Estimation of Production Functions

Aviv Nevo, Northwestern University

Winter 2009

1 Introduction

Why do we care about estimating production functions (in an IO course)?

- Production functions are components in many economic models;
- Evaluate efficiency of an industry:
  - increasing returns, economies of scale
  - complementarity / substitutability between inputs
  - cost regulation, synergies and mergers
  - economics of scope
  - learning by doing
- Productivity analysis (output per index of inputs)
  - measurement of productivity differences across firms or over time
  - effects of policy (deregulation, tariffs, etc.)
  - returns of R&D (private and social)
  - returns to adoption of new technology

As a side: I have an additional reason for covering this material, especially early in the course. It is an excellent subject to motivate many very important issues in empirical micro. Unlike, many of the models we will see later, the models of productions functions are linear. It is important to cover many of the issues in linear models before we move on to non-linear estimation.

2 Data

The data should include:

- An output measure (typically either number of units, or value);
- Input measures (labor, capital, R&D, material, energy);
- Being able to follow the same firms over time (i.e., panel data) is very helpful for estimation.

Common sources of data are:

- Compustat: provides the annual and quarterly Income Statement, Balance Sheet, Statement of Cash Flows, and supplemental data items on most publicly held companies in North America. Financial data items are collected from a wide variety of sources including news wire services, news releases, shareholder reports, direct company contacts, and quarterly and annual documents filed with the Securities and Exchange Commission. Compustat files also contain information on aggregates, industry segments, banks, market prices, dividends, and earnings. Depending upon the data set, coverage may extend as far back as 1950 through the most recent year-end. (For more info see http://www.kellogg.northwestern.edu/researchcomputing/compustat.htm
- Longitudinal Research Database (LRD): provides a plant-level database containing detailed statistics on research and development activities of US firms. The database contains detailed company-level research and development information compiled from the annual Industrial Research and Development survey starting in 1972 through 2001. (For more info see http://www.census.gov/econ/overview/ma0800.html)

  Other countries have similar data sets.
- Regulated (or previously regulated) industries: As part of the regulation process many firms had to report detailed cost and production data. For a recent example see Markiewicz, Rose and Wolfram (2004) who use data from FERC to estimate how restructuring improved efficiency in US electricity generating plants.
- Special data sets that were made available for research through a special circumstance (e.g., consulting or personal connections).

3 The Model

A typical starting point is the Cobb-Douglas model

$$Q_{it} = e^{\alpha} L_{it}^{\beta_l} K_{it}^{\beta_k} e^{u_{it}}$$

where Q_{it} is the output of plant (or firm) i at time t, L_{it} is labor (or more generally a variable input), K_{it} is capital (a quasi fixed input), u_{it} is an error term, α, β_l and β_k are parameters to be estimated.

We might include additional right-hand side variable like material, energy, different types of labor (blue/white collar), different types of capital or R&D;

The output measure might be in physical units but in most cases, since it involves aggregation across products will be measured in dollars. In some cases it will be value added rather than sales (i.e., cost of material, energy and other short-term inputs would have already been subtracted from sales).

The error term, u_{it}, includes:

- technology or management differences
- measurement errors
- variation in external factors (e.g., weather)

Taking logs we obtain

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + u_{it}$$

where: $y_{it} \equiv \log(Q_{it})$, $l_{it} \equiv log(L_{it})$, $k_{it} \equiv log(K_{it})$.

4 Econometric Issues

There are several econometric issues we have to deal with in estimating the above equation.

4.1 Specification:

- Functional form? Here we will focus only on Cobb Douglas functions (following much of the applied work). There is a significant body of working examining more flexible functional forms (e.g., Fuss and McFadden)
- What does the equation mean?
  - Aggregation across products
  - Aggregation across plants/firms (or even industries)
  - Since the dependent variable is dollar amount: potentially change in pricing not productivity

4.2 Data:

- measurement error in outputs/inputs
- measurement of capital (different types, depreciation, etc.)
- measurement of labor (types, wage vs work force, etc.)

4.3 Simultaneity:

Observed inputs may be correlated with unobserved shock and therefore OLS will yield biased and inconsistent estimates.

Example 1 *Suppose we observe a cross section of firms; some are more productive (better managers); these firms might also need less labor to produce the same output (and assume they know this therefore hire less). The bottom line these firms will produce more with less labor, thus OLS will underestimate* β_l.

Example 2 *Suppose we observe the same firms over time; in a period the firm gets a higher productivity shock (which it observes) it will hire more labor. OLS will attribute all the increase in output to the change in labor, thus overestimating* β_l.

For most of what follows we will focus on the simultaneity of labor, assuming capital is pre-determined.

4.4 Selection:

Firms observed in the market are not necessarily a random draw from the population of interest. This is especially problematic in panel data, where we observe firms over time. The sample of firms that survive over time might not be random. This will introduce bias.

In what follows we will focus almost exclusively on the simultaneity issue.

5 Solutions to the Simultaneity Problem

In what follows we will focus on the variable input, labor.

Here are some of the solutions that have been offered in the literature:

5.1 Instrumental variable (in a cross section)

look for a variable that is correlated with the variable input, labor, and uncorrelated with the shock. For example, input prices.

Problems:

- input prices might not be observed;

- Might not vary by firm (in a cross section this means this cannot be used, in a panel structure basically a time dummy)
- Even if input prices vary by firm, the variation might be correlated with the error: market power in input market; matching process between firms and inputs;

5.2 Panel Data (to the rescue)

I assume you have seem most of the material in this sub-section before. I will therefore just review it quickly.

- Assume $u_{it} = \alpha_i + \varepsilon_{it}$, where α_i is a firm specific shock and ε_{it} is "white" noise.
- Furthermore, assume that the data include multiple observation for each firm (and are balanced).

Fixed Effects The firm specific effect is a parameter to be estimated (we place no restrictions on it besides the fact that it is fixed).

The model can be estimated in several ways:

Dummy variables estimate a dummy variable for each firm;

First differences estimate by OLS

$$y_{it} - y_{i,t-1} = \beta_l(l_{it} - l_{i,t-1}) + \beta_k(k_{it} - k_{i,t-1}) + \varepsilon_{it} - \varepsilon_{i,t-1}$$

Note: we can also use longer differences (more on this later).

"Within" transformation

$$y_{it} - \overline{y}_i = \beta_l(l_{it} - \bar{l}_i) + \beta_k(k_{it} - \overline{k}_i) + \varepsilon_{it} - \overline{\varepsilon}_i)$$

where $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{l}_i$ and $\overline{k}_i$ are defined similarly. OLS estimation of this equation will yield identical estimates to the dummy variables approach, but will be easier to compute (and does not suffer from the incidental parameters problem).

Note: $\widehat{\alpha}_i$ can be estimated by $\overline{y}_i - \beta_l\bar{l}_i - \beta_k\overline{k}_i$.
Issues:

- Is α_i really fixed? If not the estimates might still be biased

- The "within" transformation reduces the signal to noise ratio, thus measurement error will become more of an issue and the estimates will be biased towards zero (assuming classical ME).
- Requires strict exogeneity: l_{it} must be uncorrelated with $\varepsilon_{i\tau}$ for all t and τ, since $\overline{\varepsilon}_i$ is in the error term.

5.2.1 Random Effects

An alternative approach assumes that α_i is not correlated with labor in capital and has a known (up to parameters) distribution. Therefore, in our model OLS is unbiased and results only in loss of efficiency. For unbiased estimates one can use Total (OLS on the full sample), within, between (OLS on the firm means) or GLS. GLS is the efficient estimator (if the model is correctly specified), but biased if the fixed effects model is correct. On the other hand, the within estimator is unbiased in both cases but not efficient. Thus, we can construct a Hausman test to test between the models.

Note:

- The idea behind random effects is to try to model the heterogeneity across firms. However, in a way this is not "true" heterogeneity since it is only ex-post not ex ante and it does not impact the firm behavior (in the inputs markets);
- Under the RE model there is no bias in LS estimates, in this linear static model. This will not be the case in non-linear or dynamic models.
- Relative to the FE model this model requires more assumption and is less general.

5.3 Correlated Random Effects

This model tries to combine the FE and RE model. Like RE assume a distribution for the firm-specific shocks but allow for correlation with the inputs, like the FE model. The main problem is that we need a model for the correlation. Advantage over FE is that it requires less data, is more efficient if correct, but yields biased estimates if mis-specified. Relative to RE, deals with potential correlation.

5.3.1 Π – Matrix Approach

An approach proposed by Chamberlain (1982) that includes all the above cases, and many more as private cases. The approach has two steps. First, you recover from the data "reduced form" coefficients. Second, you impose the restrictions of the model on these coefficients.

Step 1: Suppose, for ease of exposition, that $T = 3$ and that there is only labor in the model. The idea here is to regress quantity in each period on all leads and lags.

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix} = \text{const} + \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix} * \begin{bmatrix} l_{i1} \\ l_{i2} \\ l_{i3} \end{bmatrix} + \begin{bmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \end{bmatrix}$$

These are not structural equations, just a way of summarizing the data. The coefficients can be estimated using OLS.

Step 2: We now impose the model to obtain structural restrictions on the above coefficients.

Assume, for example, that $u_{it} = \alpha_i + \varepsilon_{it}$, where ε_{it} is white noise and $E(a_i \mid l_{i1}, l_{i2}, l_{i3}) = 0$ (as in the random effects model) then

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix} = \begin{bmatrix} \beta_l & 0 & 0 \\ 0 & \beta_l & 0 \\ 0 & 0 & \beta_l \end{bmatrix}$$

Alternatively assume that $E(a_i \mid l_{i1}, l_{i2}, l_{i3}) = \lambda_1 l_{i1} + \lambda_2 l_{i2} + \lambda_3 l_{i3}$, which is one way to interpret the fixed effects model, then

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix} = \begin{bmatrix} \beta_l + \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_1 & \beta_l + \lambda_2 & \lambda_3 \\ \lambda_1 & \lambda_2 & \beta_l + \lambda_3 \end{bmatrix}$$

The structural coefficients can be estimated by a Minimum Distance procedure which yields

$$\widehat{\theta} = \left(G'V^{-1}G\right)^{-1} G'V^{-1} vec(\widehat{\Pi})$$

where $vec(\widehat{\Pi})$ is a vector containing all the coefficients estimated in Step 1,V is the covariance matrix of these estimates, $\widehat{\theta}$ are the estimated structural parameters and G is the matrix of restrictions such that $vec(\Pi) = G\theta$.

For example, in the random effects model $\theta = \beta_l$, $vec(\widehat{\Pi}) = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & ... & \pi_{33} \end{bmatrix}'$ and $G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}'$.

In the fixed effects model $\theta = \begin{bmatrix} \beta_l & \lambda_1 & \lambda_2 & \lambda_3 \end{bmatrix}'$,

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Chamberlain also provides a test that can be used to test if the restrictions are "close enough" to the data.

Note: The model's restrictions can be imposed directly and in some sense there is no need for the two stage approach. However, the 2-stage approach is appealing:

- It requires estimating the first stage only once. In some cases this might have computational advantages. It also has the advantage of only requiring a few moments from the data (which might be useful if access to the data is restricted);
- It let's one eyeball the data and get a feel for where the model might be failing.

5.4 GMM/ Differencing/Dynamic Panel

As we saw before differencing the data can help get rid of the additive firm-specific effect. There are several possible differences on could take (for example if $T = 4$, three first- differences, two 2nd differences, one 3rd or "long" difference). Which should we take? Can we combine these?

GMM is an easy way to combine the moments. Griliches-Hausman (JOE, 1985) explore this and show that one can use the different differences to learn about the effect of measurement error.

[talk about strict exog – why differencing might be preffered to FE. maybe talk about "system" estimates. mention that we need additional moments to identify rho (lagged output)]

Alternatively, we could consider more complex models. Assume that $u_{it} = \alpha_i + \omega_{it} + \varepsilon$, where

α_i and ω_{it} are "transmitted" (i.e., impact the choice of l_{it});

ω_{it} is not autocorrelated;

In words the "fixed" effect is not really fixed. Taking differences we get

$$y_{it} - y_{i,t-1} = \beta_l(l_{it} - l_{i,t-1}) + \beta_k(k_{it} - k_{i,t-1}) + (\omega_{it} - \omega_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

The difference $(\omega_{it} - \omega_{i,t-1})$ is correlated with $(l_{it} - l_{i,t-1})$, therefore we need an IV. Arellano and Bond (ReStud, 1991) suggest using lagged values of output and the inputs as IV (i.e., use , $l_{i,t-\tau}$ and $k_{i,t-\tau}$, for $\tau \geq 2$). The estimation is based on the conditional moment

$$E\left[u_{it} - u_{i,t-1} \mid (l_{i\tau}, k_{i\tau})_{\tau=1}^{t-2}\right] = 0$$

Notes:

- In practice this approach has performed poorly both with real data and in Monte Carlo studies.
- The reason seems to be that lagged values are weak instruments for the differences.
- The assumption that ω_{it} is not autocorrelated seems strong;

Blundell and Bond (JOE, 1998, Econometric Reviews, 1999) propose getting around these problems. Assume that $u_{it} = \alpha_i + \omega_{it} + \varepsilon$, where
α_i and ω_{it} are "transmitted" (i.e., impact the choice of l_{it});
$\omega_{it} = \rho\omega_{it} + v_{it}$ AR(1) process;

B-B propose using $l_{i,t-\tau}$ and $k_{i,t-\tau}$, for $\tau \geq 2$ as IV for the (quasi-difference) equation

$$(y_{it} - \rho y_{i,t-1}) - (y_{it-1} - \rho y_{i,t-2}) = \beta_l \left((l_{it} - \rho l_{i,t-1}) - (l_{it-1} - \rho l_{i,t-2})\right) + \beta_k \left((k_{it} - \rho k_{i,t-1}) - (k_{it-1} - \rho k_{i,t-2})\right) + \varepsilon$$

where

$$\varepsilon_{it}^* = v_{it} - v_{it-1} + (\varepsilon_{it} - \rho\varepsilon_{i,t-1}) - (\varepsilon_{it-1} - \rho\varepsilon_{i,t-2})$$

The estimation is based on the conditional moment

$$E\left[(u_{it} - \rho u_{i,t-1}) - (u_{it-1} - \rho u_{i,t-2}) \mid (l_{i\tau}, k_{i\tau})_{\tau=1}^{t-2}\right] = 0$$

The idea here can be extended to higher order *linear* processes, but generally not to non-linear process (first, or higher order);

5.5 The Olley Pakes Approach

5.5.1 The model

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

where ω_{it} is transmitted and ε_{it} is white noise.

Assumption 1 (first-order Markov): $P(\omega_{it} \mid I_{it-1}) = P(\omega_{it} \mid \omega_{it-1})$, where I_{it-1} is the firm's information set at time $t-1$.

This assumption is more general than the linear AR(1) assumed by Arellano and Bond. However, it rules out a fixed effect.

Assumption 2 (timing): (i) labor is a non-dynamic input (choice of labor at time t does not impact future profits); (ii) capital choice is dynamic and evolves according to $k_{it} = K(k_{it}, i_{it-1})$, where $K()$ is a deterministic function and $i_{it-1}$is investment at time $t-1$.

The second part of the assumption (together with Assumption 1) implies that $\omega_{it} - E(\omega_{it} \mid I_{it-1}) = \omega_{it} - E(\omega_{it} \mid \omega_{it-1})$ is not correlated with k_{it} (since i_{it-1} is determined at $t-1$).

Assumption 3 (strict monotonicity): $i_{it} = f_t(\omega_{it}, k_{it})$ is strictly increasing in ω_{it}.

Note

- labor is non-dynamic so it does not enter the investment function
- only ω_{it}, no other unobserved shocks, enter the function. An example of such shocks are firm specific input prices.
- f is indexed by t to allow for changes in market conditions (e.g., the macro economy) that change over time and impact all firms.

Assumption 3 allows us to invert the investment decision to recover the unobservable:

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

So if we know f_t^{-1}, we could control for ω_{it}

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + f_t^{-1}(i_{it}, k_{it}) + \varepsilon_{it} \equiv \beta_l l_{it} + \phi_t(i_{it}, k_{it}) + \varepsilon_{it}$$

Since we do not know f_t^{-1} (we need to solve a fairly complex dynamic problem in order to get it), we will estimate it non-parametrically.

5.5.2 Estimation

Step 1 Estimate β_l by regression of y_{it} on l_{it} and a non-parametric estimate of ϕ_t, get $\widehat{\beta_l}$ and $\widehat{\phi_{it}}$.

(For the purpose of the discussion here you can think of a non-parametric estimate as a polynomial expansion of pre-set order, say 4th order).

Step 2 (recovering β_k) We note that

$$\omega_{it} = E(\omega_{it} \mid I_{it-1}) + \xi_{it} = E(\omega_{it} \mid \omega_{it-1}) + \xi_{it}$$

where ξ_{it} is the unexpected innovation in ω_{it}. The second equality follows from Assumption 1.

By construction $E(\xi_{it} \mid I_{it-1}) = 0$, which combined with Assumption 2 implies that

$$E(\xi_{it} \mid k_{it}) = 0$$

How can we use this?
The Step 1 estimates allow us to write $\omega_{it}(\beta_k) = \widehat{\phi_{it}} - \beta_k k_{it}$ and therefore

$$\xi_{it}(\beta_k) = \omega_{it}(\beta_k) - \psi(\omega_{it-1}(\beta_k))$$

where $\psi(\omega_{it-1}) \equiv E(\omega_{it} \mid \omega_{it-1})$. Since we allow for a general (unknown) first-order Markov process ψ is unknown and needs to be estimated, which we will do non-parametically. If we assumed an AR(1) process ψ will just be a linear function.
This can used in several ways. For example, we could choose $\widehat{\beta_k}$ to

$$Min \sum_{it} \xi_{it}^2(\beta_k)$$

Alternatively, we can construct a moment estimator by: (i) for a given β_k; computing $\omega_{it}(\beta_k)$; (ii) non-parametrically regressing ω_{it} on $\omega_{it-1}$to get estimates of ψ; (iii) use these estimates to compute ξ_{it} and (iv) search over β_k to satisfy the sample analog of the above moment condition $\frac{1}{T}\frac{1}{N}\sum_{it}\xi_{it}(\beta_k)k_{it} = 0$.

In summary, we identified β_l by controlling for ω_{it} using the investment decision, which brought in new information. β_k is identified off of the timing decisions.

Note, the model in the paper is more general in two ways:

- They include age in the model. Does not change much.

- Allow for selection, which depends on ω_{it}. This adds a step, between Step 1 and 2, where the probability of exit is estimated non-parametrically. Denote this probability by $\widehat{p_{it}}$. Step 2 is then slightly modified by writing $\psi(\omega_{it-1}, \widehat{p_{it}})$

5.5.3 Levinsohn-Petrin

In many data sets investment is lumpy. In particular many observations with $i_{it} = 0$.

For these observations f_t is not strictly increasing and cannot be inverted.

In principle, the OP approach can still work but it requires that we ignore all the observations with $i_{it} = 0$, which could mean ignoring a lot of data. For this reason Levinsohn-Petrin propose to modify the OP approach.

The consider

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

where m_{it} are intermediate inputs like material, fuel or electricity.

Assume that $m_{it} = f_t(\omega_{it}, k_{it})$ and the f_t is invertible.

Therefore

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

The estimation follows OP very closely.

In Step 1, regress y_{it} on l_{it} and a non-parametric estimate of $\phi_t(m_{it}, k_{it})$, get $\widehat{\beta}_l$ and $\widehat{\phi_{it}}$.

In Step 2 exploit the conditional moment condition:

$$E(\xi_{it}(\beta_k, \beta_m) \mid k_{it}, m_{it-1}) = 0$$

Comments:

- Nice idea to bring in more proxies (allows for test, and possibly a richer model);
- Why if f_t nonparametric? Unlike OP we do not need a dynamic model to specify it. Given the production function and assumptions about input and output prices it could be specified.
- In some sense no new information - what is unique about m_{it}? Why couldn't we use labor?
- Timing assumptions.

5.5.4 Ackerberg-Caves-Frazer

They critique the LP (and to a lesser extant the OP) method. Furthermore, they offer an alternative. I'll just quickly summarize there main criticism and review their alternative.

"Colinearity" Issues Consider Step 1 in LP

$$y_{it} = \beta_l l_{it} + \phi_t(m_{it}, k_{it}) + \varepsilon_{it}$$

The question is why (in the model) will l_{it} vary independently of $\phi_t(m_{it}, k_{it})$? Obviously, they vary independently when we estimate the model (otherwise, we could not get first stage estimates). But the question is this real variation or just a function of us not allowing for a general enough function for ϕ_t.

ACF consider various options on timing.

1) l_{it} chosen at the same time as m_{it} in this case, just like the we described m_{it} we can write

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

No independent variation in l_{it}

2) l_{it} chosen either before or after m_{it} (and ω_{it} evolves between the choices) If l_{it} is chosen after m_{it}, we will get variation in l_{it} (because in includes additional information), but inverting m_{it} will not recover the correct productivity shock.

If l_{it} is chosen before m_{it}, then $m_{it} = f_t(l_{it}, \omega_{it}, k_{it})$ and we will not get the required variation

3) A couple of assumption that might work measurement in l_{it}, but not in m_{it} will generate the required variation and still allow the inversion to recover the productivity shock (note that we need no measurement error in m_{it} for the inversion to work).

Another alternative is that l_{it} is chosen after m_{it}, ω_{it} does *not* evolves between the choices, and there is an additional unexpected error that impact the choice of labor.

The last two alternatives will work but are very specific and it is not clear we want to build on them to justify the procedure

Similar argument can be made for OP (see the ACF paper for details). The bottom line is that similar issues exist in OP, but that the assumptions required to justify the procedure are easier to believe.

Alternative method In addition to pointing out the problem, ACF offer an alternative method (they actually offer more than one, but I will discuss the main one).

Consider the value added production function:

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

The timing is as follows:

- i_{it} (and k_{it}) set at time $t-1$;
- l_{it} set at time $t-b$, $0 < b < 1$;
- m_{it} set at time t; and the production occurs.

The productivity shocks evolves according to $P(\omega_{it} \mid I_{it-b}) = P(\omega_{it} \mid \omega_{it-b})$ and $P(\omega_{it-b} \mid I_{it-1}) = P(\omega_{it-b} \mid \omega_{it-1})$.

Finally, given the timing materials are determined by $m_{it} = f_t(l_{it}, \omega_{it}, k_{it})$, and f_t is invertible.

Therefore:

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

The estimation is in two steps

Step 1, regress y_{it} on a non-parametric function of l_{it}, k_{it} and m_{it}. Note, that unlike before no parameters are estimated only $\widehat{\phi_{it}}$ is recovered (basically the whole purpose of this step is to net out ε_{it}).

Step 2, estimate the coefficients β_l and $\beta_k$using the conditional moment

$$E(\xi_{it}(\beta_l, \beta_k) \mid k_{it}, l_{it-1}) = 0$$

where as before $\xi_{it} = \omega_{it} - E(\omega_{it} \mid \omega_{it-1})$.

The advantage of this approach is that it makes assumptions about the timing that seem reasonable and are consistent with the estimation.

5.5.5 Final Comments

- The OP/LP/ACF estimators can also be estimated in a single step, by stacking up all the relevant moments. the advantage of a one step approach is in computing the standard errors.

- Relation to the dynamic panel literature

  - OP/LP/ACF allow for general first order process (not just a linear AR(1))
  - dynamic panel approach can allow for a fixed effect in addition to the AR(1) process (i.e., more persistence in the productivity shock), while OP/LP/ACF cannot. If a fixed-effect exists the moment condition used in the second step would not be valid.
  - dynamic panel approach does not require scalar unobservable or monotonicity condition.

6 Results

Olley-Pakes

Griliches-Mairesse

Ackerberg-Caves-Frazer

TABLE VI

ALTERNATIVE ESTIMATES OF PRODUCTION FUNCTION PARAMETERS[a]
(STANDARD ERRORS IN PARENTHESES)

Sample:	Balanced Panel		Full Sample[c, d]						
								Nonparametric F_ω	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation Procedure	Total	Within	Total	Within	OLS	Only P	Only h	Series	Kernel
Labor	.851	.728	.693	.629	.628			.608	
	(.039)	(.049)	(.019)	(.026)	(.020)			(.027)	
Capital	.173	.067	.304	.150	.219	.355	.339	.342	.355
	(.034)	(.049)	(.018)	(.026)	(.018)	(.02)	(.03)	(.035)	(.058)
Age	.002	−.006	−.0046	−.008	−.001	−.003	.000	−.001	.010
	(.003)	(.016)	(.0026)	(.017)	(.002)	(.002)	(.004)	(.004)	(.013)
Time	.024	.042	.016	.026	.012	.034	.011	.044	.020
	(.006)	(.017)	(.004)	(.017)	(.004)	(.005)	(.01)	(.019)	(.046)
Investment	—	—	—	—	.13	—	—	—	—
					(.01)				
Other Variables	—	—	—	—	—	Powers of P	Powers of h	Full Polynomial in P and h	Kernel in P and h
# Obs.[b]	896	896	2592	2592	2592	1758	1758	1758	1758

[a]The dependent variable in columns (1) to (5) is the log of value added, while in columns (6) to (10), the dependent variable is the log of value added $-b_l * \log(\text{labor})$.

[b]The number of observations in the balanced panels of regressions 1 and 2 are the observations for those plants that have continuous data over the period, with zero investment observations removed. The 2592 observations used in columns (3), (4), and (5) are all observations in the full sample except those with zero investment. Approximately 8% of the full data set had observations with zero investment. Columns (6) to (10) have fewer observations because the sampling procedures for the Annual Survey of Manufactures forced us to drop observations in years 1978, 1983, and the last year, 1987. See note c.

[c]The number of observations in the last four columns decreases to 1758 because we needed lagged values of some of the independent variables in estimation. This rules out using the first observation on each plant and the first year of the rotating five-year panels that make up the Annual Survey of Manufactures. To check that the difference between the estimates in columns (6)–(9) and those in columns (3)–(5) are not due to the sample, we ran the estimating equations in columns (3)–(5) on the 1758 plant sample and got almost identical results.

[d]Consult the text for details of the estimation algorithm for columns (6) to (10).

Table 3: Alternative Estimates of Production Function Parameters[1]:
U.S. R&D Performing Firms, 1973, 1978, 1983, 1988
(standard errors in parentheses)

Variables[1]	**Sample**					
	Balanced Panel		**Full Sample**[3]			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Total	**Within**	**Total OLS**		**Nonparametric** F	
Labor	.496 (.022)	.685 (.030)	.578 (.013)	.551 (.013)	.591 (.013)	
Physical Capital	.460 (.014)	.180 (.027)	.372 (.009)	.298 (.012)	.321 (.016)	.320 (.017)
R&D Capital	.034 (.015)	.099 (.027)	.038 (.007)	.027 (.007)	.081 (.016)	.077 (.019)
Investment	-	-	-	.110 (.011)	-	
Other Variables[4]	-	-	-	-	Powers of h	Polynomial in P and h
# Observations[2]	856		2971		1571	

(1) The dependent variable in columns 1 to 4 is the log of sales, while in column 5 and 6, the dependent variable is the log(value added) - β*log(labor).

(2) The number of observations in the balanced panel for regressions in columns 1 and 2 are the observations for those firms that have continuous data over the period. Similarly, the 2971 observations in columns 3 and 4 are all observations in the full sample. (Only six observations had to be discarded because of zero investment.) The number of observations in the last two columns (5) and (6) decreased to 1571 because lagged values of some of the independent variables are needed in estimation.

TABLE 1

	Industry 311					
	Capital		Labor		Returns to Scale	
	Estimate	SE	Estimate	SE	Estimate	SE
OLS	0.336	0.025	1.080	0.042	1.416	0.026
FE	0.081	0.038	0.719	0.055	0.800	0.066
ACF – M	0.371	0.037	0.842	0.048	1.212	0.034
ACF – E	0.379	0.031	0.865	0.047	1.244	0.032
ACF – F	0.395	0.033	0.884	0.046	1.279	0.028
LP – M	0.455	0.038	0.676	0.037	1.131	0.035
LP – E	0.446	0.032	0.764	0.040	1.210	0.034
LP – F	0.410	0.032	0.942	0.040	1.352	0.036
DP	0.391	0.026	0.987	0.043	1.378	0.028
	Industry 321					
	Capital		Labor		Returns to Scale	
	Estimate	SE	Estimate	SE	Estimate	SE
OLS	0.256	0.035	0.953	0.056	1.210	0.034
FE	0.204	0.068	0.724	0.087	0.927	0.108
ACF – M	0.242	0.041	0.893	0.063	1.135	0.040
ACF – E	0.272	0.037	0.832	0.060	1.104	0.039
ACF – F	0.272	0.038	0.873	0.061	1.145	0.040
LP – M	0.320	0.037	0.775	0.059	1.094	0.049
LP – E	0.241	0.037	0.978	0.065	1.219	0.047
LP – F	0.254	0.039	1.008	0.062	1.262	0.048
DP	0.320	0.042	0.837	0.064	1.157	0.041
	Industry 331					
	Capital		Labor		Returns to Scale	
	Estimate	SE	Estimate	SE	Estimate	SE
OLS	0.236	0.047	1.038	0.074	1.274	0.052
FE	-0.028	0.103	0.897	0.095	0.869	0.136
ACF – M	0.196	0.064	0.923	0.085	1.119	0.076
ACF – E	0.195	0.065	0.897	0.088	1.092	0.073
ACF – F	0.212	0.062	0.915	0.086	1.127	0.075
LP – M	0.352	0.056	0.678	0.077	1.030	0.072
LP – E	0.305	0.059	0.786	0.086	1.090	0.075
LP – F	0.241	0.052	0.993	0.079	1.234	0.071
DP	0.252	0.054	0.998	0.073	1.249	0.061
	Industry 381					
	Capital		Labor		Returns to Scale	
	Estimate	SE	Estimate	SE	Estimate	SE
OLS	0.223	0.025	1.160	0.045	1.383	0.033
FE	0.036	0.056	0.783	0.077	0.819	0.098
ACF – M	0.262	0.033	1.010	0.053	1.273	0.040
ACF – E	0.250	0.030	1.002	0.053	1.252	0.040
ACF – F	0.259	0.028	1.022	0.051	1.280	0.039
LP – M	0.342	0.038	0.803	0.053	1.145	0.056
LP – E	0.306	0.033	0.944	0.047	1.251	0.044
LP – F	0.265	0.031	1.090	0.049	1.355	0.041
DP	0.275	0.034	1.056	0.053	1.331	0.037