Dynamic Linear Panel Data

Bryan S. Graham, UC - Berkeley & NBER

March 18, 2021

The recovery of a firm's cost function from price, quantity and/or input data features widely in empirical IO analyzes of industry competition. The structure of cost/production functions has important regulatory implications. Finally, understanding the nature of productivity differences across firm, as well as its growth over time, has profound implications for human well-being.

Panel data feature widely in empirical analyses of firm production. Indeed some of the earliest applications of panel data were to production function analysis (e.g., Mundlak, 1961). A useful overview of issues, and survey of work through the mid-1990s, is provided by Griliches & Mairesse (1998).

This note uses the production function example to introduce some basic ideas in dynamic panel data analysis. The Arellano (2003) book provides excellent coverage of this material. A second goal of this lecture note is to develop some connections between the Olley & Pakes (1996) approach to production function estimation and more more traditional dynamic panel data approaches.

Olley and Pakes approach

We will work with the following Cobb-Douglas production function

$$O_{it} = L_{it}^{\alpha} K_{it}^{\beta} \exp\left(A_{it}^{*} + U_{it}\right),$$

where O_{it} is firm i's period t output, L_{it} its labor input and K_{it} its capital input. Productivity has two components: (i) U_{it}, an unforecastable stochastic input / productivity shock which is revealed after all period t decisions are made and (ii) a systematic productivity component which, as in Olley & Pakes (1996), follows a first-order Markov process:

$$\Pr\left(A_{t}^{*} \leq a \middle| \mathcal{I}_{t}\right) = \Pr\left(A_{t}^{*} \leq a \middle| A_{t-1}^{*}\right), \tag{1}$$

with $\mathcal{I}_t$ denoting a firm's beginning of period information set (as usual I suppress the i subscript when doing so causes no confusion). This process is stochastically increasing such that the CDF function shifts rightward with last period's productivity (firms that were more productive yesterday are more likely to more productive today). Restriction (1), as we will see, is *the* identifying assumption for what follows. Furthermore it is far from innocuous; it rules out the possibility that firms engage in behaviors which influence the direction of TFP (e.g., R&D expenditures). Even if the researcher is willing to maintain the assumption of an exogenous productivity process, the Markov assumption is strong (as we shall see below).

The Markov assumption delivers the L_2 projection

$$\begin{aligned} A_t^* &= \mathbb{E}\left[A_t^* \middle| \mathcal{I}_t\right] + \varepsilon_t \\ &= h\left(A_{t-1}^*\right) + \varepsilon_t \end{aligned}$$

with

$$\mathbb{E}\left[\varepsilon_t \middle| \mathcal{I}_t\right] = 0. \tag{2}$$

by construction. In words we can decompose productivity into a component which is predictable by the firm (before input choices are made) and productivity *innovation* which is not predictable. Restriction (2) will be the basis for estimation. Since information sets grow over time such that $\mathcal{I}_0 \subset \mathcal{I}_1 \subset \cdots \mathcal{I}_T$, equation 2 implies a sequential set of conditional moment restrictions. Chamberlain (1992) provides a discussion of conditional moment restrictions of this type.

Each period the firm can adjust its labor input by hiring or laying off workers. Unlike in Olley & Pakes (1996), we will assume that adjusting the labor stock is costly (see Bond & Söderbom (2005) for interesting discussion). This makes total employment a state variable, and the period-specific amount of hiring/firing a control. Capital stock is determined one period in advance by the firm's choice of capital expenditure, I_t, (i.e., investment level). Capital is a second state variable, with investment a corresponding control. Firms choose the amount of hiring and investment each period to maximized the discounted sum of expected profits. Solving this dynamic program will result in the policy functions

$$\begin{aligned} L_{it} &= g_L\left(L_{it-1}, K_{it}, A_{it}^*; W_{it}\right) \\ I_{it} &= g_K\left(L_{it}, K_{it}, A_{it}^*; W_{it}\right), \end{aligned}$$

with W_{it} equal to input prices (i.e., wage rate, rental price of capital etc.). We might also include other state variables in these functions (e.g., firm-specific demand-shifters). Note we assume that the firm can adjust is current labor supply (albeit it is costly to do so), whereas adjustments to the capital stock are not instantaneous, but made one period in advance.

Olley & Pakes (1996) showed that when the TFP process is stochastically increasing *and* firms face the same input prices and demand shocks (which means that $W_{it} \equiv W_t$ for all i), that $g_K\left(L_{it}, K_{it}, A_{it}^*; W_t\right) \overset{def}{\equiv} g_{Kt}\left(L_{it}, K_{it}, A_{it}^*\right)$ is invertible. Taking logs of the production function, and defining $Y_t = \ln O_t$, $X_t^L = \ln L_t$, and $X_t^K = \ln K_t$, therefore yields

$$\begin{aligned}
Y_{it} &= \alpha X_{it}^L + \beta X_{it}^K + A_{it}^* + U_{it} \\
&= \alpha X_{it}^L + \beta X_{it}^K + g_{Kt}^{-1}\left(L_{it-1}, K_{it}, I_{it}\right) + U_{it} \\
&= \phi_t\left(X_{it}^L, X_{it}^K, I_{it}\right) + U_{it}
\end{aligned}$$

for $\phi_t\left(X_{it}^L, X_{it}^K, I_{it}\right) \overset{def}{\equiv} \mathbb{E}\left[\left.Y_{it}\right| X_{it}^L, X_{it}^K, I_{it}\right]$.

Next define

$$A_{it}^*\left(\alpha, \beta, \phi_t\left(\cdot\right)\right) \overset{def}{\equiv} \phi_t\left(X_{it}^L, X_{it}^K, I_{it}\right) - \alpha X_{it}^L - \beta X_{it}^K.$$

If α and β were known, it would be possible to recover the systematic component of TFP from the conditional expectation function $\phi_t\left(X_{it}^L, X_{it}^K, I_{it}\right)$. If the $h\left(\cdot\right)$ function, which maps A_{it-1}^* into (expected) A_{it}^*, were also known, then we could further recover the innovation to productivity as

$$\varepsilon_{it}\left(\alpha, \beta, \phi_t\left(\cdot\right), \phi_{t-1}\left(\cdot\right), h\left(\cdot\right)\right) = A_{it}^*\left(\alpha, \beta, \phi_t\left(\cdot\right)\right) - h\left(A_{it-1}^*\left(\alpha, \beta, \phi_{t-1}\left(\cdot\right)\right)\right)$$

with

$$h\left(A_{it-1}^*\left(\alpha, \beta, \phi_{t-1}\left(\cdot\right)\right)\right) \overset{def}{\equiv} \mathbb{E}\left[\left.A_{it}^*\left(\alpha, \beta, \phi_t\left(\cdot\right)\right)\right| A_{it-1}^*\left(\alpha, \beta, \phi_{t-1}\left(\cdot\right)\right)\right].$$

Using (2) above, then delivers the following sequential conditional moment conditions

$$\mathbb{E}\left[\left.\varepsilon_{it}\left(\alpha, \beta, \phi_t\left(\cdot\right), \phi_{t-1}\left(\cdot\right), h\left(\cdot\right)\right)\right| \mathcal{I}_t\right] = 0,\ t = 1, \ldots, T. \tag{3}$$

Note I am assuming all parameters and functions are being evaluated at their "true" population values. This problem contains $T+2$ infinite dimensional (nuisance) parameters: $\phi_0\left(\cdot\right), \phi_1\left(\cdot\right), \ldots, , \phi_T\left(\cdot\right), h\left(\cdot\right)$. We could also make the Markov process non-stationary, which means we'd need to estimate a different $h\left(\cdot\right)$ function for each period. Building on Chamberlain (1992) and others, Ai & Chen (2012) study semi-parametric efficiency bounds in problems like 3.

In practice the econometrician is only able to form moment conditions based on a subset of the firm's actual period t information set. Depending on the exact data available the common practice is to work with $\left(K_{i0}^{t}, L_{i0}^{t-1}, I_{i0}^{t-1}, Y_{i0}^{t-1}\right) \subset \mathcal{I}_t$:

$$\mathbb{E}\left[\varepsilon_{it}\left(\alpha, \beta, \phi_t\left(\cdot\right), \phi_{t-1}\left(\cdot\right), h\left(\cdot\right)\right) \middle| K_{i0}^{t}, L_{i0}^{t-1}, I_{i0}^{t-1}, Y_{i0}^{t-1}\right] = 0,\ t = 1, \ldots, T$$

Other conditioning variables can be added if they are available. Recall that $Z_0^t = (Z_0, Z_1, \ldots, Z_t)'$. Let $J = \frac{T}{2}\left(T+3\right) + \frac{3T}{2}\left(T+1\right) = 2T^2 + 3T$ and define the following $T \times J$ instrument matrix:

$$\underset{T\times J}{Z} = \begin{bmatrix} \left(X_0^K, X_1^K, X_0^L, I_0, Y_0\right) & & & 0 \\ & \left(X_0^K, X_1^K, X_2^K, X_0^L, X_1^L, I_0, I_1, Y_0, Y_1\right) & & \\ & & \ddots & \\ 0 & & & \left(X_0^K, \ldots, X_T^K, X_0^L, \ldots, X_{T-1}^L I_0, \ldots, I_{T-1}, Y_0, \ldots, Y_{T-1}\right) \end{bmatrix}$$

Next define the vector of productivity innovations

$$\underline{\varepsilon}_i\left(\alpha, \beta, \phi_0\left(\cdot\right), \ldots, \phi_T\left(\cdot\right), h\left(\cdot\right)\right) = \begin{pmatrix} \varepsilon_{i1}\left(\alpha, \beta, \phi_t\left(\cdot\right), \phi_{t-1}\left(\cdot\right), h\left(\cdot\right)\right) \\ \varepsilon_{i2}\left(\alpha, \beta, \phi_t\left(\cdot\right), \phi_{t-1}\left(\cdot\right), h\left(\cdot\right)\right) \\ \vdots \\ \varepsilon_{iT}\left(\alpha, \beta, \phi_T\left(\cdot\right), \phi_{T-1}\left(\cdot\right), h\left(\cdot\right)\right) \end{pmatrix}.$$

Putting things together then yields the J unconditional moment restrictions:

$$\mathbb{E}\left[Z_i'\underline{\varepsilon}_i\left(\alpha, \beta, \phi_0\left(\cdot\right), \ldots, \phi_T\left(\cdot\right), h\left(\cdot\right)\right)\right] = 0. \tag{4}$$

Estimation

Equation (4) defines a fairly complicated semiparametric estimation problem. In practice estimation proceeds according to some variant of the following iterative procedure (see for example de Loecker & Warzynski (2012)).

1. Construct estimates of $\phi_0\left(\cdot\right), \ldots, \phi_T\left(\cdot\right)$. This is a low-dimensional non-parametric estimation problem, so any number of non-parametric regression estimators are suitable.

2. Let $\alpha^{(s)}, \beta^{(s)}$ be the current values of the labor and capital elasticities.

(a) Compute the pooled nonparametric regression of

$$A_{it}^{*}\left(\alpha^{(s)},\beta^{(s)},\hat{\phi}_{t}\left(\cdot\right)\right)$$

onto

$$A_{it-1}^{*}\left(\alpha^{(s)},\beta^{(s)},\hat{\phi}_{t-1}\left(\cdot\right)\right).$$

This generates an estimate of the $h\left(\cdot\right)$ function. The fitted residuals from this regression fit correspond to estimates of the productivity innovations

$$\varepsilon_{it}\left(\alpha^{(s)},\beta^{(s)},\hat{\phi}_{t}\left(\cdot\right),\hat{\phi}_{t-1}\left(\cdot\right),\hat{h}\left(\cdot\right)\right),\ i=1,\ldots,N,\ t=1,\ldots,T.$$

(b) Construct the quadratic form:

$$\begin{aligned}&\left[\frac{1}{N}\sum_{i=1}^{N}Z_{i}'\underline{\varepsilon}_{i}\left(\alpha^{(s)},\beta^{(s)},\hat{\phi}_{0}\left(\cdot\right),\ldots,\hat{\phi}_{T}\left(\cdot\right),\hat{h}\left(\cdot\right)\right)\right]'\\ \times W&\left[\frac{1}{N}\sum_{i=1}^{N}Z_{i}'\underline{\varepsilon}_{i}\left(\alpha^{(s)},\beta^{(s)},\hat{\phi}_{0}\left(\cdot\right),\ldots,\hat{\phi}_{T}\left(\cdot\right),\hat{h}\left(\cdot\right)\right)\right],\end{aligned}$$

with W a $J\times J$ weight matrix.

3. Repeat steps 2a and 2b, varying $\alpha^{(s+1)},\beta^{(s+1)}$ to minimize the quadratic form criterion function.

The outer-minimization, over α and β, can be completed using a grid search. In practice $\phi_{t}\left(\cdot\right),\phi_{t-1}\left(\cdot\right),h\left(\cdot\right)$ are often assumed to be low order polynomials. While it is somewhat tedious to set-up the problem as a giant GMM problem, doing so makes inference straightforward. de Loecker (2013) and de Loecker & Warzynski (2012) are recent empirical examples.

Dynamic panel data approach

The development of dynamic panel data methods in the 1990s was motivated in part, by estimation problems arising in the context of firm-level panel data (see, for example Arellano & Bond (1991) and Blundell & Bond (1998)). Blundell & Bond (2000) discuss the application to production function estimation. It turns out that there is a close relationship between dynamic panel data production function estimators and the Olley-Pakes one. Developing this relationship provides insight into the strengths and weaknesses of both approaches.

The dynamic panel data approach requires a parametric restriction on the productivity process. We will start by assuming that the Markov process is linear such that

$$A_{it}^{*}=\lambda_{t}+\rho A_{it-1}^{*}+\varepsilon_{t}.$$

This is restrictive relative to the nonparametric specification used above (although we now allow for some non-stationarity in the form of time-specific intercepts). As above we can write the productivity innovation as

$$\begin{aligned}\varepsilon_{it}\left(\alpha,\beta,\phi_{t}\left(\cdot\right),\phi_{t-1}\left(\cdot\right)\right)=&A_{it}^{*}\left(\alpha,\beta,\phi_{t}\left(\cdot\right)\right)-\lambda_{t}-\rho A_{it-1}^{*}\left(\alpha,\beta,\phi_{t-1}\left(\cdot\right)\right)\\ =&\left(\phi_{t}\left(X_{it}^{L},X_{it}^{K},I_{it}\right)-\alpha X_{it}^{L}-\beta X_{it}^{K}\right)-\lambda_{t}\\ &-\rho\left(\phi_{t-1}\left(X_{it-1}^{L},X_{it-1}^{K},I_{it-1}\right)-\alpha X_{it-1}^{L}-\beta X_{it-1}^{K}\right).\end{aligned}$$

Now the parametric component of the parameter is

$$\theta=\left(\lambda_{1},\ldots,\lambda_{T},\rho,\alpha,\beta\right)'$$

The linear Markov assumptions simplifies things; but working with it has an even larger payoff. Note that our basic model also implies the sequential conditional moment restriction

$$\mathbb{E}\left[\left.\varepsilon_{t}+U_{t}\right|\mathcal{I}_{t}\right]=0. \tag{5}$$

Next observe that since $Y_{it}=\phi_{t}\left(X_{it}^{L},X_{it}^{K},I_{it}\right)+U_{it}$ we have

$$\begin{aligned}\varepsilon_{it}=&\left(Y_{it}-U_{it}-\alpha X_{it}^{L}-\beta X_{it}^{K}\right)-\lambda_{t}-\rho\left(Y_{it-t}-U_{it-t}-\alpha X_{it-1}^{L}-\beta X_{it-1}^{K}\right)\\ \varepsilon_{it}+U_{it}-\rho U_{it-1}=&\left(Y_{it}-\alpha X_{it}^{L}-\beta X_{it}^{K}\right)-\lambda_{t}-\rho\left(Y_{it-t}-\alpha X_{it-1}^{L}-\beta X_{it-1}^{K}\right)\\ \overset{def}{\equiv}&\ \rho_{it}\left(\lambda_{t},\rho,\alpha,\beta\right),\end{aligned}$$

with

$$\mathbb{E}\left[\left.\rho_{it}\left(\lambda_{t},\rho,\alpha,\beta\right)\right|\mathcal{I}_{it}^{*}\right]=0.$$

The notation $\mathcal{I}_{it}^{*}$ is a bit awkward; it includes everything that was in $\mathcal{I}_{it}$ except Y_{it-1}. Because Y_{it-1} and U_{it-1} covary we need to remove it from the conditioning set; however since U_{it-1} is an unforecastable productivity shock, realized after any period $t-1$ decisions are made, it will be mean independence of any other component of $\mathcal{I}_{it}$.

Relative to Olley-Pakes we are working with both (i) a parametric restriction on the productivity process and (ii) a slightly different set of sequential moment restrictions. The approaches are not strictly nested, making formal comparisons – for example in terms of efficiency – subtle.[1] For estimation purposes we define the instrument matrix

$$\underset{T\times J}{Z} = \begin{bmatrix} \left(X_0^K, X_1^K, X_0^L, I_0\right) & & & 0 \\ & \left(X_0^K, X_1^K, X_2^K, X_0^L, X_1^L, I_0, I_1, Y_0\right) & & \\ & & \ddots & \\ 0 & & & \left(X_0^K, \ldots, X_T^K, X_0^L, \ldots, X_{T-1}^L I_0, \ldots, I_{T-1}, Y_0, \ldots, Y_{T-2}\right) \end{bmatrix}$$

with $J = \frac{T}{2}(T+3) + \frac{2T}{2}(T+1) + \frac{1}{2}T(T-1) = 2(T^2+T)$ and base estimation on

$$\mathbb{E}\left[Z_i' \underline{\rho}_i(\lambda_1, \ldots, \lambda_T, \rho, \alpha, \beta)\right] = 0$$

with $\underline{\rho}_i(\lambda_1, \ldots, \lambda_T, \rho, \alpha, \beta) = (\rho_{i1}(\lambda_1, \rho, \alpha, \beta), \ldots, \rho_{iT}(\lambda_T, \rho, \alpha, \beta))'$. In terms on computation and inference, this is much "easier" than Olley-Pakes (although, in fairness, Olley-Pakes is pretty manageable in the context of modern empirical IO).

The connection to dynamic panel data is easiest to see by noting that, since $A_{it}^* = \lambda_t + \rho A_{it-1}^* + \varepsilon_t$ we have $\rho Y_{it-1} = \alpha\rho X_{it-1}^L + \beta\rho X_{it-1}^K + \rho A_{it-1}^* + \rho U_{it-t}$. The quasi-difference $Y_{it} - \rho Y_{it-1}$ thus eliminates the forecastable component of productivity A_{it-1}^*, yielding (after bringing ρY_{it-1} to the write of the equality)

$$Y_{it} = \lambda_t + \rho Y_{it-t} + \alpha X_{it}^L - \alpha\rho X_{it-1}^L + \beta X_{it}^K - \beta\rho X_{it-1}^K + \varepsilon_{it} + U_{it} - \rho U_{it-t},$$

which is a linear AR(1) dynamic panel data model (with a common factor and an MA(1) "error" process). See, for example, Arellano & Bond (1991), Arellano & Bover (1995) and Blundell & Bond (2000).

The first order Markov assumption is restrictive. For example, the sum of two Markov processes will typically not be Markov. This implies that a productivity process consisting of a permanent firm-specific effect and an AR(1) process can not be accommodated in the Olley-Pakes approach. An advantage of the dynamic panel data approach is that a firm-

[1] See Ackerberg et al. (2014) for related discussion.

specific effect can be included. Let

$$A_{it}^{*} = A_i + V_{it}$$

with

$$V_{it} = \lambda_t + \rho V_{it-1} + \varepsilon_t.$$

This productivity assumption is simultaneously more restrictive (linearity) and more flexible (firm-specific productivity component) than the Olley-Pakes one. This implies an linear AR(1) dynamic panel data model (with a common factor), as before, but now with the addition of a firm-specific intercept:

$$\begin{aligned} Y_{it} &= \alpha X_{it}^{L} + \beta X_{it}^{K} + A_{it}^{*} + U_{it} \\ &= \lambda_t + \rho Y_{it-t} + \alpha X_{it}^{L} - \alpha\rho X_{it-1}^{L} + \beta X_{it}^{K} - \beta\rho X_{it-1}^{K} + (1-\rho) A_i + \varepsilon_{it} + U_{it} - \rho U_{it-1} \end{aligned}$$

To eliminate the firm-specific intercept we can take first differences. Re-arranging this yields

$$\triangle Y_{it} - \rho\triangle Y_{it-1} - (\lambda_t - \lambda_{t-1}) - \alpha\triangle X_{it}^{L} + \alpha\rho\triangle X_{it-1}^{L} - \beta\triangle X_{it}^{K} + \beta\rho\triangle X_{it-1}^{K} = \triangle\varepsilon_{it} + \triangle U_{it} - \rho\triangle U_{it-1}$$

with the conditional moment restriction

$$\mathbb{E}\left[\triangle\varepsilon_{it} + \triangle U_{it} - \rho\triangle U_{it-1} \middle| \mathcal{I}_{it-1}^{*}\right] = 0.$$

Re-define $\rho_{it}(\lambda_{t-1}, \lambda_t, \rho, \alpha, \beta) = \triangle Y_{it} - \rho\triangle Y_{it-1} - (\lambda_t - \lambda_{t-1}) - \alpha\triangle X_{it}^{L} + \alpha\rho\triangle X_{it-1}^{L} - \beta\triangle X_{it}^{K} + \beta\rho\triangle X_{it-1}^{K}$, normalize $\lambda_1 = 0$, and set $\underline{\rho}_i(\lambda_2, \ldots, \lambda_T, \rho, \alpha, \beta) = (\rho_{i2}(\lambda_2, \rho, \alpha, \beta), \ldots, \rho_{iT}(\lambda_{T-1}, \lambda_T, \rho, \alpha, \beta))'$. Our instrument matrix is now

$$\underset{T-1\times J}{Z} = \begin{bmatrix} \left(X_0^K, X_1^K, X_0^L, I_0\right) & & & 0 \\ & \left(X_0^K, X_1^K, X_2^K, X_0^L, X_1^L, I_0, I_1, Y_0\right) & & \\ & & \ddots & \\ 0 & & & \left(X_0^K, \ldots, X_{T-1}^K, X_0^L, \ldots, X_{T-2}^L I_0, \ldots, I_{T-2}, Y_0, \ldots, Y_{T-3}\right) \end{bmatrix}$$

with $J = \frac{T-1}{2}(T+2) + \frac{2(T-1)}{2}T + \frac{(T-1)}{2}(T-2) = 2T(T-1)$. We base estimation on

$$\mathbb{E}\left[Z_i'\underline{\rho}_i(\lambda_2, \ldots, \lambda_T, \rho, \alpha, \beta)\right] = 0.$$

This is a variant of the famous Arellano-Bond dynamic panel data moment condition. In the language of that literature we are using Y_{it-3} as an instrument for $\triangle Y_{it-1}$. You may be more familiar with the idea of using "Y_{it-2} as an instrument for $\triangle Y_{it-1}$". We cannot do that here because the presence of the linear Markov productivity process induces a MA(1) error process (hence the need to lag an additional period.

It is well-known in the literature on dynamic panel data that lagged levels of the outcome may only be weakly correlated with first differences of the outcome (see Blundell & Bond (1998, 2000) for an accessible discussion). Indeed this model was an important reference point in the literature on "many weak instruments". In the context of our specific production function example it is possible that the inclusion of the "outside" investment instruments may help alleviate this concern (or other outside components of the firm's information set – like past demand shifters). In applications of dynamic panel data methods to production function estimation such outside instruments are not typically used. Their use here can be viewed as one advantage of building up this estimation procedure in terms of an explicit productivity process and information set assumption.

Before turning to the topic of endogenous productivity, it is worth making a few observations.

1. Both the Olley-Pakes and Arellano-Blundell-Bond approaches to production function estimation – as formulated here – are based on the assumption that the innovation to productivity is unforecastable. The fundamental economic restriction is $\mathbb{E}\left[A_{it}^*\middle|\,\mathcal{I}_{it}\right] = \mathbb{E}\left[A_{it}^*\middle|\,A_{it-1}^*\right] = 0$ in both approaches.

2. They make different non-nested assumptions about the productivity process. Olley-Pakes allows for a general first-order Markov process, but their set-up cannot incorporate permanent differences in productivity across firms.

3. In addition to modeling productivity differently, they use slightly different conditional moment restrictions. This makes efficient comparisons complicated.

Sometimes Arellano-Bond type instrument sets are referred to as "internal instruments". Technically this is correct – past outcomes and regressors are used as instruments; but I also think it is mis-leading. The instruments are implications of a "structural" economic assumption about the firm's productivity process and information set when making period t decisions. The economics are delivering the restrictions upon which estimation is based. There is nothing "automatic" or magical about the instruments. As the above discussion makes clear, the underlying economics of the Olley-Pakes and Arellano-Blundell-Bond approaches are, more or less, the same. They differ in how they operationalize those assumptions to create a workable approach to estimation.

Both approaches pose practical difficulties. The Olley-Pakes approach requires estimating numerous infinite dimensional nuisance functions. Here cross-fitting, cross-validation and associated ideas may be quite useful in practice. Second there is abundant evidence that Hansen's two-step GMM estimator performs poorly in the context of many overidentifying "weak" moment restrictions. This could be an issue for the last approach discussed – a variant of Arellano & Bond (1991) where the presence of the firm-specific productivity component necessitates first differencing. Solutions to this problem include ad hoc reductions in the instrument set, modeling the optimal instruments as in Arellano (2016), or using GEL instead of two-step GMM.

Endogenous productivity

Ample evidence indicates that productivity differs dramatically across firms, even among those in the same (narrowly defined) industry (cf., Syverson, 2011). Why? This is a great question with lots of recent work (e.g., Bloom & Reenen, 2010). In turns out that we can adapt the above framework to allow firms to make decisions which might influence productivity. Assume now that the productivity is given by

$$A_t^* = h\left(A_{t-1}^*, R_{t-1}\right) + \varepsilon_t$$

where R_t is an additional control variable for the firm. In Doraszelski & Jaumandreu (2013) this variable is R&D expenditure, in de Loecker (2013) it is the decision to export or not.

We will consider adapting the Olley-Pakes approach first.[2] Two changes are made. First, we can include R_0^{t-1}as components of the firm's beginning-of-period information set; this generates more instruments. Second, in step 2.a of our estimation procedure we now regress $A_{it}^*\left(\alpha^{(s)}, \beta^{(s)}, \hat{\phi}_t(\cdot)\right)$ onto *both* $A_{it-1}^*\left(\alpha^{(s)}, \beta^{(s)}, \hat{\phi}_{t-1}(\cdot)\right)$ *and* R_{it-1}.

For the dynamic panel data estimators we could work with the parametric productivity process

$$A_{it}^* = \lambda_t + R_{it-1}'\eta + \rho A_{it-1}^* + \varepsilon_t$$

[2] The introduction of a control variable into the productivity process, may involve complications to the Olley-Pakes invertibility result (cf., Buettner, 2004). I ignore this issue here.

and proceed using the same logic as before. We could also introduce a firm-specific component to the productivity process if desired.

This discussion also highlights a potential problem with *not* modeling decisions made by firms which plausibly influence productivity. If firms can make decisions which influence productivity, then the underlying Markov assumption

$$\Pr\left(A_t^* \leq a \middle| \mathcal{I}_t\right) = \Pr\left(A_t^* \leq a \middle| A_{t-1}^*\right)$$

will not hold. Instead we will have

$$\Pr\left(A_t^* \leq a \middle| \mathcal{I}_t\right) = \Pr\left(A_t^* \leq a \middle| A_{t-1}^*, X_{t-1}^A\right)$$

with X_{t-1}^A denoting any past actions by the firm which could plausibly increase productivity "today". If, in step 2.a of our estimation procedure, we fail to include firm choices which influence productivity, then we will erroneously be proceeding "as if"

$$\mathbb{E}\left[\left\{h\left(A_{t-1}^*, R_{t-1}\right) - \mathbb{E}\left[A_t^* \middle| A_{t-1}^*\right]\right\} + \varepsilon_t \middle| \mathcal{I}_t\right] = 0.$$

This condition is, of course, not true since R_{it-1} is an element of $\mathcal{I}_t$. A similar issue arises in the dynamic data panel data case.

Using the Olley-Pakes and/or dynamic panel data approaches to understanding how firm choices might influence productivity is attractive. At the same time this discussion illustrates how sensitive these methods are to the underlying assumption about the productivity process. The assumption that productivity is an exogenous first order Markov process is – in a very real sense – an identifying one. It may not be especially plausible is some settings.

References

Ackerberg, D., Chen, X., Hahn, J., & Liao, Z. (2014). Asymptotic efficiency of semiparametric two-step gmm. *Review of Economic Studies*, 81(3), 919 – 943.

Ai, C. & Chen, X. (2012). The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, 1970(2), 442 – 457.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.

Arellano, M. (2016). Modelling optimal instrumental variables for dynamic panel data models. *Research in Economics*, 70(2), 238 – 261.

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

Arellano, M. & Bover, O. (1995). Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics*, 68(1), 29 – 51.

Bloom, N. & Reenen, J. V. (2010). Why do management practices differ across firms and countries? *Journal of economic perspectives 24.1 (2010): 203-24.*, 24(1), 203 – 224.

Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115 – 143.

Blundell, R. & Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), 321 – 340.

Bond, S. & Söderbom, M. (2005). *Adjustment costs and the identification of Cobb Douglas production functions*. IFS Working Paper W05/04, Institute for Fiscal Studies.

Buettner, T. (2004). *The dynamics of firm profitability, growth, and exit*. PhD thesis, London School of Economics and Political Science.

Chamberlain, G. (1992). Comment: sequential moment restrictions in panel data. *Journal of Business and Economic Statistics*, 10(2), 20 – 26.

de Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, 5(3), 1 – 21.

de Loecker, J. & Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6), 2437 – 2471.

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

Griliches, Z. & Mairesse, J. (1998). *Econometrics and Economic Theory in the 20th Century*, chapter Production functions: the search for identification, (pp. 169 – 203). Cambridge University Press: Cambridge.

Mundlak, Y. (1961). Empirical production function free of management bias. *Journal of Farm Economics*, 43(1), 44 – 56.

Olley, G. S. & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6), 1263 – 1297.

Syverson, C. (2011). What determines productivity. *Journal of Economic Literature*, 49(2), 326 – 365.

