strong></td>
<td><strong>ACF vs OLS</strong></td>
</tr>
<tr>
<td>K</td>
<td>0.864</td>
<td>0.754</td>
<td>0.266</td>
</tr>
<tr>
<td>L</td>
<td>0.739</td>
<td>0.874</td>
<td>1.000</td>
</tr>
<tr>
<td>RTS</td>
<td>0.980</td>
<td>0.975</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>ACF vs LP</strong></td>
<td><strong>ACF vs LP</strong></td>
<td><strong>ACF vs LP</strong></td>
<td><strong>ACF vs LP</strong></td>
</tr>
<tr>
<td>K</td>
<td>1.000</td>
<td>1.000</td>
<td>0.925</td>
</tr>
<tr>
<td>L</td>
<td>0.000</td>
<td>0.000</td>
<td>0.352</td>
</tr>
<tr>
<td>RTS</td>
<td>0.005</td>
<td>0.005</td>
<td>0.754</td>
</tr>
<tr>
<td><strong>ACF vs DP</strong></td>
<td><strong>ACF vs DP</strong></td>
<td><strong>ACF vs DP</strong></td>
<td><strong>ACF vs DP</strong></td>
</tr>
<tr>
<td>K</td>
<td>0.709</td>
<td>0.307</td>
<td>0.040</td>
</tr>
<tr>
<td>L</td>
<td>0.990</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>RTS</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Value is the % of bootstrap reps where ACF coeff is less than OLS, LP, or DP coef. A value either above 0.95 or below 0.05 indicates that coefficients are significantly different from each other.