# 1 Lecture Notes - Production Functions - 1/5/2017 D.A.

# 2 Introduction

- Production functions are one of the most basic components of economics
- They are important in themselves, e.g.
  - What is the level of returns to scale?
  - How do input coefficients on capital and labor change over time?
  - How does adoption of a new technology affect production?
  - How much heterogeneity is there in measured productivity across firms, and what explains it?
  - How does the allocation of firm inputs relate to productivity
- Also can be important as inputs into other interesting questions, e.g. dynamic models of industry evolution, evaluation of firm conduct (e.g. collusion)
- For this lecture note, we will work with a simple two input Cobb-Douglas production function

$$Y_i = e^{\beta_0} K_i^{\beta_1} L_i^{\beta_2} e^{\varepsilon_i}$$

  where $i$ indexes firms, $K_i$ is units of capital, $L_i$ is units of labor, and $Y_i$ is units of output. $(\beta_0, \beta_1, \beta_2)$ are parameters and $\varepsilon_i$ captures unobservables that affects output (e.g. weather, soil quality, management quality)
- Take natural logs to get:

$$y_i = \beta_0 + \beta_1 k_i + \beta_2 l_i + \varepsilon_i$$

- This can be extended to
  - Additional inputs, e.g. R&D (knowledge capital), dummies representing discrete technologies, different types of labor/capital, intermediate inputs.
  - Later we will see more flexible models

$$y_i = f(k_i, l_i; \beta) + \varepsilon_i$$

$$y_i = f(k_i, l_i, \varepsilon_i; \beta) \quad \text{(with scalar/monotonic } \varepsilon_i\text{)}$$

$$y_i = \beta_0 + \beta_{1i} k_i + \beta_{2i} l_i + \varepsilon_i$$

# 3 Endogeneity Issues

- Problem is that inputs $k_i$, $l_i$ are typically choice variables of the firm. Typically, these choices are made to maximize profits, and hence will often depend on unobservables $\varepsilon_i$.
- Of course, this dependence depends on what the firm knows about $\varepsilon_i$ when they make these input choices.

- Example: Suppose a firm operating in perfectly competitive output and input markets (with respective prices $p_i, r_i$, and $w_i$) perfectly observes $\varepsilon_i$ before optimally choosing inputs. Profit maximization problem is:

$$\max_{K_i, L_i} \quad p_i e^{\beta_0} K_i^{\beta_1} L_i^{\beta_2} e^{\varepsilon_i} - r_i K_i - w_i L_i$$

- FOC's will imply that optimal choices of $K_i$ and $L_i$ ($k_i$ and $l_i$) will depend on $\varepsilon_i$. Intuition: $\varepsilon_i$ positively affects marginal product of inputs. Hence firms with higher $\varepsilon_i$'s will want to use more inputs.

- As a result, one cannot estimate

$$y_i = \beta_0 + \beta_1 k_i + \beta_2 l_i + \varepsilon_i$$

  using OLS because $k_i$ and $l_i$ are correlated with $\varepsilon_i$. Generally one would expect coefficients to be positively biased.

- Similar problems would arise in more complicated models (e.g. non-perfectly competitive output or input markets, $\varepsilon_i$ only partially observed), except the special case where the firm has no knowledge of $\varepsilon_i$ when choosing inputs

- If $k_i$ is a "less variable" input than $l_i$, one might expect the firm to have less knowledge about $\varepsilon_i$ when choosing $k_i$ (relative to $l_i$). Generally, this will imply $k_i$ will be less correlated with $\varepsilon_i$ than $l_i$ is. So one might expect more bias in the labor coefficient.

- Note: we will generally assume that the unobservables $\varepsilon_i$ are generated or evolve exogenously, i.e. they are not choice variables of the firm. Things get considerably harder when the unobservables are choice variables of the firm.

- WLOG, lets think about $\varepsilon_i$ having two components, i.e.

$$y_i = \beta_0 + \beta_1 k_i + \beta_2 l_i + \omega_i + \epsilon_i$$

  where $\omega_i$ is an unobservable that is predictable (or partially predictable) to the firm when it makes its input decisions, and $\epsilon_i$ is an unobservable that the firm has no information about when making input decisions (e.g. $\omega_i$ represents average weather conditions on a particular farm, $\epsilon_i$ represents deviations from that average in a given year (after inputs are chosen)). $\epsilon_i$ could also represent measurement error in output.

- In this formulation, $\omega_i$ is causing the endogeneity problem, not $\epsilon_i$. Let's call $\omega_i$ the "productivity shock".

# 4 Traditional Solutions

- Two traditional solutions to endogeneity problems can be used here: instrumental variables and fixed effects model. I will discuss these before moving to more recent methodological approaches.

### 4.1 Instrumental Variables

- Want to find "instruments" that are correlated with the endogenous inputs, but do not directly determine $y_i$ and are not correlated with $\omega_i$ (and $\epsilon_i$).

- Good news is that theory can provide us with such instruments.

- Specifically, consider input and output prices $w_i$, $r_i$, and $p_i$.Theory tells us that these prices will affect firms' optimal choices of $k_i$ and $l_i$. Theory also says that these prices are excluded from the production function as they do not directly determine output $y_i$ conditional on the inputs.

- Last requirement is that $w_i$, $r_i$, and $p_i$ are not correlated with the productivity shock $\omega_i$. When will this be the case (or not be the case)?

- One key issue is the form of competition in input and output markets.

  - If output markets are imperfectly competitive (i.e. firms face downward sloping demand curves), then a higher $\omega_i$ will increase a firm's output, driving $p_i$ down. In other words, $p_i$ will be positively correlated with $\omega_i$, invalidating $p_i$ as an instrument.
  - If input markets are imperfectly competitive (i.e. firms face upward sloping supply curves), then a higher $\omega_i$ will increase a firm's input demand, driving $w_i$ and/or $r_i$ up. So $w_i$ and/or $r_i$ are now correlated with $\omega_i$, invalidating them as instruments.

- So for these instruments want firms operating in perfectly competitive input or output markets. Typically, this is more believable for input markets than for output markets.

- Unfortunately, even if willing to make these assumptions, IV solutions haven't been that broadly used in practice. First, one needs data on $w_i$ and $r_i$. Second, there is often very little variation in $w_i$ and $r_i$ across firms (often there is a real question of whether firms actually operate in different input markets?). Third, one often wonders whether observed variation in e.g. $w_i$, actually represents firms facing different input prices, or whether it represents things like variation in unobserved labor quality (i.e. the firm with the higher $w_i$ is employing workers of higher quality). If the latter, then $w_i$ is not a valid instrument.

- While there might be "true" variation in input prices across time, this is usually not helpful, because if one has data across time, one often wants to allow the production function to change across time, e.g.
$$y_{it} = \beta_{0t} + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it}$$
(though there could be exceptions)

- That said, I think if one can find a market where there is convincing exogenous input price variation, IV approach is probably more convincing than the approaches I will talk about in the rest of this lecture note, as there seem to be less auxiliary assumptions.

- Notes:

  - Randomized experiments - either directly manipulating inputs, or manipulating input prices.

- As is typically done in this literature, I have implicitly made a "homogeneous treatment effects" assumption. A heterogeneous treatment effects model would be

$$y_i = \beta_0 + \beta_{1i} k_i + \beta_{2i} l_i + \omega_i + \epsilon_i$$

  This affects the interpretation of IV estimators, e.g. Heckman and Robb (1985), Angrist and Imbens (1994, Ecta)

- If there are unobserved firm choice variables in $\omega_i$, it becomes quite hard to find valid instruments, even with the above assumptions.

### 4.2 Fixed Effects

- This approach relies on having panel data on firms across time, i.e.

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

- Assume that $\epsilon_{it}$ is independent across $t$ (this is consistent with $\epsilon_{it}$ not being predictable by the firm when choosing $k_{it}$ and $l_{it}$)

- Suppose one is willing to assume that the productivity shock is contant over time (fixed effect assumption), i.e.

$$\omega_{it} = \omega_i$$

- Then one can either mean difference

$$y_{it} - \overline{y}_i = \beta_1 \left(k_{it} - \overline{k}_i\right) + \beta_2 \left(l_{it} - \overline{l}_i\right) + \left(\epsilon_{it} - \overline{\epsilon}_i\right)$$

  or first difference

$$y_{it} - y_{it-1} = \beta_1 \left(k_{it} - k_{it-1}\right) + \beta_2 \left(l_{it} - l_{it-1}\right) + \left(\epsilon_{it} - \epsilon_{it-1}\right)$$

  Since the problematic unobservable $\omega_{it}$ have been differenced out of these expressions (recall that we have assumed that the $\epsilon_{it}$'s are uncorrelated with input choices) these equations can be estimated with OLS.

- Problems:

  - 1) $\omega_{it} = \omega_i$ is a strong assumption
  - 2) These estimators often produce strange estimates. In particular, they often generate very small (or even negative) capital coefficients. Perhaps this is due to measurement error in capital (Griliches and Hausman (1986, JoE))?

- Other notes:

  - The mean difference approach requires all the input choices to be uncorrelated with all the $\epsilon_{it}$ (strict exogeneity). The first difference approach only requires current and lagged inputs to be uncorrelated with current and lagged $\epsilon_{it}$. Using $k_{it-1}$ and $l_{it-1}$ (or other lags) as instruments for $(k_{it} - k_{it-1})$ and $(l_{it} - l_{it-1})$, one can allow current inputs to be arbitrarily correlated with past $\epsilon_{it}$'s (sequential exogeneity)

- Panel data approach can be extended to richer error structures (Arellano and Bond (1991, ReStud), Arellano and Bover (1995. JoE), Blundell and Bond (1998, JoE, 2000, ER), Arellano and Honore (2001, Handbook)) e.g.

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

or

$$\omega_{it} = \alpha_i + \lambda_{it} \qquad \text{where } \lambda_{it} = \rho\lambda_{it-1} + \xi_{it}$$

I will talk further about these these later.

### 4.3 First Order Conditions

- A third approach to estimating production functions is based on information in first order conditions of optimizing firms.

- For example, for a firm operating in perfectly competitive input and output markets, static cost minimization implies that

$$\begin{aligned} \frac{\partial Y}{\partial L}\frac{L}{Y} &= \frac{wL}{pY} \\ \frac{\partial Y}{\partial K}\frac{K}{Y} &= \frac{rK}{pY} \end{aligned}$$

i.e. the output elasticity w.r.t. an input must equal its (cost) share in revenue.

- In a Cobb-Douglas context, these output elasticities are the production function coefficients $\beta_1$ and $\beta_2$, so observations on these revenue shares across firms could provide estimates of the coefficients.

- Note that $r$ can often be assumed known and often one directly observes $wL$ and $pY$ (rather than $L$ and $Y$ - i.e. labor input and output are measured in terms of dollar units (that are implicitly assumed to be comparable across firms))

- But:

  - This assumes static cost minimization - i.e. it assumes away dynamics, adjustment costs, etc.. At the very least we often think about the capital input being subject to a dynamic accumulation process, e.g. $K_{it} = \delta K_{it-1} + i_{it-1}$
  - There are additional terms when firms are not operating in perfectly competitive markets, e.g. when firms face downward sloping demand curve

$$\begin{aligned} \frac{\partial Y}{\partial L}\frac{L}{Y} &= \mu\frac{wL}{pY} \\ \frac{\partial Y}{\partial K}\frac{K}{Y} &= \mu\frac{rK}{pY} \end{aligned}$$

  where $\mu = \frac{p}{mc}$,i.e. percentage markup. Note that profit maximization implies $\frac{p}{mc} = \frac{\epsilon}{1+\epsilon}$, where $\epsilon$ is the elasticity of demand. So, for example, one could still identify production coefficients using this method if the elasticity of demand was known (this is done in Hsieh and Klenow (2009, QJE)). Or, one might be able to identify both with additional restrictions, e.g. Constant Returns to Scale (related to Hall (1988, JPE)).

## 5 Olley and Pakes (1996, Ecta)

- Alternative approach to estimating production functions. I will argue that key assumptions are timing/information set assumptions, a scalar unobservable assumption, and a monotonicity assumption.
- Setup:
$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \tag{1}$$
Again, the unobserved productivity shocks $\omega_{it}$ are potentially correlated with $k_{it}$ and $l_{it}$.but the unobservables $\epsilon_{it}$ 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 $\omega_{it}$ is unobserved by the econometrician. If some other equation can tell us what $\omega_{it}$ is (i.e. making it "observable"), then the endogeneity problem would be eliminated.
- Olley and Pakes will use observed firms' investment decisions $i_{it}$ to "tell us" about $\omega_{it}$.
- Assumptions:
- 1)The productivity shock $\omega_{it}$ follows 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 $t$ (which includes current and past $\omega_{it}$'s). Note:
  - This is both an assumption on the stochastic process governing $\omega_{it}$ **and** an assumption on firms' information sets at various points in time. Essentially, firms are moving through time, observing $\omega_{it}$ at $t$, and forming expectations about future $\omega_{it}$ 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 $\omega_{it} = \omega_i$ or $\omega_{it}$.following an AR(1) process.
  - This assumption implies that
  $$E\left[\omega_{it+1}|\ I_{it}\right] = g(\omega_{it})$$
  and that we can write
  $$\omega_{it+1} = g(\omega_{it}) + \xi_{it+1} \qquad \text{where by construction } E\left[\xi_{it+1}|\ I_{it}\right] = 0$$
  - $g(\omega_{it})$ can be thought of as the "predictable" component of $\omega_{it+1}$, $\xi_{it+1}$ 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 $\omega_{it}$).
- 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. This rules out, e.g. labor adjustment costs like firing or hiring costs.

- 4) 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 $\omega_{it}$). Importantly, note that $k_{it}$ 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 $k_{it}$ was actually decided by the firm at time $t-1$. This is what I refer to as 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 affect 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:

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

Note that:

  - $k_{it}$ and $\omega_{it}$ 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 $\omega_{it}$ 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 $\omega_{it}$. OP prove this formally under a set of assumptions that include the assumption that $p(\omega_{it+1}|\omega_{it})$ is stochastically increasing in $\omega_{it}$. This result is fairly intuitive.

- 2) (scalar unobservable) $\omega_{it}$ 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 $f_t$). 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 $\omega_{it}$ 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 $f_t^{-1}$ (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 $\beta_1$ from $\gamma_{1t}$. 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 $\beta_2$ 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

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

  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, $\beta_1$, we need a "second stage".
- Recall that we can write

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

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

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

  and therefore

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

  This moment condition can be used to estimate the capital coefficient

- More specifically, consider the following procedure:
  - 1) Guess a candidate $\beta_1$
  - 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" $\omega_{it}$'s given the guess of $\beta_1$. If our guess is the true $\beta_1$, $\widehat{\omega}_{it}(\beta_1)$ will be the true $\omega_{it}$'s (asymptotically). If our guess is not the true $\beta_1$, the $\widehat{\omega}_{it}(\beta_1)$'s **will not be** the true $\omega_{it}$'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 $\omega_{it}$ i.e. implied $\xi_{it}$'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 $\omega_{it}$. Again, if our guess is the true $\beta_1$, $\widehat{\xi}_{it}(\beta_1)$ will be the true $\xi_{it}$'s (asymptotically). If our guess is not the true $\beta_1$, then the $\widehat{\xi}_{it}(\beta_1)$'s **will not be** the true $\xi_{it}$'s.
  - 4) Lastly, evaluate the sample analogue of the moment condition $E\left[\xi_{it}\ k_{it}\right] = 0$, i.e.

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

    Since $E\left[\xi_{it}\ k_{it}\right] = 0$, this sample analogue should be approximately zero if we have guessed the true $\beta_1$. For other $\beta_1$, this will generally not equal zero (identification)
  - 5) Use a computer to do a non-linear search for the $\widehat{\beta}_1$ that sets

$$\frac{1}{N}\frac{1}{T}\sum_i \sum_t \widehat{\xi}_{it}(\widehat{\beta}_1) k_{it} = 0$$

- 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 $\omega_{it}$ (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 $\omega_{it}$ (this can be strengthened or relaxed - see Ackerberg (2016))
    * Strict monotonicity of investment demand in $\omega_{it}$ (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)

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} && (7) \\ y_{it} - \beta_1 k_{it} - \beta_2 l_{it} &= \beta_0 + g(\omega_{it-1}) + \xi_{it} + \epsilon_{it} && (8) \\ y_{it} - \beta_1 k_{it} - \widehat{\beta}_2 l_{it} &= \beta_0 + g(\widehat{\Phi}_{it-1} - \beta_0 - \beta_1 k_{it-1}) + \xi_{it} + \epsilon_{it} && (9) \\ y_{it} - \beta_1 k_{it} - \widehat{\beta}_2 l_{it} &= g(\widehat{\Phi}_{it-1} - \beta_1 k_{it-1}) + \xi_{it} + \epsilon_{it} && (10) \end{aligned}$$

    So given a guess of $\beta_1$, 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

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

    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 $\beta_1$.
  - 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 ($\beta_1$ plus the parameters of $g$ and $f_t^{-1}$), whereas the above two step formulations only require a non-linear search for $\beta_1$ (or $\beta_1$ and $g$)
  - 4) Note that there are additional moments generated by the model. The assumptions of the model imply that $E\left[\xi_{it} | \ I_{it-1}\right] = 0$. This means that the implied $\xi_{it}$'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 $\omega_{it}$) , but different $l_{it}$. This variation in $l_{it}$ is uncorrelated with the remaining unobservables determining $y_{it}$ ($\epsilon_{it}$), and so it identifies the labor coefficient. (But again, see ACF section below)
  - Second stage: Compare output of firms with same $\omega_{it-1}$, but different $k_{it}$'s (note that firms can have the same $\omega_{it-1}$, 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 $y_{it}$ ($\xi_{it}$ and $\epsilon_{it}$), so it identifies the capital coefficient (However, note that the "comparison of firms with same $\omega_{it-1}$" 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

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

  in the entire population of firms,

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

  Specifically, if a firm's exit decision at $t$ depends on $\omega_{it}$ (and thus $\xi_{it}$), 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 $\omega_{it}$ or $\xi_{it}$). 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 $I_{it-1}$.

# 6 Levinsohn and Petrin (2003, ReStud)

- Levinsohn and Petrin worry about the assumption that investment is strictly monotonic in $\omega_{it}$. Intuitively, this assumption implies that any two firms with the same $k_{it}$ and $i_{it}$ must have the same $\omega_{it}$.

- But in many datasets, especially in developing countries, $i_{it}$ is often 0 (e.g. in LP's Chilean dataset, approximately 50% of observations have 0 investment)

- It seems like a strong assumption that all these firms have the same $\omega_{it}$ (given $k_{it}$). It seems more likely that there is some threshold $\underline{\omega}_{it}$ below which firms invest 0.

- One can extend OP to allow weak monotonicity for the observations where $i_{it} = 0$, but this requires discarding these observations from the analysis (Aside: in this case, there is no selection issue as long as one uses the second stage moment $E\left[(\xi_{it} + \epsilon_{it})\, k_{it}\ \right] = 0$ rather than $E\left[\xi_{it} k_{it}\ \right] = 0$ (see Gandhi, Navarro and Rivers (GNR, 2015)). This is because one cannot compute implied $\xi_{it}$'s for observations for which $i_{it} = 0$ (but one can compute implied $(\xi_{it} + \epsilon_{it})$ for these observations))

- Anyway, given these problems with 0 investment and an unwillingness to throw away data, the basic idea of LP is to use a different "control" variable to learn about $\omega_{it}$, one that is more likely to be strictly monotonic in $\omega_{it}$. They use an intermediate input, e.g. inputs like materials, fuel, or electricity. These types of inputs rarely take the value 0.

- Production Function:

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 m_{it} + \omega_{it} + \epsilon_{it} \tag{11}$$

  where $m_{it}$ is an intermediate input. $m_{it}$ is assumed to be a variable, non-dynamic input, like labor.

- Consider a firm's optimal choice of $m_{it}$. Like investment in OP, $m_{it}$ will be chosen as a function of the state variables $k_{it}$ and $\omega_{it}$, i.e.

$$m_{it} = f_t(k_{it}, \omega_{it}) \tag{12}$$

- Assuming strict monotonicity, this can be inverted and substituted into the production function

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 m_{it} + f_t^{-1}(k_{it}, m_{it}) + \epsilon_{it} \tag{13}$$

- The rest follows exactly as in OP
  - Estimate $\widehat{\beta}_2$ in first stage ($\beta_1$ and $\beta_3$ cannot be identified because they are in $f_t^{-1}$)
  - Estimate $\widehat{\beta}_1$ and $\widehat{\beta}_3$ in second stage (Need additional moment here to identify the second parameter. LP use. $E\left[(\xi_{it} + \epsilon_{it})\, m_{it-1}\right] = 0$ or $E\left[\xi_{it} m_{it-1}\right] = 0$, though see Bond and Soderbom (2005) and GNR )

# 7 Ackerberg, Caves, and Frazer (2015)

## 7.1 Critique

- This paper examines the first stage of LP and OP
- Our question: Under what conditions is the labor coefficient $\widehat{\beta}_2$ identified in the first stage?
- The LP first stage regresses $y_{it}$ on $l_{it}$ and a non-parametric function of $k_{it}$ and $m_{it}$. (call this non-parametric function $np(k_{it}, m_{it})$)

$$y_{it} = \beta_2 l_{it} + np(k_{it}, m_{it}) + \epsilon_{it} \tag{14}$$

- There is no endogeneity problem here (since $\epsilon_{it}$ assumed uncorrelated with everything). But our question is whether the labor input $l_{it}$ moves around independently of $np(k_{it}, m_{it})$. In other words, can two firms with the same $k_{it}$ and $m_{it}$ have different $l_{it}$? To analyze this, we need to think about a model of how firms chose $l_{it}$

- Most natural model seems as follows. Since $m_{it}$ and $l_{it}$ are both non-dynamic, variable inputs, and since LP have already assumed that

$$m_{it} = f_t(k_{it}, \omega_{it}) \tag{15}$$

it seems logical to treat $l_{it}$ symmetrically and assume

$$l_{it} = h_t(k_{it}, \omega_{it}) \tag{16}$$

Of course, these will be different functions.

- If this is the case, then note that

$$\begin{aligned} l_{it} &= h_t(k_{it}, \omega_{it}) \\ &= h_t(k_{it}, f_t^{-1}(k_{it}, m_{it})) \\ &= \widetilde{h}_t(k_{it}, m_{it}) \end{aligned}$$

The last line implies that $l_{it}$ is a *deterministic* function of $k_{it}$ and $m_{it}$

- But this is a problem for the first stage estimating equation

$$y_{it} = \beta_2 l_{it} + np(k_{it}, m_{it}) + \epsilon_{it} \tag{17}$$

since it implies that $l_{it}$ is functionally dependent on ("collinear" with) $np(k_{it}, m_{it})$, i.e. $l_{it}$ doesn't move independently of $np(k_{it}, m_{it})$.

- Another way of saying this is as follows: LP want to condition on $k_{it}$ and $m_{it}$ (i.e. condition on $\omega_{it}$) and look at remaining variation in $l_{it}$ to identify $\beta_2$. But according to the above, $l_{it}$ is a deterministic function of $k_{it}$ and $m_{it}$. Hence, there is no remaining variation in $l_{it}$ once we condition on $k_{it}$ and $m_{it}$!!!

- Can also think about both $\widetilde{h}_t(k_{it}, m_{it})$ and $np(k_{it}, m_{it})$ being polynomials.

- So if (16) is correct, then $\beta_2$ **should not** be identified in the first stage. If OLS does in fact produce an estimate of $\widehat{\beta}_2$, then some assumption of the model must be incorrect.

- To get the first stage of LP to work, we need to find something that moves around $l_{it}$ independently of $np(k_{it}, m_{it})$. Unfortunately, this is hard to do within the context of LP's other maintained assumptions. For example, suppose one assumes that there is some firm-specific unobserved shock to the price of labor, $v_{it}$. This will clearly affect firms' optimal labor choices, i.e..

$$l_{it} = h_t(k_{it}, \omega_{it}, v_{it})$$

The problem is that this will also generally affect the firms' optimal choice of materials

$$m_{it} = f_t(k_{it}, \omega_{it}, , v_{it}) \tag{18}$$

which then violates the scalar unobservable assumption necessary to invert $f_t$ and write $\omega_{it}$ as a function of observables.

- Our paper thinks about various alternative models of $l_{it}$ (i.e. various data-generating processes (DGPs)) that might "break" this functional dependence problem. We can only come up with 2 such DGPs, and neither seems all that general.

- 1) Suppose there is "optimization" error in $l_{it}$, i.e.

$$l_{it} = h_t(k_{it}, \omega_{it}) + u_{it}$$

where $u_{it}$ is independent of $(k_{it}, \omega_{it})$. In other words i.e. for some exogenous reason firms do not get the optimal choice of labor correct. This breaks the functional dependence problem (and does not seem completely unreasonable). On the other hand, one simultaneously needs to assume that there is 0 optimization error in $m_{it}$ (otherwise, the scalar unobservable assumption is violated). It seems challenging to argue that there is a significant amount of optimization error in $l_{it}$, but almost no optimization error in $m_{it}$. (One example might be if one's data measures "planned" or ordered materials, but actual labor (e.g. subject to sick days or unexpected quits))

- 2) Suppose that $m_{it}$ is chosen at some point in time **prior** to $l_{it}$, and that:
  - a) The firm knows $\omega_{it}$ when choosing $m_{it}$
  - b) Between these two points in time, there is a shock to the price of labor, $v_{it}$, that varies across firms.
  - c) $v_{it}$ is independent across time (and other variables in the model)

- This second DGP also allows the labor coefficient to be identified in the first stage, because the shock $v_{it}$ moves $l_{it}$ around conditional on $m_{it}$ and $k_{it}$. Note that $v_{it}$ needs to be independent across time, otherwise the choice of $m_{it}$ at $t$ will optimally depend on $v_{it-1}$, violating the scalar unobservable assumption needed for invertibility.

- Again, this DGP does not seem very general. Why is $m_{it}$ chosen before $l_{it}$ (if anything, I would tend to think the reverse)? And it seems like a stretch to assume that there are no unobserved firm specific input price shocks except for this very special $v_{it}$ shock that must be realized between these two points in time.

- Notes:

- 1) Parametric treatment of the intermediate input demand function does not rescue the LP first stage identification - see ACF for details (though unlike with $i_{it}$ it is not hard to do this, and it likely adds efficiency)

- 2) The OP estimator is also affected by this critique. However, there is a 3rd DGP that breaks the functional dependence problem. This involves $l_{it}$ being chosen with incomplete knowledge of $\omega_{it}$, e.g. prior to the realization of $\omega_{it}$ - see ACF for details.

- 3)Bond and Soderbom (2005) make a related argument that criticizes the second stage identification of $\beta_3$ (the coefficient on the intermediate input) in LP. The crux of the implications of their argument is that under the assumptions of the LP model the moment condition $E\left[(\xi_{it} + \epsilon_{it})\, m_{it-1}\ \right] = 0$ or $E\left[\xi_{it} m_{it-1}\ \right] = 0$ is not informative about the coefficient $\beta_3$. The intuition is that under the assumptions of the LP model, $m_{it-1}$ is not correlated with $m_{it}$ conditional on $k_{it}$ and $\omega_{it-1}$- hence it is uninformative as an instrument. A serially correlated, firm-specific, unobserved price shock to the intermediate input could generate such correlation, but it violates the LP scalar unobservable assumption. More generally, Bond and Soderbom show that without firm specific input price variation, coefficients on perfectly variable, non-dynamic inputs are not identified in Cobb-Douglas production functions. GNR extend this argument to more general production functions, essentially showing that if a perfectly variable, non-dynamic input is being used as a proxy/control variable in the context of an OP/LP like procedure, its effect on output cannot generally be identified using the above moments (and they argue that FOC approaches are needed for these inputs, see below).

### 7.2 Alternative estimator suggested by ACF

- Given that we feel these identification arguments rely on very particular data generating processes, we suggest a slightly different approach where we abandon trying to identify the labor coefficient in the first stage. Instead, we try to identify $\beta_2$ along with the capital coefficient in the second stage.

- To do this, assume that $m_{it}$ is chosen either at the same time or after $l_{it}$ is chosen. This implies we can write
$$m_{it} = f_t(k_{it}, \omega_{it}, l_{it}) \tag{19}$$

- This can be thought of as a "conditional" (on $l_{it}$) intermediate input demand function, in contrast to LP's "unconditional" (on $l_{it}$) intermediate input demand function:

$$\begin{array}{rrcl} & \text{Unconditional interm. input demand (LP)} & : & m_{it} = f_t(k_{it}, \omega_{it}) \\ \text{vs.} & \text{Conditional interm. input demand (ACF)} & : & m_{it} = f_t(k_{it}, \omega_{it}, l_{it}) \end{array}$$

- Assuming strict monotonicity, this conditional intermediate input demand function can be inverted and substituted into the production function to get

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + f_t^{-1}(k_{it}, m_{it}, l_{it}) + \epsilon_{it} \tag{20}$$

- Treating $f_t^{-1}$ non-parametrically, it is obvious that now not even $\beta_2$ can be identified in the first stage. However, we can still identify the composite term:

$$\widehat{\Phi}_{it} = \beta_0 + \beta_1 k_{it} \widehat{+} \beta_2 l_{it} + \omega_{it}$$

(these are just the predicted values of $y_{it}$ from the regression)

- Just like in OP, given guesses of $\beta_1$ and $\beta_2$, we can a) compute implied $\widehat{\omega}_{it}(\beta_1, \beta_2)$'s, then b) regress the $\widehat{\omega}_{it}(\beta_1, \beta_2)$'s on the $\widehat{\omega}_{it-1}(\beta_1, \beta_2)$'s to obtain implied $\widehat{\xi}_{it}(\beta_1, \beta_2)$'s (the residuals from the regression). and c) compute the sample moment

$$\frac{1}{N}\frac{1}{T}\sum_i \sum_t \widehat{\xi}_{it}(\beta_1, \beta_2) k_{it} = 0$$

- But since there is an additional parameter to estimate in the second stage, we need another moment condition. We suggest

$$\frac{1}{N}\frac{1}{T}\sum_i \sum_t \widehat{\xi}_{it}(\beta_1, \beta_2) l_{it-1} = 0$$

which should be approximately zero at the true $\beta_1$ and $\beta_2$ since $l_{it-1} \in I_{it-1}$ and hence $E\left[\xi_{it} l_{it-1}\right] = 0$.

- So in summary, the second stage estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are defined by

$$\frac{1}{N}\frac{1}{T}\sum_i\sum_t \begin{pmatrix} \widehat{\xi}_{it}(\widehat{\beta}_1,\widehat{\beta}_2)k_{it} \\ \widehat{\xi}_{it}(\widehat{\beta}_1,\widehat{\beta}_2)l_{it-1} \end{pmatrix} = 0$$

- Alternatively, if we are willing to assume that like capital, $l_{it}$ is "fixed" and decided at $t-1$ (and thus $l_{it} \in I_{it-1}$), we could use

$$\frac{1}{N}\frac{1}{T}\sum_i\sum_t \begin{pmatrix} \widehat{\xi}_{it}(\widehat{\beta}_1,\widehat{\beta}_2)k_{it} \\ \widehat{\xi}_{it}(\widehat{\beta}_1,\widehat{\beta}_2)l_{it} \end{pmatrix} = 0$$

  One can see how various other timing/information assumptions would determine the different moment conditions here (e.g. Ackerberg (2016)).

- Notes:

- 1) Even though first stage does not identify any parameters, it is still crucial in that it "separates" $\omega_{it}$ from $\epsilon_{it}$.

- 2) The procedure does not rely on labor being a non-dynamic input, i.e. labor choices could have dynamic implications (e.g. hiring or firing costs).

- 3) The procedure allows firm specific, serially correlated, unobserved shocks to the price of labor (as well as to capital costs). We cannot allow such shocks to the price of intermediate inputs (it would violate the scalar unobservable assumption necessary for the inversion), but in many cases intermediate inputs are commodities where we would expect very little price variation across firms.

  - OP - rules out serially correlated, unobserved, firm specific shocks to all input prices ($i_{it}$, $l_{it}$,$m_{it}$) (note: can allow non-serially correlated shocks to prices of $l_{it}$ and $m_{it}$)
  - LP - allows serially correlated, unobserved, firm specific input price shocks to $i_{it}$, but not to ($l_{it}$,$m_{it}$)
  - ACF (with intermediate input proxy) allows serially correlated, unobserved, firm specific input price shocks to $i_{it}$ and $l_{it}$, but not to $m_{it}$

- 4) Bond and Soderbom (2005) and GNR argument implies that we actually need some degree of 2) or 3) for identification of the labor coefficient.

- 5) Can use $i_{it}$ rather than $m_{it}$ as the proxy variable in ACF procedure, but lose ability to allow serially correlated, unobserved, firm specific input price shocks to $i_{it}$ and $l_{it}$.

- 6) Can also overidentify the model by adding further lags of inputs as instruments.

- 7) Also note that the production function (20) does not include $m_{it}$. This is because the Bond-Soderbom (and GNR) arguments imply that we cannot use $m_{it-1}$ as an instrument to identify the coefficient on $m_{it}$. Not including $m_{it}$ is implicitly using a value-added production function (i.e. $y_{it}$ is deflated revenues minus deflated costs of intermediate inputs). One structural interpretation of this is that the production function is Leontief in the intermediate input (e.g. materials), i.e.

$$Y_{it} = \min\left\{\beta_0 K_{it}^{\beta_1} L_{it}^{\beta_2} e^{\omega_{it}}, \beta_3 M_{it}\right\} e^{\epsilon_{it}}$$

which implies that

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \tag{21}$$

An alternative would be to follow Appendix B of Levinsohn and Petrin, or Gandhi, Navarro and Rivers (2015) and use a first order condition to obtain the coefficient on the intermediate input.

- 8) Less parametric generalizations
  - Ackerberg and Hahn (2015) show that in this "value added model"

$$\begin{aligned} Y_{it} &= \min \left\{ F\left(K_{it}, L_{it}, \omega_{it}\right), \beta_3 M_{it} \right\} e^{\epsilon_{it}} \\ y_{it} &= \min \left\{ f(k_{it}, l_{it}, \omega_{it}), \ln(\beta_3) + m_{it} \right\} + \epsilon_{it} \end{aligned}$$

    if $f$ is strictly monotonic in the scalar markov process $\omega_{it}$, it can be fully non-parametrically identified. We formally consider the generic model $y_{it} = f\left(x_{it}, \omega_{it}\right)$ and show conditions on timing and information sets under which $f$ is non-parametrically identified. We describe the result as showing how the timing/information set assumptions crucial to OP have power in a non-parametric context. Note that these assumptions are starting to be used in other literatures, e.g. demand with endogenous product characteristics.
  - Gandhi, Navarro and Rivers (2015) extend the first order condition approach to estimating the effect of the proxy variable ($m_{it}$) to a non-parametric setting and show that in

$$y_{it} = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \epsilon_{it}$$

    $f$ is non-parametrically identified. I would describe this approach as combining the timing/information set assumption identification approach (for $k_{it}$ and $l_{it}$) with the first order condition identification approach (for $m_{it}$)

- 9) More generally can mix-and-match the different identification strategies (for different inputs), i.e.
  - timing/information set assumptions (though as detailed above, this does not work for a non-dynamic, variable inputs that is being used to proxy for unobserved productivity).
  - first order conditions (at least for static inputs)
  - observed firm specific input price shocks as instruments

# 8 Dynamic Panel Approaches

- These are econometric procedures (Arellano and Bond (1991, ReStud), Arellano and Bover (1995. JoE), Blundell and Bond (1998, JoE, 2000, ER), Arellano and Honore (2001, Handbook)) that generalize the fixed effects model to allow $\omega_{it}$ to vary across time. These have been used in many different applied contexts, including production functions (e.g. Blundell Bond 2000, and emprical work by John Van Reenan, Nick Bloom and coauthors).

- "Dynamic Panel" is somewhat of a misnomer in the context that I am using these methodologies, as there is no lagged dependent r.h.s. variable. Many of these methods were developed in that context.

- I will focus on one very simple example, to try to highlight the similarities and differences between this literature and the literature stemming from OP.

- Production function
$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \tag{22}$$
where
$$\omega_{it} = \rho\omega_{it-1} + \xi_{it}$$

- Suppose that $\epsilon_{it}$ satisfies strict exogeneity, i.e. $\epsilon_{it}$'s are uncorrelated with all input choices.

- Suppose that $\omega_{it}$ not observed until $t$, that $k_{it}$ is chosen at $t-1$, and that $l_{it}$ is chosen at $t$.

- These assumptions are analagous to the timing/information set assumptions made in OP, and imply the following orthogonality conditions
$$\begin{aligned} E\left[\epsilon_{it}k_{is}\right] &= E\left[\epsilon_{it}l_{is}\right] = 0 \ \ \forall t, s \\ E\left[\xi_{it}k_{is}\right] &= 0 \text{ for } s \leq t \\ E\left[\xi_{it}l_{is}\right] &= 0 \text{ for } s < t \end{aligned}$$

- Consider "$\rho$-differencing" the production function, i.e.
$$y_{it} - \rho y_{it-1} = (1-\rho)\beta_0 + \beta_1\left(k_{it} - \rho k_{it-1}\right) + \beta_2\left(l_{it} - \rho l_{it-1}\right) + \xi_{it} + \left(\epsilon_{it} - \rho\epsilon_{it-1}\right)$$
or
$$y_{it} = \rho y_{it-1} + (1-\rho)\beta_0 + \beta_1\left(k_{it} - \rho k_{it-1}\right) + \beta_2\left(l_{it} - \rho l_{it-1}\right) + \xi_{it} + \left(\epsilon_{it} - \rho\epsilon_{it-1}\right)$$

- Now, given a guess of the parameters $(\rho, \beta_0, \beta_1, \beta_2)$ one can compute the implied values of the term $\xi_{it} + (\epsilon_{it} - \rho\epsilon_{it-1})$, i.e.
$$\xi_{it} + (\widehat{\epsilon_{it} - }\rho\epsilon_{it-1})\left(\rho, \beta_0, \beta_1, \beta_2\right) = y_{it} - \rho y_{it-1} - (1-\rho)\beta_0 - \beta_1\left(k_{it} - \rho k_{it-1}\right) - \beta_2\left(l_{it} - \rho l_{it-1}\right)$$
Then estimation can proceed using, e.g. by setting the sample moment
$$\frac{1}{N}\frac{1}{T}\sum_i\sum_t\left(\xi_{it} + (\widehat{\epsilon_{it} - }\rho\epsilon_{it-1})\left(\rho, \beta_0, \beta_1, \beta_2\right) \otimes \begin{pmatrix} 1 \\ k_{it} \\ k_{it-1} \\ l_{it-1} \end{pmatrix}\right) = \underrightarrow{0}$$

- Again, there are actually many more potential moment conditions, since all values of $k, l$, and $y$ prior to $(k_{it}, l_{it-1}, y_{it-2})$ are also uncorrelated with $\xi_{it} + (\epsilon_{it} - \rho\epsilon_{it-1})$.

- Note:
  - There are implicit assumptions here about what makes lags strong instruments. This depends on dynamic issues (e.g. adjustment costs) and issues regarding serial correlation in input prices. See Blundell and Bond (1999) and Bond and Soderbom (2005) for discussion of when they are strong instruments, and when they are not so strong instruments.

- One can extend the model to allow for an additional unobservable that is fixed across time, i.e.
$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \alpha_i + \omega_{it} + \epsilon_{it} \tag{23}$$
where $\alpha_i$ is allowed to be correlated with all input choices. This model requires "double differencing", i.e.
$$\begin{aligned}(y_{it} - \rho y_{it-1}) - (y_{it-1} - \rho y_{it-2}) &= \beta_1 \left[(k_{it} - \rho k_{it-1}) - (k_{it-1} - \rho k_{it-2})\right] + \beta_2 \left[(l_{it} - \rho l_{it-1}) - (l_{it-1} - \rho \right. \\ &\quad + \xi_{it} - \xi_{it-1} + (\epsilon_{it} - \rho \epsilon_{it-1}) - (\epsilon_{it-1} - \rho \epsilon_{it-2})\end{aligned}$$
Again, one can appropriately lag $k, l$, and $y$ to find valid moments. One problem is that double differencing can be demanding on the data, and estimates can be imprecise. Arellano and Bover (1995) and Blundell and Bond (1998, 2000) suggest some additional moment conditions based on stationarity assumptions that can help here, though these assumptions may be strong.

- So how do these dynamic panel approaches compare to the Olley-Pakes literature? I'd say the main tradeoff is the following
  - The dynamic panel approach does not require the scalar unobservable and strict monotonicity assumptions that are required for the OP/LP/ACF inversions. So in the dynamic panel literature, for example, one does not need to worry about unobserved firm specific input prices, nor other sorts of unobservables like optimization error.
  - On the other hand, the dynamic panel approach requires that the serial correlation in $\omega_{it}$ is linear, e.g. an AR or MA process. This is essential to being able to construct usable moments. In contrast, the OP/LP/ACF literature can allow the productivity shock to follow a completely general first order markov process.
  - Dynamic panel literature can also allow additional fixed effects, although precise estimation appears to be challenging. Intermediate assumption of Arellano-Bover Blundell-Bover may be helpful.

- Given that theory may provide little guidance between choosing between these two sets of assumptions (and because they are both strong), I would suggest trying both approaches.

# 9 Other Issues

- 1) Often observe firm revenues as the output measure, not physical quantities. As pointed out by Klette and Griliches (1996, JAE), this can be problematic when firms operate in distinct imperfectly competitive output markets (and one does not observe output price).
  - Intuition: Suppose observe that firms that (exogenously) use double the inputs of others produce less than double revenue of other. There are two explanations - 1) declining returns to scale (but, e.g., perfect competition), 2) constant returns to scale with a downward sloping demand curve.
  - Even if one doesn't care about separating the above two effects, there may now be two distinct sources of unobservables in the (revenue) production function, which can be problematic for the proxy based approaches.

- See, e.g. Klette (1999, JIE), Foster, Haltiwanger, and Syverson (2007, AER), DeLoecker (2011, Ecta), DeLoecker and Warzynski (2012, AER).
- Note that just observing quantites is not a complete panacea - need to be equivalent across firms for these quantities to be meaningful.

- 2) Other types of information structures (e.g. Greenstreet (2007)).

- 3) Additional inputs (e.g. Griliches Knowledge Capital model, Doraszelski and Jaumandreu (2014, ReStud))

  - Griliches Knowledge capital

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 c_{it} + \omega_{it} + \epsilon_{it} \tag{24}$$

    where

$$c_{it} = \sum_{\tau=0}^{t} \delta_c^{t-\tau} d_{i\tau}$$

    where $d_{i\tau}$ is firm's chosen R&D expenditures in period $\tau$
  - Doraszelski and Jaumandreu (2014)

$$y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \tag{25}$$

    where

$$\omega_{it} = g(\omega_{it-1}, d_{it-1})$$

  - Advantages DJ
    * Doesn't have has initial $t = 0$ R&D stock issue that Griliches has (assuming use OP/LP/ACF related methods to estimate)
    * Explicitly has uncertainty in the contribution of R&D to productivity
  - Disadvantages DJ
    * Unobserved component of "productivity" and R&D component of "productivity" affect future through scalar - this is restrictive, e.g.

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

      $d_{it-1}$ and $\xi_{it}$ forced to depreciate at same rate. This is not the case in Griliches - i.e. $\delta_c$ different than $\rho$.
    * Alternative model of physical capital stock

$$y_{it} = \beta_0 + \beta_2 l_{it} + \omega_{it} + \epsilon_{it} \tag{26}$$

      where

$$\omega_{it} = g(\omega_{it-1}, d_{it-1}, i_{it-1})$$

  - Note: both "endogenize productivity" if think about defining productivity as $(\beta_3 c_{it} + \omega_{it})$ in the Griliches model

- 4) Empirical questions -

- Determinants of productivity - deregulation (OP), trade openess, exporting - similar to above
  * generally best to include these factors as inputs in the production function
- Allocative efficiency - recently Hsieh and Klenow (2009, QJE), Asker, Collard-Wexler, and DeLoecker (2014, JPE)