Olley and Pakes (1996, Ecta)

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 312, Spring 2023

- Alternative approach to estimating production functions.
- Key assumptions are timing/information set assumptions, a scalar unobservable assumption, and a monotonicity assumption.
- Setup: Cobb-Douglas

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \tag{1}$$

- Again, the unobserved productivity shocks ω_{it} are potentially correlated with k_{it} and l_{it}.
- But the unobservables ϵ_{it} are measurement errors or unforecastable shocks that are not correlated with inputs k_{it} and l_{it}.

- **Basic Idea:** Endogeneity problem is due to the fact that ω_{it} is unobserved by the econometrician.

- If some other equation can tell us what ω_{it} is (i.e. making it "observable"), then the endogeneity problem would be eliminated.

- Olley and Pakes will use observed investment decisions i_{it} to "tell us" about ω_{it}.

- Assumptions:

1) The productivity shock ω_{it} follows a first order markov process, i.e.

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

- I_{it} is information set.

- This is both an assumption on the stochastic process governing ω_{it} and an assumption on firms' information sets at various points in time.

- Essentially, firms are moving through time, observing ω_{it} at t, and forming expectations about future ω_{it} using $p(\omega_{it+1}|\omega_{it})$.

- The form of this first order markov process is completely general, e.g. it is more general than $\omega_{it} = \omega_{it}$ or ω_{it} following an AR(1) process.

- This assumption implies that

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

- Then we can write

$$\omega_{it+1} = g(\omega_{it}) + \xi_{it+1}$$

where by construction $E[\xi_{it+1}|I_{it}] = 0$.

- $g(\omega_{it})$ can be thought of as the "predictable" component of ω_{it+1}, ξ_{it+1} can be thought of as the "innovation" component, i.e. the part that the firm doesn't observe until $t+1$.

- This can be extended to higher order Markov processes (see ABBP Handbook article and Ackerberg and Hahn (2015)).

2) Labor is a perfectly variable input, i.e. l_{it} is chosen by the firm at time t (after observing ω_{it}).

3) Labor has no dynamic implications. In other words, my choice of l_{it} at t only affects profits at period t, not future profits.

4) This rules out, e.g. labor adjustment costs like firing or hiring costs.

5) On the other hand, k_{it} is accumulated according to a dynamic investment process. Specifically

$$K_{it} = \delta K_{it-1} + i_{it-1}$$

where i_{it} is the investment level chosen by the firm in period t (after observing ω_{it}).

- Importantly, note that kit depends on last period's investment, not current investment.
- The assumption here is that it takes full time period for new capital to be ordered, delivered, and installed.
- This also implies that kit was actually decided by the firm at time $t-1$.
- This is a "timing assumption".
- In summary:
  - labor is a variable (decided at t), non-dynamic input
  - capital is a fixed (decided at $t-1$), dynamic input
  - we could also think about including fixed, non-dynamic inputs, or variable, dynamic inputs. (see ABBP)

- Given this setup, lets think about a firm's optimal investment choice i_{it}. Given i_{it} will aspect future capital levels, a profit maximizing firm will choose i_{it} to maximize the PDV of its future
- profits.

- This is a dynamic programming problem, and will result in an dynamic investment demand function of the form:

$$i_{it} = f_t(k_{it}, \omega_{it}) \qquad (2)$$

- Note that:

- k_{it} and ω_{it} are part of the state space, but l_{it} does not enter the state space. Why?

- f_t is indexed by t. This implicitly allows investment decisions to depend on other state variables (e.g. input prices, demand conditions, industry structure) that are constant across firms.

- f_t will likely be a complicated function because it is the solution to a dynamic programming problem. Fortunately, we can estimate the production function parameters without actually solving this DP problem (this is helpful not only computationally, but also allows us to estimate the production function without having to specify large parts of the firms optimization problem (semiparametric)). This is a nice example of how semiparametrics can help in terms of computation - literature based on Hotz and Miller (1993, ReStud) is similar in nature.

- One of the key ideas behind OP is that under some conditions, the investment demand equation (2) can be inverted to obtain

$$\omega_{it} = f_t^{-1}(k_{it}, i_{it}) \tag{3}$$

i.e. we can write the productivity shock ω_{it} as a function of variables that are observed by the econometrician (though the function is unknown).

- What are these conditions/assumptions?

1) (strict monotonicity) f_t is strictly monotonic in ω_{it}. OP prove this formally under a set of assumptions that include the assumption that $p(\omega_{it+1}|\omega_{it})$ is stochastically increasing in ω_{it}. This result is fairly intuitive.

2) (scalar unobservable) ω_{it} is the only econometric unobservable in the investment equation, i.e.

- ❖ Essentially no unobserved input prices that vary across firms (if there were observed input prices that varied across firms, they could be included as arguments of ft). There is one exception to this - labor input price shocks across firms that are not correlated across time.

- ❖ No other structural unobservables that affect firms optimal investment levels (e.g efficiency at doing investment, heterogeneity in adjustment costs, other heterogeneity in the production function (e.g. random coefficients))

- ❖ No optimization or measurement error in i.

2) is a fairly strong assumption, but it is crucial to being able to write f_{it} as an (unknown) function of observables.

- Suppose these conditions hold. Substitute (3) into (1) to get

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + f_t^{-1}(k_{it}, i_{it}) + \epsilon_{it} \tag{4}$$

- Since we don't know the form of the function t (and it is a complicated solution to a dynamic programming problem), let's just treat it non-parametrically, e.g., a high order polynomial in i_{it} and k_{it}, e.g.

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \gamma_{0t} + \gamma_{1t} k_{it} + \gamma_{2t} i_{it} + \gamma_{3t} k_{it}^2 + \gamma_{4t} i_{it}^2 + \gamma_{5t} k_{it} i_{it} + \epsilon_{it} \tag{5}$$

- Main point is that under the OP assumptions, we have eliminated the unobservable causing the endogeneity problem

- In this literature, i_{it} is sometimes called a control variable and sometimes called a proxy variable. Neither is perfect terminology.

- So we can think about estimating this equation with a simple OLS regression of y_{it} on k_{it}, l_{it}, and a polynomial in k_{it} and i_{it}:

- Problem: $\beta_1 k_{it}$ is collinear with the linear term in the polynomial, so we can't separately identify β_1 from γ_{it}. Intuitively, there is no way to separate out the effect of k_{it} on y_{it} through the production function, from the effect of k_{it}on y_{it} through f_t^{-1}.

- But, there is no l_{it} in the polynomial, so β_2 can in principle be identified (though see discussion of Ackerberg, Caves, and Frazer (ACF, 2015, Ecta) below).

- In summary, the "first stage" of OP involves OLS estimation of

$$y_{it} = \beta_2 l_{it} + \widetilde{\gamma}_{0t} + \widetilde{\gamma}_{1t} k_{it} + \gamma_{2t} i_{it} + \gamma_{3t} k_{it}^2 + \gamma_{4t} i_{it}^2 + \gamma_{5t} k_{it} i_{it} + \epsilon_{it} \quad (6)$$

where $\widetilde{\gamma}_{0t} = \beta_0 + \gamma_{0t}$ and $\widetilde{\gamma}_{1t} = \beta_1 + \gamma_{1t}$. This produces an estimate of the labor coefficient

$$\widehat{\beta}_2$$

and an estimate of the "composite" term $\beta_0 + \beta_1 k_{it} + \omega_{it}$

$$\widehat{\Phi}_{it} = \widehat{\widetilde{\gamma}}_{0t} + \widehat{\widetilde{\gamma}}_{1t} k_{it} + \widehat{\gamma}_{2t} i_{it} + \widehat{\gamma}_{3t} k_{it}^2 + \widehat{\gamma}_{4t} i_{it}^2 + \widehat{\gamma}_{5t} k_{it} i_{it} = \beta_0 + \widehat{\beta_1 k_{it}} + \omega_{it}$$

- To estimate the coefficient on capital, β_1, we need a "second stage".
- Recall that we can write

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it} \qquad \text{where } E[\xi_{it} | I_{it-1}] = 0$$

- Since k_{it} was decided at $t-1, k_{it} \in I_{it-1}$. Hence

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

and therefore

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

- This moment condition can be used to estimate the capital coefficient.

- More specifically, consider the following procedure:
  - 1) Guess a candidate β_1
  - 2) Compute

$$\widehat{\omega}_{it}(\beta_1) = \widehat{\Phi}_{it} - \beta_1 k_{it}$$

    for all i and t. $\widehat{\omega}_{it}(\beta_1)$ are the "implied" ω_{it}'s given the guess of β_1. If our guess is the true β_1, $\widehat{\omega}_{it}(\beta_1)$ will be the true ω_{it}'s (asymptotically). If our guess is not the true β_1, the $\widehat{\omega}_{it}(\beta_1)$'s **will not be** the true ω_{it}'s asymptotically. (Note: Actually, $\widehat{\omega}_{it}(\beta_1)$ is really $\omega_{it} + \beta_0$, but the constant term ends up not mattering)
  - 3) Given the implied $\widehat{\omega}_{it}(\beta_1)$'s, we now want to compute the implied *innovations* in ω_{it} i.e. implied ξ_{it}'s. To do this, consider the equation

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it}$$

Think about estimating this equation, i.e. non-parametrically regressing the implied $\widehat{\omega}_{it}(\beta_1)$'s (from step 2) on the implied $\widehat{\omega}_{it-1}(\beta_1)$'s (also from step 2). Again, we can think of representing g non-parametrically using a polynomial in $\widehat{\omega}_{it-1}(\beta_1)$. Call the residuals from this regression

$$\widehat{\xi}_{it}(\beta_1)$$

These are the implied innovations in ω_{it}. Again, if our guess is the true β_1, $\widehat{\xi}_{it}(\beta_1)$ will be the true ξ_{it}'s (asymptotically). If our guess is not the true β_1, then the $\widehat{\xi}_{it}(\beta_1)$'s **will not be** the true ξ_{it}'s.

– 4) Lastly, evaluate the sample analogue of the moment condition $E[\xi_{it}\ k_{it}] = 0$, i.e.

$$\frac{1}{N}\frac{1}{T}\sum_i \sum_t \widehat{\xi}_{it}(\beta_1)k_{it} = 0$$

- This is a version of the second stage of OP. It is essentially a non-linear GMM estimator

- Notes
  - 1) Recap of key assumptions:
    * First order markov assumption on ω_{it} (again can be relaxed to higher order (but Markov)) - note, for example, that the sum of two markov processes is not generally first order markov (e.g. sum of two AR(1) processes with different AR coefficients)
    * Timing assumptions on when inputs are chosen and information set assumptions regarding when the firm observes ω_{it} (this can be strengthened or relaxed - see Ackerberg (2016))
    * Strict monotonicity of investment demand in ω_{it} (can be relaxed to weak monotonicity - see below)
    * Scalar unobservable in investment demand (tough to relax, though one can allow other **observables** to enter investment demand, e.g. input prices)

– 2) Alternative formulation of the second stage (more like OP paper)

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \quad (7)$$

$$y_{it} - \beta_1 k_{it} - \beta_2 l_{it} = \beta_0 + g(\omega_{it-1}) + \xi_{it} + \epsilon_{it} \quad (8)$$

$$y_{it} - \beta_1 k_{it} - \widehat{\beta}_2 l_{it} = \beta_0 + g(\widehat{\Phi}_{it-1} - \beta_0 - \beta_1 k_{it-1}) + \xi_{it} + \epsilon_{it} \quad (9)$$

$$y_{it} - \beta_1 k_{it} - \widehat{\beta}_2 l_{it} = g(\widehat{\Phi}_{it-1} - \beta_1 k_{it-1}) + \xi_{it} + \epsilon_{it} \quad (10)$$

So given a guess of β_1, one can regress $\left(y_{it} - \beta_1 k_{it} - \widehat{\beta}_2 l_{it}\right)$ on a polynomial in $(\widehat{\Phi}_{it-1} - \beta_1 k_{it-1})$ to recover implied $\xi_{it} + \epsilon_{it}$'s, i.e. $\widehat{\xi_{it} + \epsilon_{it}}(\beta_1)$, and then use the moment condition

$$E\left[(\xi_{it} + \epsilon_{it})\, k_{it}\right] = 0$$

and sample analogue

$$\frac{1}{N}\frac{1}{T}\sum_i \sum_t \left(\widehat{\xi_{it} + \epsilon_{it}}(\beta_1)\right) k_{it} = 0$$

to estimate β_1.

- 3) There are other formulations as well. For example, Wooldridge (2009, EcLet) suggests estimating both first stage and second stage simultaneously. This has two potential advantages: 1) efficiency (though this is not always the case, see, e.g. Ackerberg, Chen, Hahn, and Liao (2014, ReStud) , and 2) it makes it easier to compute standard errors (with two-step procedure, it is typically easiest to bootstrap). On the other hand, a disadvantage is that it requires a non-linear search over a larger set of parameters (β_1 plus the parameters of g and f_t^{-1}), whereas the above two step formulations only require a non-linear search for β_1 (or β_1 and g)
- 4) Note that there are additional moments generated by the model. The assumptions of the model imply that $E[\xi_{it}| \; I_{it-1}] = 0$. This means that the implied ξ_{it}'s should not only be uncorrelated with k_{it}, but everything else in I_{it-1}, e.g. k_{it-1}, k_{it-2}, l_{it-1},k_{it}^2..... (though not l_{it}). These additional moments can potentially add efficiency, but also result in an overidentified model, which can lead to small sample bias. The extent to which one utilizes these additional moments is typically a matter of taste.

– 5) Intuitive description of identification

* First stage: Compare output of firms with same i_{it} and k_{it} (which imply the same ω_{it}) , but different l_{it}. This variation in l_{it} is uncorrelated with the remaining unobservables determining y_{it} (ϵ_{it}), and so it identifies the labor coefficient. (But again, see ACF section below)
* Second stage: Compare output of firms with same ω_{it-1}, but different k_{it}'s (note that firms can have the same ω_{it-1}, but different i_{it-1} and k_{it-1}).

$$\begin{aligned} y_{it} - \widehat{\beta}_2 l_{it} &= \beta_0 + \beta_1 k_{it} + g(\omega_{it-1}) + \xi_{it} + \epsilon_{it} \\ &= \beta_0 + \beta_1 k_{it} + g(\widehat{\Phi}_{it-1} - \beta_0 - \beta_1 k_{it-1}) + \xi_{it} + \epsilon_{it} \end{aligned}$$

This variation in k_{it} is uncorrelated with the remaining unobservables determining y_{it} (ξ_{it} and ϵ_{it}), so it identifies the capital coefficient (However, note that the "comparison of firms with same ω_{it-1}" depends on the parameters themselves, so this is not completely transparent intuition)

– 6) OP also deal with a selection problem due to the fact that unproductive firms may exit the market. The problem is that even if

$$E\left[\xi_{it}\middle|\, k_{it}\right] = 0$$

in the entire population of firms,

$$E\left[\xi_{it}\middle|\, k_{it},\ \text{still in sample at } t\right] \qquad \text{may not equal 0 and be a function of } k_{it}$$

Specifically, if a firm's exit decision at t depends on ω_{it} (and thus ξ_{it}), then this second expectation is likely > 0 and depends negatively on k_{it} (since firms with higher k_{it}'s may be more apt to stay in the market for a given ω_{it} or ξ_{it}). OP develop a selection correction to correct for this, which I dont think I will go through (see ABBP for discussion). On the other hand, if exit decisions at t are made at time $t-1$ (a timing assumption like that already being made on capital), then there is no selection problem, since in this case the exit decision is just a function of I_{it-1}.