# ERSA Training Workshop
# Lecture 4: Estimation of Production Functions with Micro Data

Måns Söderbom

Thursday 15 January 2009

# 1 Introduction

Earlier in this course you will have seen how panel data methods can be used to estimate the parameters of a production function, which may take the following form:

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_j + u_{jt}\right),$$

where $y, k, l$ denote output (or value-added), capital, labour, respectively, $j, t$ denote firm and time (panel data), respectively, $\omega_j$ is a firm-specific unobserved effect, $u_{jt}$ is a time varying residual, and $\beta_k, \beta_l$ are unknown parameters.

The main reasons for using a panel estimator in this context are as follows:

- The researcher might suspect there is **time-invariant unobserved heterogeneity** across firms in underlying productivity - controlling for 'fixed effects', either by means of differencing, by going within, is meant to take care of this.

- The researcher might suspect that the time varying component of the residual is **serially correlated** in levels - pseudo-differencing the levels equation which results in a dynamic model is meant to take care of this. Recall: if

$$u_{jt} = \rho u_{j,t-1} + e_{jt},$$

we have

$$\begin{aligned} y_{jt} &= \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_j + u_{jt}\right) \\ y_{jt} &= \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_j + \rho u_{j,t-1} + e_{jt}\right), \end{aligned}$$

and since, by definition,

$$\rho u_{j,t-1} = \rho y_{j,t-1} - \rho \beta_k k_{j,t-1} - \rho \beta_l l_{j,t-1} - \rho \omega_j,$$

the production function can be written as a dynamic equation (with common factor restrictions):

$$y_{jt} = \rho y_{j,t-1} + \beta_k k_{jt} - \rho\beta_k k_{j,t-1} + \beta_l l_{jt} - \rho\beta_l l_{j,t-1} + \omega_j (1 - \rho) + e_{jt}.$$

- The researcher might suspect that the time varying component of the **residual is correlated with the factor inputs** (e.g. capital, labour) - using instrumental variables is meant to take care of this.

The panel data methods discussed earlier in this course are obviously very general, and not specific to production functions. In this lecture we discuss in more detail the econometrics of estimating production functions using panel data, typically at the firm level.

# 2 Why are we interested?

In so far as there is one thing on which economists appear to be able to agree it is the desirability of higher **productivity**. The production function is an important tool that can be used to analyze various aspects of productivity. Here are some research questions/issues that can be addressed using a production function approach:

- **Scale and productivity.** In most datasets on enterprises in Sub-Saharan Africa, labour productivity (usually defined as value-added per worker) is much higher large than small firms (see e.g. the survey paper by Bigsten & Söderbom, WBRO, 2006). Is this because large firms have more capital per worker, or because there are increasing returns to scale? If we believe the production function above is correctly specified, we can answer this question by estimating $\beta_k$ and $\beta_l$.

- Suppose we convince ourselves there are increasing returns to scale, i.e. $\beta_k + \beta_l > 1$. One implication would be that if a fixed set of inputs (at the national level) gets allocated to a small number of large firms this results in more aggregate output than if allocated to a large number of small firms. This may be important for policy.

- In contrast, if we convince ourselves returns to scale of constant, $\alpha + \beta = 1$, a reallocation of resources between firms of differing size may not impact on aggregate output (e.g. two small firms will produce as much output as one large firm using the same amount of inputs as the two small ones between them).

- In fact, the evidence on returns to scale in developing countries is most consistent with **constant** returns to scale (see e.g. Söderbom and Francis

Teal, 2004, for evidence on enterprises in Ghana). That is, while there are many small firms in developing countries, this does not imply foregone scale economies. You will have seen that Blundell-Bond obtain the same result for firms in the US.

- Although I won't be taking about farms in this lecture, production functions are commonly used in **agricultural economics** too. For example, a common view is that small farms are more productive than large farms; however the empirical evidence on the matter is somewhat mixed (e.g. Lamb, 2001, refutes this notion, concluding that large farms are as productive as small ones)*.

- Rates of technological change.

*Lamb, R. L. "Inverse productivity: land quality, labor markets and measurement error" Journal of Development Economics, 2003, 71: 71-95.

- Rates of return on, for example, R&D or exporting ('learning-by-exporting')

- The contribution of various forms of inputs to output (e.g..skilled & unskilled labour).

# 3 Production functions & the basic endogeneity issue

We focus on the simple 2-factor Cobb-Douglas production function:

$$Y_j = A_j K_j^{\beta_k} L_j^{\beta_l},$$

or, in natural logarithms,

$$y_j = \beta_0 + \beta_k k_j + \beta_l l_j + \epsilon_j,$$

where

$$\ln A_j = \beta_0 + \epsilon_j$$

is log TFP. $\beta_0$ is a constant, interpretable as the mean of log TFP, while $\epsilon_j$ measures the deviation in productivity from the mean, for firm $j$. TFP is typically assumed unobserved (at least partially).

Suppose we have micro data on output, capital and labour. How can the parameters of this equation be estimated?

- As you know, for OLS to consistently estimate the $\beta$-parameters, the error term must have zero mean and be uncorrelated with the explanatory variables:

$$E\left(\epsilon_j\right) = 0,$$

$$Cov\left(k_j, \epsilon_j\right) = 0, \quad (1)$$

$$Cov\left(l_j, \epsilon_j\right) = 0 \quad (2)$$

  The zero mean assumption is innocuous, as the intercept $\beta_0$ would pick up a non-zero mean in $\epsilon_j$.

- The crucial assumption is zero covariance. Is this likely to hold in the present context?

- No - because it seems quite possible that the firm's capital and labour decisions are influenced by factors that are observed to the firm's manager but unobserved to the econometrician, i.e. by $\epsilon_j$. This would set up a correlation between the regressors and the residuals, rendering the OLS estimates biased and inconsistent.

## 3.1 Illustration

Assumptions:

- Firms operate in perfectly competitive input and output markets (so that input and output prices are not affected by the actions of firm $j$);

- Capital is a fixed input (decided upon one period in advance, say) rented at rate $r$;

- Firms observe $\epsilon_j$ before hiring labour (at rate $W$), and labour is a 'flexible input' that can be altered without dynamic implications.

The firm's profit is given by

$$\begin{aligned}\pi_j &= pY_j - WL_j - rK_j \\ \pi_j &= p\left(A_j K_j^{\beta_k} L_j^{\beta_l}\right) - wL_j - rK_j,\end{aligned}$$

where $p$ is the output price. Assuming the firm maximizes profits, it will choose labour such the following first-order condition is fulfilled:

$$\beta_l p A_j K_j^{\beta_k} L_j^{\beta_l - 1} = W,$$

which implies

$$L_j = \left(\frac{\beta_l p A_j}{W}\right)^{\frac{1}{1-\beta_l}} K_j^{\frac{\beta_k}{1-\beta_l}},$$

or, in logs,

$$l_j = \frac{1}{1-\beta_l}\left[\ln \beta_l + \ln p - \ln W + \ln \beta_0 + \epsilon_j + \beta_k k_j\right].$$

- Clearly in this case $l_j$ depends on unobserved TFP (which is the interpretation assigned to the residual $\epsilon_j$) and so estimating the production function

$$y_i = \beta_0 + \beta_k k_j + \beta_l l_j + \epsilon_j.$$

by means of OLS will give biased and inconsistent results.

- Note that, since the first-order condition for labour implies a positive correlation between $l_j$ and $\epsilon_j$, we would expect the OLS estimate of $\beta_l$ to be upward biased.

## 3.2 Other endogeneity issues

- **Attrition.** Suppose the probability of exit is a negative function of the value of the firm, and suppose the value of the firm depends on unobserved productivity and the level of capital stock installed:

$$\begin{aligned} \mathrm{Pr}\left(exit_{j,t+1} = 1 | \epsilon_j, k_j\right) &= \mathrm{Pr}\left(V_j\left(\epsilon_j, k_j\right) < \Omega\right) \\ &= f\left(\epsilon_j, k_j\right), \end{aligned}$$

  where $f_1 < 0$, $f_2 < 0$. That is, the typical firm that would exit would be one with a low level of productivity and a low level of capital (this would be a low value firm).

- Think about what this means for the correlation between unobserved productivity and observed capital in the "selected sample", i.e. in the sample of survivors.

- Firms with a lot of capital are likely to survive even if they have low productivity, because they have high values.'

- However firms with little capital will only survive if they have high levels of productivity.

- Hence, in the *sample of survivors* there will be a **negative** correlation between $k_j$ and unobserved productivity $\epsilon_j$.

- Thus, if we estimate the production function

$$y_j = \beta_0 + \beta_k k_j + \beta_l l_j + \epsilon_j,$$

this mechanism would tend to yield a downward bias in the coefficient on $k_j$.

- **Measurement errors**. In general, we expect measurement errors in inputs to lead to downward bias (**attenuation bias**) in the estimated coefficients. Recall the attenuation bias formula:

$$y_{it} = \beta x_{it}^{*} + v_{it},$$

where $x_{it}^{*}$ is the true but unobserved value of the explanatory variable, and $v_{it}$ is a non-autocorrelated, homoskedastic error term with zero mean. We observe an imperfect measure of $x_{it}^{*}$ , namely $x_{it}$ such that

$$x_{it} = x_{it}^{*} + u_{it},$$

where $u_{it}$ is a random measurement error uncorrelated with $x_{it}^{*}$. Our estimable equation is

$$y_{it} = \beta x_{it} + \left(v_{it} - \beta u_{it}\right),$$

so the regressor $x_{it}$ is correlated with the error term $(v_{it} - \beta u_{it})$. It can be shown that this will lead to a downward bias in the OLS estimate of $\beta$

- that is, estimated $\beta$ is **lower** than true $\beta$. To give you an idea of what the bias looks like, consider the following formula showing the bias caused by measurement errors:

$$p\,\text{lim}\,\hat{\beta}^{OLS} = \beta\left(\frac{\sigma^2_{x^*}}{\sigma^2_{x^*} + \sigma^2_u}\right),$$

where $\sigma^2_{x^*}$ is the variance of the true, unobserved explanatory variable, and $\sigma^2_u$ is the variance of the measurement error. The operator $p\,\text{lim}$ can be thought of as showing the value of estimated $\beta$ in a large sample. Loosely speaking, this is what we can expect to get if there are measurement errors in the explanatory variable. Clearly the higher the variance of the measurement error, the more severe is the bias.

- What happens if we take first differences? Clearly,

$$p\lim \hat{\beta}^{FD} = \beta \left( \frac{\sigma^2_{dx^*}}{\sigma^2_{dx^*} + \sigma^2_{de}} \right),$$

where $d$ indicates that the variance refers to the differenced variable. Assumed that the variance is constant over time and that the mean of $z$ is zero, it can be shown that

$$p\lim \hat{\beta}^{FD} = \beta \left( \frac{\sigma^2_{x^*}}{\sigma^2_{x^*} + \sigma^2_e \frac{(1-\rho_e)}{(1-\rho_{x^*})}} \right)$$

where $\rho_e$ is the serial correlation of the measurement errors and $\rho_{x^*}$ is the serial correlation of the true values of the regressors.

Now compare the following expressions:

$$
\begin{aligned}
p\lim \hat{\beta}^{OLS} &= \beta \left( \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2} \right), \\
p\lim \hat{\beta}^{FD} &= \beta \left( \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2 \frac{(1-\rho_e)}{(1-\rho_{x^*})}} \right).
\end{aligned}
$$

Which one has the most severe bias?

The bias of the FD estimator will be more severe than that of the levels estimator if $\frac{(1-\rho_e)}{(1-\rho_{x^*})} > 1$, i.e. if $\rho_{x^*} > \rho_e$.

This is an important result. In most applications we assume that the serial correlation of the measurement errors typically is quite small or zero, while the serial correlation of the true unobserved explanatory variable is positive. In this case *first differencing the data is bound to exacerbate the measurement error* *bias*, and OLS estimation of the levels equation would be preferable to the FD model.

- In practice, estimating the coefficient on the capital stock whilst controlling for fixed effects has proved difficult - see Söderbom and Teal, 2004, for details.

# 4 Traditional solutions to the endogeneity problem

The two traditional solutions to endogeneity problems are **instrumental variables** and **fixed effects**. We are now going to write the production function as

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt},$$

i.e. we have added time subscripts reflecting the panel dimension in the data; and we have decomposed the residual $\epsilon$ into two components, $\omega_{jt} + \eta_{jt}$

- $\omega_{jt}$ represents the part of TFP observable to the firm but not to the econometrician - hence this is the source of endogeneity problems. You can think of $\omega_{jt}$ as a measure of the managerial quality of the firm. From now on, we will refer to $\omega_{jt}$ as 'unobserved productivity'.

- $\eta_{jt}$ on the other hand is assumed not to impact on the firm's input decisions. You can think of $\eta_{jt}$ as representing measurement errors in output, for example (other interpretations are possible too; see Section 2.2 in ABBP). What's important is that $\eta_{jt}$ is not a source of endogeneity bias.

## 4.1 Instrumental Variables

Our problem: We want to estimate

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt},$$

but we cannot use OLS, since

$$Cov\left(l_{jt}, \omega_{jt}\right) \neq 0.$$

(It is likely, of course, that capital is endogenous too, but we abstract from that possibility for the moment.)

Suppose an instrument $z_{jt}$ is available, that fulfills the following conditions:

1. The instrument is **valid** (or **exogenous**):

$$cov\left(z_{jt}, \omega_{jt}\right) = 0.$$

This is an **exclusion restriction** - $z_{jt}$ is excluded from the structural equation (the production function).

2. The instrument is **informative** (or **relevant**). This means that the instrument $z_{jt}$ must be correlated with the endogenous regressor (labour in the current example), conditional on all exogenous variables in the model (i.e. capital, if this is thought exogenous). That is, if we assume there is a linear relationship between $l_{jt}$ and $z_{jt}$ and $k_{jt}$,

$$l_{jt} = \delta_0 + \delta_1 k_{jt} + \theta_1 z_{jt} + r_{jt}, \tag{3}$$

   where $r_{jt}$ is mean zero and uncorrelated with the variables on the right-hand side, we require $\theta_1 \neq 0$.

Many economists take the view that, for instrumental variable estimation to be convincing, the instruments used must be motivated by theory. Recall the first-order condition for labour derived above - with my slightly modified notation we get

$$l_{jt} = \frac{1}{1-\beta_l}\left[\ln\beta_l + \ln p - \ln W + \beta_k k_{jt} + \omega_{jt}\right].$$

- This suggests the wage rate $W$ might be a useful instrument:
  - Our theory says it is (negatively) correlated with labour.
  - The wage rate also must be uncorrelated with $\omega_{jt}$. This may not be an entirely innocuous assumption to make. While the wage rate does not directly enter the production function, wages might be correlated with unobserved productivity for other reasons - e.g. if more productive firms have stronger market power in input markets - in which case the wage will not be a valid instrument.

- It also follows from the first-order condition above that the output price is a potential instrument - however, that has been used less often in the literature. Why might we be concerned about using the output price is an instrument?

- A similar way of reasoning can be applied for capital, if that is thought endogenous (i.e. use the cost of capital as an instrument).

**Five reasons why the IV approach based on prices as instruments has not been very successful**

1. **Market power.** Wages and capital prices (and output prices) could well be correlated with unobserved productivity if input (output) markets are not perfectly competitive: e.g. high unobserved productivity gives the firm market power and so enables it to influence the price.

2. **Wages and unobserved worker quality**. When labour costs are reported in firm-level datasets, they typically come in the form of average wage per worker, and you may well be concerned that the average wage in the firm is correlated with unobserved quality of the workforce. Since the unobserved quality of the workforce likely impacts on unobserved productivity, this would imply the average wage is an invalid instrument.

3. **Law of one price.** If, as is typically the case, one wants to include time dummies in the production function, there must be variation in input prices **across** firms at a given point in time for these to be useful instruments. If input markets are essentially national in scope, this seems unlikely. (If average wages indeed vary across firms in most datasets, you suspect this is at least partly picking up unobserved worker quality).

4. **Endogenous unobserved productivity.** Suppose unobserved productivity $\omega_{jt}$ actually **depends** on input choices - e.g. investment in modern technology raises productivity. In that case it will be hard to argue that input prices are valid instruments, since these surely will impact on investment.

5. **Attrition.** A different kind of endogeneity problem sometimes discussed in the literature is posed by endogenous attrition, i.e. that the firm's exit

decision depends on unobserved productivity as well as input prices (after all, these jointly determine the profitability of the firm). In such a case we will have a Heckman type selection problem, in which all variables determining the exit decision will go into the residual of the production function in the selected sample. Clearly input prices cannot be used as instruments in this case.

The common theme across these reasons is that prices are unlikely to be valid instruments.

## 4.2 Fixed Effects

A second traditional solution to the endogeneity problem is fixed effects estimation, which as you know requires panel data. One key assumption underlying this approach is that unobserved productivity is constant over time,

$$\omega_{jt} = \omega_j$$

but varies across firms. We would now write the production function as

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_j + \eta_{jt}\right),$$

and use perhaps the within estimator ('fixed effects' estimator) or the first-differenced estimator to estimate the parameters - in the latter case for example we would thus estimate

$$y_{jt} - y_{j,t-1} = \beta_k \left(k_{jt} - k_{j,t-1}\right) + \beta_l \left(l_{jt} - l_{j,t-1}\right) + \left(\eta_{jt} - \eta_{j,t-1}\right),$$

using OLS (probably with firm-clustered standard errors since the differenced residual is likely serially correlated).

Notice that the source of endogeneity bias has been eliminated, thus effectively solving the endogeneity problem (subject of course to strict exogeneity; see earlier lectures in this course).

**Three reasons why the fixed effects approach has not been very successful**

1. **Time invariant unobserved productivity.** The assumption that unobserved productivity is fixed over time is quite restrictive, especially in longer panels.

2. **Differencing may exacerbate measurement error bias.** When there are measurement errors in inputs, the fixed effects estimator may well be more severely biased than the OLS estimator. Discuss.

3. **Poor performance in practice.** Fixed effects estimates of the capital coefficient are often implausibly low, and estimated returns to scale is often (severely) decreasing ($\beta_k + \beta_l << 1$).

[EXAMPLE 1. To be discussed in class]

# 5 The Olley and Pakes (1996) approach

The Olley & Pakes (1996; henceforth OP) use a different approach to solve the endogeneity problems discussed above. Similar to the IV approach, OP derive their solution from the **input demand equations**, however OP do not require factor prices to be observed. In what follows I will discuss a simplified version of the OP model.

- The production function:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_{jt} + \eta_{jt}\right).$$

  (the original OP model also allows for an effect of firm age, but I ignore that here).

**Summary of key assumptions:**

- **Labour** is a flexible input chosen in period $t$, after observing productivity $\omega_{jt}$.

- **Capital** is a "quasi-fixed" input chosen in period $t-1$ and subject to strictly convex adjustment costs. Capital evolves according to the equation

$$K_{jt} = (1 - \delta) K_{j,t-1} + I_{j,t-1},$$

where $I_{j,t-1}$ denotes investment.

- Unobserved **productivity** $\omega_{it}$ is assumed to follow a first order Markov process,

$$p\left(\omega_{j,t+1} | \{\omega_{j\tau}\}_{\tau=0}^{t}, I_{jt}\right) = p\left(\omega_{j,t+1} | \omega_{jt}\right),$$

where $I_{jt}$ is the firm's information set in period $t$. This means that, given the present information, future states are independent of the past states - lags of the productivity variable do not provide additional information as to what might happen to productivity in the future. Examples:

- Linear process

$$\omega_{jt} = \rho\omega_{j,t-1} + \xi_{jt}.$$

- Nonlinear process

$$\omega_{jt} = \rho_1\omega_{j,t-1} + \rho_2\omega_{j,t-1}^3 + \xi_{jt}.$$

- Nonparametric process

$$\omega_{jt} = f\left(\omega_{j,t-1}\right) + \xi_{jt}.$$

Recall the linear process was adopted by Blundell and Bond in their analysis of production functions based on US data.

- The **profit** in period $t$ is defined as

$$\Pi_t = pK_{jt}^{\beta_k} L_{jt}^{\beta_l} \exp\left(\beta_0 + \omega_{jt}\right) - W_{jt}L_{jt} - p^I I_{jt} - G(I_{jt}, K_{jt}),$$

  where $p$ is the output price, $p^I$ is the price of one unit of capital, and $G(I_{jt}, K_{jt})$ is the adjustment cost for capital. Note: labour is not a function of $\eta_{jt}$ since we're assuming this term is just random noise (output measurement error).

- Since labour is assumed to be a flexible input, the static first-order condition applies:

$$\beta_l p K_{jt}^{\beta_k} L_{jt}^{\beta_l - 1} \exp\left(\beta_0 + \omega_{jt}\right) = W_{jt},$$

$$L_{jt} = \left( \frac{\beta_l p \exp\left(\beta_0 + \omega_{jt}\right)}{W_{jt}} \right)^{\frac{1}{1-\beta_l}} K_{jt}^{\frac{\beta_k}{1-\beta_l}}.$$

Using this expression for labour in the profit function above, we can rewrite profits as

$$\begin{aligned} \Pi_t &= (1-\beta_l)\, \beta_l^{\frac{\beta_l}{1-\beta_l}} \left(p \exp\left(\beta_0 + \omega_{jt}\right)\right)^{\frac{1}{1-\beta_l}} \left(W_{jt}\right)^{\frac{\beta_l}{\beta_l - 1}} K_{jt}^{\frac{\beta_k}{1-\beta_l}} \\ &\quad - p^I I_{jt} - G(I_{jt}, K_{jt}), \end{aligned}$$

or, in more reader-friendly notation,

$$\Pi_t = \varphi\left(W_{jt}, \omega_{jt}\right) K_{jt}^{\frac{\beta_k}{1-\beta_l}} - p^I I_{jt} - G(I_{jt}, K_{jt}).$$

You see how the labour variable has "disappeared" - replaced by the variables and parameters determining $L_{jt}$ as implied by the first-order condition for labour. Using a notation more similar to that in OP, we might therefore write profits as

$$\Pi_t = \pi \left( k_{jt}, \omega_{jt} \right) - c \left( I_{jt} \right)$$

where $\pi_{jt}$ is sales minus labour costs, and $c \left( I_{jt} \right)$ is the cost of investment, including strictly convex adjustment costs, for example

$$G(I_{jt}, K_{jt}) = \frac{\gamma}{2} \left( \frac{I_{jt}}{K_{jt}} \right)^2 K_{jt}.$$

where $\gamma$ is a parameter measuring the marginal adjustment cost of capital.

**The firm's objective**

- The final important assumption underlying the OP framework concerns the behaviour of the firm.

- It is assumed that the firm chooses investment and employment to maximize the present value of current and expected future net revenues. We have already seen how labour is "optimized out" at each period, which means we can write the value of the firm as a function of capital and productivity only:

$$V(k_{jt}, \omega_{jt}) = \max_{I_t} E_t \sum_{s=t}^{\infty} \psi^{(s-t)} \left[ \pi \left( k_{js}, \omega_{js} \right) - c \left( I_{js} \right) \right],$$

  where $E_t$ denotes expectation given the information available in period $t$, and $\psi$ is a discount factor. The choice variable (or control variable) here is investment in period $t$.

- Note: the fact that labour is not visible in this equation does **not** mean labour is irrelevant. Labour is not visible here because we have implicitly replaced it by the variables and parameters **determining** labour as implied by the first-order condition. Indeed, estimating the coefficient on labour in the production function is a central objective in the analysis.

- Alternatively, we can write the value of the firm recursively as a (stochastic) Bellman equation:

$$V(k_{jt}, \omega_{jt}) = \max_{I_t} \pi\left(k_{jt}, \omega_{jt}\right) - c\left(I_{jt}\right) + \psi E_t[V(k_{j,t+1}, \omega_{j,t+1})] \quad (4)$$

**The firm's investment demand**

- Key for the OP approach is the firm's investment. In a model of the form outlined above, optimal investment in period $t$ will depend on
  - the existing capital stock; and
  - expectations about the future profitability of capital.
- The first-order Markov assumption implies that expected productivity in the future depends on current, but not past, productivity.
- OP hence write down an investment demand function of the following form:

$$I_{jt} = I_t \left( k_{jt}, \omega_{jt} \right).$$

This function needs to be **strictly increasing** in unobserved productivity for the OP procedure to work - a firm with a high value of $\omega_{jt}$ will invest strictly more than a firm with a low value of $\omega_{jt}$, conditional on $k_{jt}$.

- [EXAMPLE 2. From Bond, Söderbom and Wu, 2008. To be discussed in class]

**Controlling for the endogeneity of input choice** We are now ready to discuss the estimation strategy proposed by OP. Notice that this is motivated by the theory discussed above.

- **The key "trick" in OP.** Recall that investment is assumed to be a strictly monotonic in $\omega_{jt}$. This implies that the investment demand function

  $$I_{jt} = I_{jt}\left(k_{jt}, \omega_{jt}\right)$$

  can be **inverted** so that productivity is expressed as a function of investment and capital:

  $$\omega_{jt} = h_t\left(k_{jt}, I_{jt}\right).$$

  Intuitively, capital $k_{jt}$ and investment $I_{jt}$ "tells" us what $\omega_{jt}$ must be.

- Now return (to the production function:

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_{jt} + \eta_{jt}\right).$$

Recall that unobserved productivity $\omega_{jt}$ is a source of endogeneity bias. We now use $\omega_{jt} = h_t\left(k_{jt}, I_{jt}\right)$ and rewrite the production function as

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + h_t\left(k_{jt}, I_{jt}\right) + \eta_{jt}.$$

By including the function $h_t\left(k_{jt}, I_{jt}\right)$ as an additional term on the right-hand side, we have effectively "controlled" for unobserved productivity.

- Building on this, OP proposed a **two stage procedure** to estimate the parameters $\beta_l$ and $\beta_k$ This works as follows.

- **First stage**: Define

$$\phi_t\left(k_{jt}, I_{jt}\right) = \beta_k k_{jt} + h_t\left(k_{jt}, I_{jt}\right),$$

and rewrite the production function

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_{jt} + \eta_{jt}\right).$$

as

$$y_{jt} = \beta_l l_{jt} + \phi_t\left(k_{jt}, I_{jt}\right) + \eta_{jt}.$$

- In general, the function $\phi_t$ is not linear. OP propose either approximating $\phi_t$ using a polynomial, e.g.

$$\phi_t\left(k_{jt}, I_{jt}\right) = \lambda_0 + \lambda_1 I_{jt} + \lambda_2 k_{jt} + \lambda_3\left(I_{jt} \times k_{jt}\right) + \lambda_4 I_{jt}^2 + \lambda_5 k_{jt}^2,$$

or using kernel methods (nonparametric). In any case, what is clear now is that, provided we control for $\phi_t\left(k_{jt}, I_{jt}\right)$, we may be able to identify the labour coefficient $\beta_l$ in the first stage. Indeed, if we use the polynomial above, all we have to do is to estimate the following regression

$$y_{jt} = \lambda_0 + \beta_l l_{jt} + \lambda_1 I_{jt} + \lambda_2 k_{jt} + \lambda_3 \left( I_{jt} \times k_{jt} \right) + \lambda_4 I_{jt}^2 + \lambda_5 k_{jt}^2 + \eta_{jt}$$

using OLS.

- [EXAMPLE 3: Applying the first-stage OP procedure to the Blundell-Bond data. To be discussed in class.]

- **Second stage**: We have now estimated $\beta_l$. In the second stage we shall estimate the capital coefficient $\beta_k$ - this cannot be estimated in the first stage. Note that the first-stage estimation will give us an estimate of the function $\phi_t$, e.g.

$$\hat{\phi}_t\left(k_{jt}, I_{jt}\right) = \hat{\lambda}_0 + \hat{\lambda}_1 I_{jt} + \hat{\lambda}_2 k_{jt} + \hat{\lambda}_3\left(I_{jt} \times k_{jt}\right) + \hat{\lambda}_4 I_{jt}^2 + \hat{\lambda}_5 k_{jt}^2,$$

  if we are using the polynomial above.

- It follows that

$$\hat{\omega}_{jt} = \hat{h}_t\left(k_{jt}, I_{jt}\right) = \hat{\phi}_{jt} - \beta_k k_{jt}.$$

- Now, recall that unobserved productivity follows a first-order Markov process;

this means we can decompose $\omega_{jt}$ as follows:

$$\begin{aligned}\omega_{jt} &= E_{t-1}\left(\omega_{jt}\right) + \xi_{jt} \\ \omega_{jt} &= g\left(\omega_{j,t-1}\right) + \xi_{jt},\end{aligned}$$

where $\xi_{jt}$ is the innovation (shock) to productivity. If productivity follows a linear autoregressive process, for example, we would have

$$\omega_{jt} = \rho\omega_{j,t-1} + \xi_{jt},$$

c.f Blundell-Bond.

- The production function, again:

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt},$$

which given the insights above can be written

$$y_{jt} - \beta_l l_{jt} = \beta_k k_{jt} + g\left(\omega_{j,t-1}\right) + \xi_{jt} + \eta_{jt},$$

or

$$y_{jt} - \beta_l l_{jt} = \beta_k k_{jt} + g\left(\hat{\phi}_{j,t-1} - \beta_0 - \beta_k k_{j,t-1}\right) + \xi_{jt} + \eta_{jt}. \quad (5)$$

Now, because capital is chosen one period in advance, the residual $\xi_{jt}+\eta_{jt}$ will be **uncorrelated** with all the right-hand side variables (remember we have already estimated $\beta_l$, which is why I have moved $\beta_l l_{jt}$ to the left-hand side here).

- Depending on how flexible you want to be, (5) can be estimated using either OLS (if $g$ is linear); NLLS (if $g$ is a polynomial); or kernel methods (if $g$ is treated nonparametrically).

- [EXAMPLE 4: Applying the second-stage OP procedure to the Blundell-Bond data. To be discussed in class.]

### 5.0.1 Discussion

**Scalar unobservable assumption must hold, otherwise can't invert.**

- – Suppose there are **two** stochastic components of unobserved productivity, so that

$$I_{jt} = I_t \left( k_{jt}, \omega_{jt}^1, \omega_{jt}^2 \right),$$

and suppse $\omega_{jt}^1, \omega_{jt}^2$ follow different stochastic processes; for example, let's suppose $\omega_{jt}^1$ is highly persistent whereas $\omega_{jt}^2$ exhibits only moderate serial correlation.

– In such a case, $\omega^1_{jt}$ and $\omega^2_{jt}$ will impact differently on investment. For example, conditional on capital,

$$\begin{aligned} \omega^1_{jt} &= \Delta > 0 \\ \omega^2_{jt} &= 0, \end{aligned}$$

will give a stronger investment response than

$$\begin{aligned} \omega^1_{jt} &= 0 \\ \omega^2_{jt} &= \Delta > 0, \end{aligned}$$

because the firm understands that expected future profits are higher in the former case than in the latter case (since $\omega^1_{jt}$ more persistent).

– Because $\omega^1_{jt}, \omega^2_{jt}$ are both unobserved, we are stuck. The investment demand equation cannot be inverted; put differently, we can't infer from capital and investment the values of $\omega^1_{jt}, \omega^2_{jt}$ separately. The OP approach won't work.

**Zero investment levels potentially problematic** Recall that investment needs to be a strictly monotonic function of (scalar) unobserved productivity. The presence of lots of zero investments in the data strongly indicates that this is not the case - surely it's wildly unrealistic to assume that all firms that invest nothing have precisely the same level of unobserved productivity (conditional on capital).

If investment is irreversible, for example, the investment demand function will not be a monotonic function of productivity, and there will be lots of investment zeros in the data (see the graph taken from Bond, Söderbom and Wu, 2008).

- Levinsohn & Petrin (2003) proposed using **raw materials** as a proxy for unobserved productivity in such a case. Raw materials is rarely if ever zero in datasets and so strict monotonicity might hold. Below I will discuss a generalized approach in this vein proposed by Ackerberg, Caves and Frazer (2006), so I will not discuss the Levinsohn-Petrin estimator here.

- Alternatively, we can retain the OP approach provided we simply drop all observations for which investment is equal to zero. Provided the model is correctly specified, such a procedure would control for unobserved productivity and yield consistent estimates.

**Labour really flexible?** The OP approach just described is really **only** appropriate if labour is a flexible input. If not, e.g. because firms can't easily hire and fire workers from one day to another, then the investment demand function specified as part of the OP approach,

$$I_{jt} = I_t \left( k_{jt}, \omega_{jt} \right).$$

would no longer be correct - investment would depend on capital and unobserved productivity, but it would also depend on labour:

$$I_{jt} = I_t \left( k_{jt}, \omega_{jt}, l_{jt} \right).$$

Inverting out $\omega_{jt}$ would leave you with a function of the following form

$$\omega_{jt} = \tilde{h}_t \left( k_{jt}, I_{jt}, l_{jt} \right),$$

and so it would clearly not be possible to identify anything in the first stage:

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \tilde{h}_t \left( k_{jt}, I_{jt}, l_{jt} \right) + \eta_{jt}.$$

## Collinearity and other issues

- Ackerberg, Caves and Frazer (2006) noted that the parameter $\beta_l$ on the flexible labour input is not identified by estimating the first stage unless in a pretty special case involving either serially uncorrelated wages or serially correlated optimization errors. More generally, parameters on flexible inputs in Cobb-Douglas production functions are not identified from cross-section variation if all firms face common input prices and inputs are optimally chosen (Bond & Söderbom, work in progress).

- To see this, consider the f.o.c. for labour again:

$$l_{jt} = \frac{1}{1-\beta_l}\left[\ln\beta_l + \ln p - \ln W + \beta_k k_{jt} + \omega_{jt}\right].$$

- Identification of $\beta_L$ in OP stage 1,

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \left(\omega_{jt} + \eta_{jt}\right).$$

requires variation across firms in $l_{it}$ at given levels of $k_{it}$ and $\omega_{it}$ (or, under the assumptions of OP, at given levels of $k_{it}$ and $h_t\left(k_{jt}, I_{jt}\right)$):

$$\phi_t\left(k_{jt}, I_{jt}\right) = \beta_k k_{jt} + h_t\left(k_{jt}, I_{jt}\right),$$

- Yet the structure of the conditional labour demand function indicates that variation across firms in $l_{it}$ is fully explained by $k_{it}$ and $\omega_{it}$, if the real wage is common to all firms and the labour input is optimally chosen.

- In general, identification of parameters on flexible inputs from cross-section variation thus requires either variation across firms in the real price of those inputs, or some form of optimization error in the choice of those inputs.

- As discussed in Ackerberg, Caves and Frazer (2006), identification of $\beta_L$ using the first stage of the Olley-Pakes estimation procedure further requires that any variation across firms in the real wage, or any optimization error in the choice of labour, must be **serially uncorrelated**. The reason is that either persistent variation in real wages or persistent optimization error in the choice of labour would affect the decision rule for capital, i.e.

  $$I_{jt} = I_t\left(k_{jt}, \omega_{jt}\right).$$

  would no longer be correct - instead,

  $$I_{jt} = I_t\left(k_{jt}, \omega_{jt}, W_{jt}\right).$$

  implying that the unobserved level of log TFP could no longer be adequately proxied using a function of investment and capital alone.

- For the same reason, consistent estimation of $\beta_L$ from the first stage also requires that there must be **no** variation across firms in the **cost of capital**, and no optimization error in the investment decision.

- Thus, if inputs are optimally chosen, the only form of input price variation that allows identification of $\beta_L$ using the first stage of the estimator proposed by Olley and Pakes (1996) is the presence of serially uncorrelated variation across firms in the real wage.

## 5.1 The Ackerberg, Caves and Frazer (2006) approach

- Ackerberg, Caves and Frazer (2006; henceforth ACF) suggest an alternative estimation approach that avoids some of the problems discussed above (e.g. the potential collinearity problems), and that will work under less restrictive assumptions than those underlying the OP model.

- Just like OP (and LP), the ACF estimator is a two-step estimator. The main difference between the ACF approach and the OP (and LP) approach is that, with the ACF approach, no coefficients of interest will be estimated in the first stage of estimation. Instead, all input coefficients are estimated in the second stage. As we shall see, the first stage is still important, however.

- A useful starting point for the ACF approach is a **three-factor** Cobb-Douglas output production function,

$$q_{jt} = \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_m m_{jt} + \tilde{\omega}_{jt},$$

  where $q_{jt}$ is output, and I have added log raw materials, denoted $m_{it}$, to the basic specification used above (a constant is subsumed in $\tilde{\omega}_{jt}$). Raw materials is assumed a perfectly flexible input, and so the following static first order condition applies:

$$\alpha_m \frac{pQ_{jt}}{M_{jt}} = p^m,$$

  where $Q_{jt}$ is the level of output, $M_{jt}$ is the level of raw materials and $p^m$ is the unit price of raw materials, assumed constant in the cross-section.

- Note that raw materials will be **proportional** to output. Using this in the output production, we can obtain a value-added production function as follows:

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt}.$$

Supposing that there are measurement errors in value-added, we modify this accordingly:

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt},$$

where $\eta_{jt}$ denotes the measurement error as usual (note: $\eta_{jt}$ doesn't have to be measurement error, it can be a "real" shock to output, provided it does not impact on any of the factor inputs - see ABBP for a discussion).

- So the raw materials variable has disappeared from the scene. Why did we introduce it then? The answer is that the raw materials variable will play a role similar to that played by investment in the OP model - i.e. as a **proxy** for unobserved productivity.

Having set the scene, we are now ready to consider the ACF approach. Key assumptions are as follows

- **Materials** is a flexible input chosen in period $t$, after observing productivity $\omega_{jt}$.

- **Capital** is a "quasi-fixed" input chosen in period $t-1$ and subject to strictly convex adjustment costs.

- **Labour** is chosen *before* material inputs, but *after* capital has been chosen. in period $t$. Suppose labour is chosen at time $t-0.5$.

- Unobserved **productivity** $\omega_{it}$ is assumed to follow a first order Markov process between the subperiods $t-1, t-0.5$, and $t$:

$$p\left(\omega_{jt}|I_{j,t-0.5}\right) = p\left(\omega_{jt}|\omega_{j,t-0.5}\right),$$

  and

$$p\left(\omega_{j,t-0.5}|I_{j,t-1}\right) = p\left(\omega_{j,t-0.5}|\omega_{j,t-1}\right).$$

- Capital for period $t$ production is decided in view of $\omega_{j,t-1}$ and the firm's capital in $t-1$.

- Labour for period $t$ production is decided in view of $\omega_{j,t-0.5}$ and the firm's capital in $t$ (which is already known at this point). This may be quite realistic: labour decisions need to be made in advance, since new workers need to be trained or worker to be laid off will have to be given some period of notice.

- Materials for period $t$ production is decided in view of $\omega_{jt}$ and the firm's capital and labour in period $t$, both of which are known at this point:

$$m_{jt} = f_t\left(\omega_{jt}, k_{jt}, l_{jt}\right).$$

- **Key "trick" in ACF.** Under the assumption that materials is a strictly monotonic (increasing, to be consistent with the theory) function of $\omega_{jt}$, conditional on capital and labour, we can invert this function for $\omega_{jt}$, along the same lines as in OP:

$$\omega_{jt} = f_t^{-1}\left(m_{jt}, k_{jt}, l_{jt}\right),$$

and so we rewrite the value-added function

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt},$$

as

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + f_t^{-1}\left(m_{jt}, k_{jt}, l_{jt}\right) + \eta_{jt}.$$

Note the close similarity with OP: by including the function $f_t^{-1}\left(m_{jt}, k_{jt}, l_{jt}\right)$ as an additional term on the right-hand side, we have effectively "controlled" for unobserved productivity. The remaining residual $\eta_{jt}$ is innocuous since it has no impact on factor inputs (e.g. because it's simply measurement error in output).

- The snag here is that **no** parameter of interest can be identified based on this specification. But don't let that distract you. The key goal in the first stage is to get rid of the $\eta_{jt}$ term - why this is desirable will be clearer later.

- **First stage**. Regress log value added on a polynomial function of capital, labour and raw materials, e.g.

$$\begin{aligned} y_{jt} &= \lambda_{0t} + \lambda_{1t} k_{jt} + \lambda_{2t} l_{jt} + \lambda_{3t} m_{jt} + \\ &\quad + \lambda_{4t} k_{jt}^2 + \lambda_{5t} l_{jt}^2 + \lambda_{6t} m_{jt}^2 \\ &\quad + \lambda_{7t} k_{jt} m_{jt} + \lambda_{8t} l_{jt} m_{jt} + \lambda_{9t} k_{jt} l_{jt} \\ &\quad + \eta_{jt}, \end{aligned}$$

  using OLS. The estimated $\lambda$-parameters are not the parameters of interest.

- Now define

$$\begin{aligned} \Phi_t &= \beta_k k_{jt} + \beta_l l_{jt} + f_t^{-1}\left(m_{jt}, k_{jt}, l_{jt}\right), \\ \Phi_t &= \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} \end{aligned}$$

  which represents log value added, **net of the term** $\eta_{jt}$. Having estimated the first stage regression, we can thus estimate $\Phi_t$ simply by requesting the predicted values.

- The next task is to decompose unobserved productivity:

$$\omega_{jt} = E_{t-1}\left(\omega_{jt}\right) + \xi_{jt},$$

or

$$\omega_{jt} = E\left(\omega_{jt}|\omega_{j,t-1}\right) + \xi_{jt},$$

(remember first order Markov property), where $\xi_{jt}$ is independent of all information known in period $t-1$.

- Given the timing assumption that capital was decided in period $t-1$, it must then be that

$$E\left[\xi_{jt}k_{jt}\right] = 0, \tag{6}$$

which you recognize is an orthogonality condition that can be used to estimate the parameters of interest.

- Labour on the other hand is chosen in $t-0.5$, and so at that time part of $\xi_{jt}$ has been observed - hence, $l_{it}$ is not uncorrelated with $\xi_{jt}$:

$$E\left[\xi_{jt}l_{jt}\right] \neq 0.$$

  However, lagged labour, $l_{i,t-1}$, was chosen in period $t-0.5-1$, and so at that point nothing was known about the innovation to productivity in period $t$:

$$E\left[\xi_{jt}l_{j,t-1}\right] = 0. \tag{7}$$

- The two moments (6) and (7) can therefore be used to identify $\beta_k$ and $\beta_l$. This is what happens in the second stage.

- **Second stage.** We have the following population moments:

$$\begin{aligned} E\left[\xi_{jt}k_{jt}\right] &= 0 \\ E\left[\xi_{jt}l_{j,t-1}\right] &= 0, \end{aligned}$$

- Provided we have a random sample, we can appeal to the **analogy principle** and replace population moments by sample moments. We can then obtain consistent estimates of $\beta_k$ and $\beta_l$ by minimizing the criterion function

$$\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\xi_{jt}\left[\begin{array}{c} k_{jt} \\ l_{j,t-1} \end{array}\right]\right]' \cdot C \cdot \left[\sum_{t=1}^{T}\sum_{i=1}^{N}\xi_{jt}\left[\begin{array}{c} k_{jt} \\ l_{j,t-1} \end{array}\right]\right]$$

(1 x 2) (2 x 2) (2 x 1)

with respect to the parameters $\beta_k, \beta_l$. Since the model is exactly identified, the choice of $C$ is irrelevant - the minimum will always occur at zero.

- The computation of $\xi_{jt}$. We saw above that

$$\Phi_t = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt},$$

hence

$$\omega_{jt} = \Phi_t - \beta_k k_{jt} - \beta_l l_{jt}.$$

Also, remember that we have an estimate of $\Phi_t$ from the first stage; thus, conditional on the parameters $\beta_k, \beta_l$ we can compute $\omega_{jt}$.

- Moreover, remember that

$$\omega_{jt} = E\left(\omega_{jt} | \omega_{j,t-1}\right) + \xi_{jt},$$

which we write as a nonparametric regression:

$$\omega_{jt} = \varphi\left(\omega_{j,t-1}\right) + \xi_{jt}.$$

This suggests the following estimation recipe, for the second stage:

1. Guess $\beta_k, \beta_l$ - denote these by $\beta_k^G, \beta_l^G$

2. Compute

$$\omega_{jt}^G = \hat{\Phi}_t - \beta_k^G k_{jt} - \beta_l^G l_{jt}$$

   (remember $\hat{\Phi}_t$ is fixed since estimated in the first stage).

3. Regress $\omega_{jt}^G$ on $\omega_{j,t-1}^G$ using some suitable technique - e.g. linear regression (perhaps allowing for a polynomial), or nonparametric techniques. Compute the productivity innovation, based on the results:

$$\xi_{jt}^G = \omega_{jt}^G - \hat{\varphi}\left(\omega_{j,t-1}^G\right).$$

4. Compute the criterion function

$$\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\xi_{jt}\left[\begin{array}{c} k_{jt} \\ l_{j,t-1} \end{array}\right]\right]' \cdot C \cdot \left[\sum_{t=1}^{T}\sum_{i=1}^{N}\xi_{jt}\left[\begin{array}{c} k_{jt} \\ l_{j,t-1} \end{array}\right]\right]$$

5. Check if this looks like the global minimum; if it does, then STOP (you have obtained your estimates); if not, judiciously change $\beta_k^G, \beta_l^G$ and go back to step 2 above.

Of course, you'd use some pre-programmed minimization routine to do this.

To get standard errors, the easiest procedure is probably to rely on bootstrapping (you should include the first stage as well).

## Relation between ACF and DPD approach

- ACF discuss how their estimator compares with the type of estimator ("DPD") used by Blundell and Bond. The identify distinct advantages and disadvantages of both approaches.

- The main advantage of ACF:

  - Unobserved productivity can follow an arbitrary first order Markov process. That is, ACF can accommodate a nonparametric process, such as $\omega_{jt} = f\left(\omega_{j,t-1}\right) + \xi_{jt}$. This is not possible with the DPD approach. ACF can do this because they recover the unobserved productivity term $\omega_{jt}$. 1st stage estimation is important in this context, as this procedure eliminates the "irrelevant" part of the residual (e.g. output measurement error).

- The main advantage of DPD:
  - Easy to allow for firm fixed effects - i.e. unobserved time invariant heterogeneity across firms. Recall we have suggested earlier in this course that this is arguably the main advantage of having panel data. This is not possible in the ACF approach, if one were modelling the dynamics of $\omega_{jt}$ nonparametrically.

- However, the estimators can be made very similar to each other - if we were using a linear AR1 model for productivity, so that $\omega_{jt} = \rho\omega_{j,t-1} + \xi_{jt}$, then the second stage of ACF would literally be the dynamic COMFAC model adopted by Blundell & Bond.