Productivity and Markup Estimation with Heterogeneous Firms

Class notes

Carlo Altomonte

Bocconi University

Presentation Outline

- Definition of Total Factor Productivity (TFP)
  - Aggregate productivity
  - Dealing with heterogeneous firms: labor market and productivity
  - Foster-type and Olley-Pakes decompositions: definition & applications
- TFP Estimation: the simultaneity bias
  - OLS and fixed effects
  - The Levinsohn-Petrin (2003) semi-parametric approach
- TFP Distribution: testing the Pareto assumption
- Measuring markups (PCM) of heterogeneous firms
- Applications to policy-relevant examples

The aggregate production function

- Consider an industry-specific generic production function of the form

$$Y = AK^{\beta_1}L^{\beta_2}$$

  where Y is the total output of a given industry, A is the technology, K and L the inputs of production, β_i the production coefficients. If $\sum \beta_i = 1$ => Constant Returns to Scale.

- We can more conveniently express our production function in log-form, using a vector notation for the production coefficients:

$$y = \beta' x + \Omega \tag{1}$$

  where y is the log-output, $\beta' x = \beta_1 k + \beta_2 l$ is the production function with its production coefficients β_i and log inputs $x = (k, l)$, and Ω is the log of total factor productivity A (TFP, or Solow residual)

Production function and firm heterogeneity

- One implication of Equation (1) is that any redistribution of inputs across plants (i.e. any linear combination of $\beta' x$) results in the same aggregate output y, which is not a realistic assumption: for example, if firms are hetereogeneous in productivity levels and new inputs flow to the most productive firms, different linear combinations of $\beta' x$ would yield different results in terms of y.

- Hence, it is better to start from firm-level productivity estimates of the form

$$y_i = \beta' x_i + \omega_i \tag{2}$$

where the sub-index denotes firm i, and then aggregate Equation (2) back to (1).

- Question: how to aggregate firm level measures from our estimates? how to calculate ω_i?

Retrieving Aggregate Total Factor Productivity

- Suppose you have correctly estimated your production function coefficients $\widehat{\beta}$ (see next section). Then, you have that

$$\widehat{\omega}_i = y_i - \widehat{y}_i = y_i - \widehat{\beta}' x_i \tag{3}$$

i.e. ω_i is the "Solow residual" of the production function estimate (the "measure of our ignorance"....).

- Aggregate productivity for all firms $i: \ \Omega = \sum_i s_i \widehat{\omega}_i$, obtained as a weighted average of the firm-specific productivity ω_i, using output or input shares s_i as weights. Note again that Ω is (the log of) our aggregate productivity term A in the original production function.
- Now, denote

$$\Delta\Omega_t = \sum_{i=1}^{N} s_{it}\omega_{it} - \sum_{i=1}^{N} s_{it-1}\omega_{it-1}$$

i.e. a decomposition of changes in productivity $\Delta\Omega_t$ on the basis of firm-specific (i) terms embedding allocative (s_i) and productive/dynamic efficiency (ω_i)

TFP Decompositions - Foster

- To precisely identify these terms, we manipulate the expression as in Foster et al. (2002) in order to obtain:

$$\Delta\Omega_t = \sum_{i\epsilon C} s_{it-k}\Delta\omega_{it} + \sum_{i\epsilon C} \Delta s_{it}(\omega_{it-k} - \Omega_{t-k}) + \sum_{i\epsilon C} \Delta s_{it}\Delta\omega_{it} + \\ + \sum_{i\epsilon E} s_{it}(\omega_{it} - \Omega_{t-k}) - \sum_{i\epsilon X} s_{it-k}(\omega_{it-k} - \Omega_{t-k}) \tag{4}$$

  where $N =$ total number of plants; C = plants who continue their business over time; E = plants who enter at a given time and X = plants who exit, while Ω_{t-k} is the weighted average productivity at the beginning of the period

- The first three terms of the decomposition are known as the "within", "between" and "covariance" component of firms'contribution in productivity, while the last two terms account for the net entry effects.
- Note on the covariance term $\Delta s_{it}\Delta\omega_{it}$: if this is positive, it means that firms who are becoming more (less) productive over time are also able to attract more (less) workers; if it is negative or non significant, then the functioning of the labor market (wage-setting mechanism) contributes negatively to productivity growth.

TFP Decompositions - Foster

- the within-firm effect is the change attributable to the productivity within a firm given its initial market share: a positive sign implies that firms, controlling for their size, are growing more productive over time;
- the between-firms effect accounts for the switch of market shares between firms, keeping the productivity constant w.r. to a benchmark, that is it captures the gains in aggregate productivity coming from the expanding market of high productivity firms, or from low-productivity firms' shrinking shares;
- the covariance (or cross) term gives information about the underlying market adjustment in size and productivity: a positive sign would indicate that market shares and productivity are changing in the same direction, that is firms able to increase (decrease) their productivity are also able to grow larger (smaller) in size/employment, with positive effects for overall growth; a negative sign would show that productivity and market shares are moving in different directions, ie firms whose productivity is decreasing are growing larger and vice versa, with negative effects for underlying productivity growth.
- the net entry effect indicates the extent to which the market is able to select firms in accordance with their competitiveness, that is whether firms less productive than a given threshold are forced to exit the market, while relatively productive firms are able to enter.

TFP Decompositions - Olley and Pakes

- An alternative decomposition has been proposed by Olley and Pakes (1996):

$$\Omega_t = \bar{\Omega}_t + \sum_i \Delta s_{it} \Delta \omega_{it}$$

where $\bar{\Omega}_t$ is the unweighted average of firms productivities ($\frac{1}{N} \sum_i \omega_i$), $\Delta s_{it} = s_{it} - \bar{s}_{it}$ and $\Delta \omega_{it} = \omega_{it} - \bar{\omega}_{it}$

- To derive it, note that $\Omega_t = \sum_i s_{it} \omega_{it} = \sum_i (\bar{s}_{it} + \Delta s_{it})(\bar{\omega}_{it} + \Delta \omega_{it})$
- In this case, if $\Delta s_{it} \Delta \omega_{it}$ is positive, it implies that firms with above average productivity compared to other firms (not over time!) display above average market shares in a given year
- So, the OP decomposition compares productivity allocation **across firms** in a given year; Foster-type decompositions compare productivity growth **within firms** over time

Production function estimation - OLS

- Start from the following firm-specific Cobb-Douglas production function:

$$Y_{it} = A_{it} L_{it}^{\alpha} K_{it}^{\beta} M_{it}^{\gamma}$$

where A_{it} = Total Factor Productivity, Y is output, L is labour, K is capital and M.are material costs.

- Log-linearizing we get:

$$y_{it} = \alpha l_{it} + \beta k_{it} + \gamma m_{it} + \omega_{it}$$

- Ideally, you want to estimate via OLS this equation, i.e. you estimate the values of $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\gamma}$. Then you can calculate the value of the predicted output, based on the estimated PF parameters

$$\widehat{y_{it}} = \widehat{\alpha} l_{it} + \widehat{\beta} k_{it} + \widehat{\gamma} m_{it}$$

and finally retrieve the Solow residual

$$\omega_{it} = y_{it} - \widehat{y_{it}} = y_{it} - \widehat{\alpha} l_{it} - \widehat{\beta} k_{it} - \widehat{\gamma} m_{it}$$

Simultaneity bias

- The term ω_{it} in the previous equation is essentially an error term of an econometric regression, while from an economic point of view it is a "shock" (e.g. a discovery) that allows the firm to better use its inputs.
- Are we sure that, given its nature of "shock", we can properly measure ω_{it} in an unbiased way?
- Actually, if you work with balance data on firms, these are year-end data (e.g. December). If the shock is observed by the firm, say, in March (but not by you...you can only measure the shock indirectly, as discussed above) the firm can change the input decision in the same year => i.e. there is a relationship between the productivity shock and the input decisions
- From an econometric point of view, this means that the error term ω_{it} is correlated with the other independent variables
- **biased OLS estimates** of $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\gamma}$

Simultaneity bias and fixed effects

- In principle, to the extent that the term ω_{it} is firm-specific, the simultaneity bias can be easily dealt with by using fixed effects in a panel-type OLS regresssion, thus delivering consistent estimates of the parameters and thus TFP.

- There are however two drawbacks to this method:

1. A substantial part of the information in the data is left unused. A fixed-effect estimator identifies only through the within (across time) variation, which in micro datasets tends to be much lower than the cross-sectional one (across firms), or possibly not even big enough to allow for proper identification => at best the coefficients will be weakly identified.
2. The assumption that ω_{it} is fixed over time goes against the (macroeconomic) evidence of business cycles in productivity, thus making the whole use of fixed-effects invalid

The Levinsohn and Petrin (2003) solution

- Let y_t denote (the log of) a firm's output (as proxied by revenue) in a Cobb-Douglas production function of the form

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \beta_m m_t + \omega_t + \varepsilon_t \tag{A1.1}$$

  where y_t is output, as proxied by revenues, l_t and m_t denote the (freely available) labour and intermediate inputs in logs, respectively, and k_t is the logarithm of the state variable capital.
- The error term has two components: ε_t, which is uncorrelated with input choices, and ω_t, a productivity shock unobserved by the econometrician, but observed by the firm.
- Since the firm adapts its input choice as soon as it observes ω_t, inputs turn out to be correlated with the error term of the regression, and thus OLS estimates of production functions yield inconsistent results, typically a lower bias of the labor coefficient.

- To correct for this problem, Levinsohn and Petrin (2003), from now on LP, assume the demand for intermediate inputs m_t (e.g. material costs) to depend on the firm's capital k_t and productivity ω_t, and show that the same demand is monotonically increasing in ω_t. Thus, it is possible for them to write ω_t as $\omega_t = \omega_t(k_t, m_t)$, expressing the unobserved productivity shock ω_t as a function of two observables, k_t and m_t.
- To allow for identification of ω_t, LP follow Olley and Pakes (1996) and assume ω_t to follow a Markov process of the form $\omega_t = E[\omega_t|\omega_{t-1}] + \xi_t$, where ξ_t is a change in productivity uncorrelated with k_t.
- Through these assumptions it is then possible to rewrite Equation (A1.1) as

$$y_t = \beta_l l_t + \phi_t(k_t, m_t) + \varepsilon_t \tag{A1.2}$$

where $\phi_t(k_t, m_t) = \beta_0 + \beta_k k_t + \beta_m m_t + \omega_t(k_t, m_t)$.

- By substituting a third-order polynomial approximation in k_t and m_t in place of $\phi_t(k_t, m_t)$, LP show that it is possible to consistently estimate the parameter $\widehat{\beta_l}$ and $\widehat{\phi_t}$ in Equation A1.2.
- For any candidate value β_k^* and β_m^* one can then compute a prediction for ω_t for all periods t, since $\widehat{\omega_t} = \widehat{\phi_t} - \beta_k^* k_t - \beta_m^* m_t$ and, using these predicted values, estimate $E[\widehat{\omega_t|\omega_{t-1}}]$. Then the residual generated by β_k^* and β_m^* with respect to y_t can be written as

$$\widehat{\eta_t + \xi_t} = y_t - \widehat{\beta_l} l_t - \beta_k^* k_t - \beta_m^* m_t - E[\widehat{\omega_t|\omega_{t-1}}] \tag{A1.3}$$

- Equation (A1.3) can then be used to identify β_k^* and β_m^* using the following two instruments:
  - if the capital stock k_t is determined by the previous period's investment decisions, it then does not respond to shocks to productivity at time t, and hence $E[\eta_t + \xi_t | k_t] = 0$;
  - also, if the last period's level of intermediate inputs m_t is uncorrelated with the error period at time t (which is plausible, e.g. proxying intermediate inputs with material costs), then $E[\eta_t + \xi_t | m_{t-1}] = 0$.
- Through these two moment conditions, it is then possible to write a consistent and unbiased estimator for β_k^* and β_m^* simply by solving

$$\min_{(\beta_k^*, \beta_m^*)} \sum_h [\sum_t (\widehat{\eta_t + \xi_t}) Z_{ht}]^2 \qquad \text{(A1.4)}$$

  with $Z_t \equiv (k_t, m_{t-1})$ and h indexing the elements of Z_t.
- All this algorithm is implemented by the STATA routine *levpet* which automatically calculates this semiparametric derivation of TFP.
- Note that you might need to install the routine in STATA by looking for it in the 'Help => Search' menu

- In their paper LP also show how to estimate a simplified algorithm in which, rather than revenues y_t, information is available on a firm value-added, which can be constructed as revenues minus material costs, that is $y_t - m_t$.
- If the latter hold, the polynomial expansion $\phi_t(k_{t,} m_t)$ simplifies and becomes easier to estimate. This is the default in STATA when running the routine. If you want to run the LP algorithm in revenues, type *revenue* at the end of the command line
- The syntax of the *levpet* command looks as follows

levpet *depvar*, **free**(*labor variables*) **proxy**(*material var*) **capital**(*capital var*)

- The complete syntax is

```
levpet depvar [if] [in] , free(varlist) proxy(varlist) capital(varname) [ [
   valueadded | revenue ] justid grid i(varname) t(varname) reps(#)
   level(#) ]
```

The algorithm also utomatically calculates your vector of productivity, generating a new variable (you can call it as you wish...) where the residuals of the estimation (TFP) are stored, *already in their exponential form*

predict *newvar*, **omega**

- This is a typical output, as shown in the STATA manual of the command written by Petrin, Levinsohn and Poi (and posted in the course website)

```
. tsset ppn year
. levpet lnva, free(lnb lnw) proxy(lne) capital(lnk) /*
    */ valueadded reps(250)
```

The output looks very similar to most of Stata's [XT] commands:

```
Levinsohn-Petrin productivity estimator

Dependent variable represents value added.      Number of obs      =      2713
Group variable (i): ppn                         Number of groups   =       556
Time variable (t): year
                                                Obs per group: min =         1
                                                               avg =       4.9
                                                               max =        10

------------------------------------------------------------------------------
        lnva |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnb |   .4659176   .0443553    10.50   0.000     .3789828    .5528524
         lnw |    .420271   .0345444    12.17   0.000     .3525653    .4879768
         lnk |   .2250087   .0646426     3.48   0.000     .0983115    .3517059
------------------------------------------------------------------------------
Wald test of constant returns to scale: Chi2 =    2.96 (p = 0.0856).
```

- Note that here the free variable input (labor) is split in two: blue collars (lnb) and white collars (lnw); the proxy instrumental variable, material costs, are measured as the cost of energy (lne), while capital is standard (lnk).
- Note how the estimate in value added yields coefficients on labor and capital, being material costs already embedded in the depvar (value added is defined as revenues - material costs). An estimate with revenues as depvar would yield an estimate also for the material costs coefficient

Estimating the 'shape' of a Pareto-distribution

- A convenient parameterization of TFP distribution, consistent with the available empirical evidence, is the Pareto-distribution.
- A variable is Pareto-distributed with skewness k and support $[X_m, \infty]$ when its cumulative distribution is

$$F(X) = 1 - \left(\frac{X}{X_m}\right)^{-k}$$

- After a logarithmic transformation, the latter can be rewritten as

$$\ln(1 - F(X)) = k \ln(X_m) - k \ln(X)$$

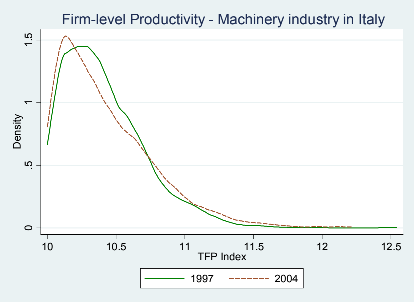
- As shown by Norman, Kotz and Balakrishnan (1994), the OLS estimate of the slope parameter k in the regression of $\ln(1 - F(X))$ on $\ln(X)$ plus a constant is a consistent estimator of k and the corresponding R^2 is close to one.

Interpreting the 'shape' of a Pareto-distribution

- The estimated k can be used to trace the extent of the reallocation effect induced by a trade shock:
- Larger k implies a less dispersed distribution, i.e. one in which a relatively larger share of small and unproductive firms operate, thus with higher room for reallocation
- In other words, a trade shock leading to reallocation of market shares towards the most productive firms should yield a decreasing k over time

The reallocation effect at work

- The aggregate average productivity of this industry is decreasing, but Italian market shares are not falling, rather, they are increasing in value. A paradox? No. As firm level data show, the 'relevant' competitiveness of the industry (the right tail of firms) is actually improving
- Competitiveness-related policies should promote the 'thickening' of the right hand tail of firms via selection and reallocation of resources; policies aimed at social cohesion should deal with the exiting firms => two objectives = two distinct policies: there is no 'average' policy for the industry



Methods to estimate markups

The **markup** is defined as the the amount by which the cost of a product is increased in order to derive the selling price (i.e. a %).

The literature reports two methods to estimate **firm-specific markups**:

- an accounting method, which aims at measuring the price-cost margin (**PCM**) directly from balance sheet data
- De Loecker & Warzynski (2012), which yields markup estimates through the production function estimation (**DLW**, henceforth).

Economic meaning of markups

The markup can be considered:

- a **rivalry indicator** of competition**:** it relates to the average profitability of a given industry and gives an inverse measure of the intensity of competition
- a **structural indicator** if looked from the firm level perspective**:** firms set PCM strategically on the basis of the *residual demand*. This strategic reaction is function of the market structure.

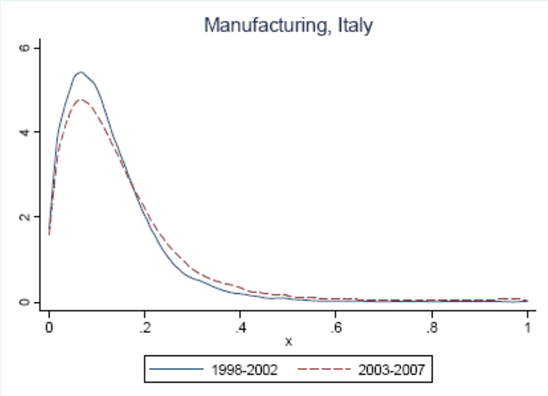
Construction of PCM - accounting method

- The PCM, or Lerner Index, is calculated as $(P - MC)/P$ and tells us how far a firm's price is from its marginal cost.
- Unfortunately MC are not retievable from balance sheets and so the literature (eg Tybout, 2003) proxies them with variable costs thus obtaining:

$$PCM_{it} \simeq \frac{\text{sales}_{it} - \text{variable\_costs}_{it}}{\text{sales}_{it}} = \frac{(p*q)_{it} - (c*q)_{it}}{(p*q)_{it}} = \frac{p_{it} - c_{it}}{p_{it}}$$

where variable_costs = costs for materials + costs for employees

- An example of a firm-level PCM distribution



DLW markup estimation method

- De Loecker & Warzynski (AER, 2012) introduced an alternative approach to obtain consistent estimates of firm-specific markups
- The idea is to interpret the markup as the ratio of an input's output elasticity and its revenue share:

$$\mu_t = \theta_t^X \left(\frac{P_t^X X_t}{P_t Y_t} \right)^{-1} = \theta_t^X (\alpha_t^X)^{-1}$$

where:

- Y_t is the total output produced by the firm
- P_t is the price charged by the firm
- X_t is a generic input employed in production, labor for instance
- P_t^X is the cost of the input X_t
- θ_t^X is the output elasticity of input X_t
- α_t^X is the revenue share of input X_t

DLW markup estimation method

If we consider the labor input L_t:

- α_t^L can be easily computed using balance sheet data about sales, number of employees and cost of employment of firms
- θ_t^L can be recovered by production function estimation.

If we assume a Cobb-Douglas production function of the type:

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \beta_m m_t + \omega_t + \varepsilon_t$$

the output elasticity of labor is equal to:

$$\theta_t^L = \frac{\partial y_t}{\partial l_t} = \beta_l$$

DLW markup estimation method

- The DLW method employs the so-called Ackerberg-Caves-Fraser (2006) algorithm to estimate the production function: in general, methods other than LP are used throughout the literature to estimate productivity when markup estimation has to be performed (e.g. Wooldridge 2009).
- However, the estimated production function coefficients obtained from LP and other estimation methods are comparable. The following table summarizes production function coefficients for the food sector (NACE15) obtained both with LP and Wooldridge algorithms within EFIGE.

	LP	**Wooldridge**	*Delta*
$\widehat{\beta}_l$	0.5568***	0.5102***	0.0466
$\widehat{\beta}_k$	0.2916***	0.3520***	-0.0604

DLW markup computation

We use Efige-Amadeus data to compute the markup charged by a generic firm. If we assume a Cobb-Douglas technology for this firm and consider L_t as the input to calculate markups, the formula for μ_t becomes:

$$\mu_t = \theta_t^L \left(\frac{P_t^L L_t}{P_t Y_t}\right)^{-1} = \widehat{\beta}_l \left(\frac{costs\_of\_employees_t}{sales_t}\right)^{-1}$$

where:

- $\widehat{\beta}_l$ is the estimated coefficient for labor, equal to 0.5102 in our example
- $costs\_of\_employees_t$ are the total costs of labor at t, equal to 371 for the firm considered
- $sales_t$ are the total sales made by the firm in year t, equal to 850.93 in our case

PCM vs. DLW markups

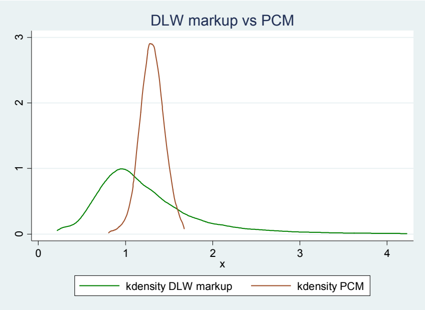
The markup computed with the DLW method for the firm considered in our example is equal to:

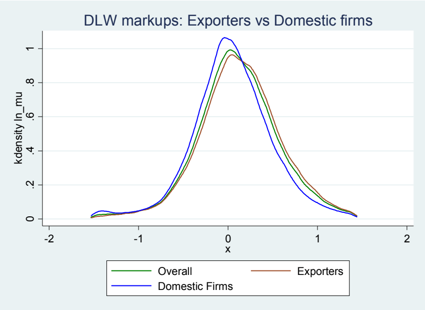
$$\mu_t = \widehat{\beta}_l \left(\frac{costs\_of\_employees_t}{sales_t} \right)^{-1} = 0.5102 \left(\frac{371}{850} \right)^{-1} = 1.1702$$

While, the markup computed with the PCM method for the firm considered in our example is equal to:

$$1 + PCM_{it} = 1 + \frac{\text{sales}_{it} - \text{variable\_costs}_{it}}{\text{sales}_{it}} = 1 + \frac{1482 - 1001}{1482} = 1.3242$$

PCM vs. DLW markups





A comparison: PCMs vs DLW markups

As we can note from the graphs above:

- The PCM calculated with the accounting method (Tybout, 2003) is always higher than the markup computed with the DLW method. The upward bias of the PCM can be attributed to the partial variability of the cost of capital: the numerator of the PCM would decrease if variable capital costs were included in variable costs, thus lowering the PCM.
- Exporters exhibit higher markups compared to domestic firms. This finding is related to the positive relation between productivity and markups, i.e. the most productive firms are able to enter foreign markets and set higher prices.

References

1. Ackerberg, D, Caves, K. Frazer, G. (2006) "Structural identification of production functions," MPRA Paper 38349, University Library of Munich, Germany.
2. Altomonte, C., Aquilante, T., Bekes, G., Ottaviano, G.I.P. (2013) "Internationalization and innovation of firms: evidence and policy", Economic Policy, 28(76): 663-700.
3. Altomonte, C., Barba Navaretti, G., di Mauro, F. and Ottaviano, G.I.P. (2011) "Assessing competitiveness: how firm-level data can help", Bruegel Policy Contribution n. 16, November 2011, Brussels.
4. Barba Navaretti, G., Bugamelli, M., Schivardi, F., Altomonte, C., Horgos, D. and Maggioni, D. (2011) "The Global Operations of European Firms", Bruegel Blueprint IV, Brussels.
5. De Loecker, J., Warzynski, F. (2012) "Markups and Firm-Level Export Status" American Economic Review, 102(6): 2437-71.
6. Lucia Foster, L., Haltiwanger, J., Krizan, C.J. (2006) "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s," The Review of Economics and Statistics, 88(4):748-758.
7. Levinsohn, J. and A. Petrin (2003) "Estimating Production Functions Using Inputs to Control for Unobservables", The Review of Economic Studies, 70(2), pp. 317-341.

References

8. Norman, L., S. Kotz and N. Balakrishnan (1994) Continuous Univariate Distributions, Volume 1, 2nd Edition, Wiley.
9. Olley, G. S. and A. Pakes (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry", Econometrica, 64(6), pp. 1263-1297.
10. Tybout, J. (2003). Plant- and Firm-level Evidence on the New Trade Theories. Handbook of International Trade, Oxford: Basil-Blackwell
11. Wooldridge, J. M.(2009) "On estimating firm-level production functions using proxy variables to control for unobservables," Economics Letters, 104(3): 112-114.