PhD Course: Development Economics (Micro)
Firms and Farms

Måns Söderbom, University of Gothenburg

19 November 2009

1 Introduction

Over the last two decades there have been radical changes in economic policy in many developing countries. A common factor in these changes has been a transition from economies where government controls were extensive to more open, market-oriented, regimes.

The private sector plays a crucial role in this process, and it is important that this sector performs well. In fact, the main reason we are interested in understanding the determinants of performance in the private sector is that we believe better performance raises the incomes and the standards of living of people in poor countries. In this lecture we have a look at the microeconomics of firms and farms in developing countries, focussing on issues related to productivity (a key indicator of performance).

Core references:

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

Martin, W. and D. Mitra “Productivity growth and convergence in agriculture versus manufacturing”, Economic Development and Cultural Change, 2001, 49 (2).

Söderbom, M. and F. Teal “Size and Efficiency in African Manufacturing Firms: Evidence from Firm-Level Panel Data” Journal of Development Economics, 2004, 73, pp. 369-394.

Sleuwaegen, Leo, and Micheline Goedhuys. 2002. “Growth of Firms in Developing Countries, Evidence from Côte d’Ivoire.” Journal of Development Economics 68: 117-135.

Optional:

Tybout, J. R., 2000. Manufacturing firms in developing countries: how well do they do and why? Journal of Economic Literature 38, 11-44.

2 Estimating and Analyzing the Productivity of Firms & Farms

- 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 role of **scale** is important:

  – Looking at labour productivity (defined as some measure of output per worker) across **firms** of differing size, it's very clear it is higher in large than small firms. Söderbom and Teal, Ghana (2004).

  – In contrast, a widespread view is that small **farms** are more productive than large farms. As we shall see, the paper by Lamb (2001) refutes this notion, concluding that large farms are as productive as small ones.

- The most common analytical tool used to assess productivity differences at the micro level is the **production function**. I will now discuss some methodological issues that arise whenever we estimate production functions using micro data.

2.1 The production function

You know from basic macroeconomics that the notion of technology, or productivity, is central. Consider the production function

$$y_i = \alpha l_i + \beta k_i + a_i,$$

where y_i, l_i, k_i denote log output, labour and capital, respectively, and a_i is total factor productivity (TFP). TFP is typically assumed unobserved (at least partially).

Returns to scale Why are the parameters α and β of economic interest? We said above that one of the issues in this literature concerns how performance varies by firm or farm size. In particular, if the economy's fixed set of inputs is allocated to a small number of large firms results in more aggregate output than if allocated to a large number of small firms, it would be efficient to steer resource allocation towards large firms. This is essentially a discussion of returns to scale.

Consequently, one null hypothesis often tested for in production function studies is $H_0 : \alpha + \beta = 1$.

- If we cannot reject this hypothesis then we cannot reject the hypothesis that the production function exhibits constant returns to scale.

- If $\hat{\alpha} + \hat{\beta} > 1$ this is evidence for increasing returns.

- In fact, the evidence on returns to scale in developing countries is most consistent with constant returns to scale (see Söderbom and Francis Teal, 2004). That is, while there are many small firms in developing countries, this does not imply foregone scale economies.

Total factor productivity (TFP) An important measure of firm performance is **total factor productivity**. Is it the case, for example, that foreign owned firms can get more output out of a given set of inputs than domestic firms? Are firms that export more efficient in transforming inputs into output than non-exporters? To answer questions like these we need to estimate TFP (because, as already noted, TFP is generally not observed). Clearly this requires reliable estimation of α and β.

So you see production functions are useful in several ways. Estimation of production functions is, however, not entirely straightforward. The two most likely sources of bias are

- Omitted variables correlated with the inputs

- Measurement errors in explanatory variables

Omitted variables It seems quite possible that the firm's capital and labour decisions are influenced by factors that are unobserved to the econometrician. This would set up a correlation between the regressors and the residuals. To illustrate, consider a simple Cobb Douglas production (no capital, for simplicity)

$$Y_i = A_i L_i^{\beta} \tag{1}$$

where $\beta < 1$, Y_i, L_i are observed measures of output, and labour respectively, and A_i is a TFP, is *observed by the firm but not by the econometrician*. The firm chooses inputs and output to maximize net revenue

$$R_i = P_i Y_i - W_i L_i \tag{2}$$

where W_i is the wage rate and P_i the unit output prices. Assuming that both prices are exogenous, the first order condition for optimal labour is

$$P_i \left(\frac{\partial Y_i}{\partial L_i} \right) = W_i, \tag{3}$$

which can be written as

$$\ln L_i = \left(\frac{1}{1-\beta}\right)(\ln \beta + \ln P + \ln A_i - w_i)\,.$$

Clearly $\ln L_i$ and $\ln A_i$ are correlated, so if we estimate the production function

$$\ln Y_i = \beta \ln L_i + \ln A_i$$

by OLS we will not obtain reliable (unbiased) estimates of β. This is the basic endogeneity problem for estimation of production functions.

If the omitted variable is **time invariant** within firms (but varying across firms), a panel data approach will solve the problem as the unobservables will be absorbed by the fixed effect. One plausible example of an unobserved time invariant factor is 'managerial quality'.

However, there may be time varying unobservables which are correlated with the inputs. In such a case we cannot identify the technology parameters by OLS or fixed effects. In that case, an instrumental variable approach appears to be the best way of addressing the problem. Possible instruments: factor prices if observed (have to be uncorrelated with TFP of course); possibly lagged inputs.

Measurement errors The problems posed by measurement errors are different to those posed by omitted variables. 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

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

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

$$p \lim \hat{\beta}^{OLS} = \beta \left(\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_u^2} \right),$$

where $\sigma_{x^*}^2$ is the variance of the true, unobserved explanatory variable, and σ_u^2 is the variance of the measurement error. The operator $p \lim$ can be thought of as showing the value of estimated β 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. In particular estimating the coefficient on the capital stock whilst controlling for fixed effects has proved difficult - see Söderbom and Teal, 2004, for details.

3 Size and productivity in Ghana's manufacturing sector

Reference: Söderbom and Teal (2004).

- Manufacturing - an engine of growth? Even though manufacturing is far from the largest sector in African economies, it is often considered "special", for reasons discussed above (engine of growth argument).

- But with a few exceptions (Mauritius), African manufacturing is not doing well:

- The share of SSA in world mfg value-added: 0.8 percent; the share in world mfg exports: 0.7 percent.

- The share in world income: 1.1 percent; the share in world population: 11 percent.

- 1972-2002: mfg value-added per capita in SSA fell from USD 98 to USD 85 in constant 1995 values.

• More generally, foreign investors do not see Africa as a promising location for investment (extensive outsourcing to Asia - very little to Africa). Africans themselves keep a large share of their wealth outside Africa (40% according to Collier, Hoeffler and Pattillo, 2001).

• Still, the view that Africa's private sector can contribute to poverty reduction by generating more jobs remains widespread. How generate success?

- In the early 1990s extensive manufacturing surveys were initiated in Africa as part of the World Bank's Regional Program on Enterprise Development (RPED). Generated rich firm-level data, which opened up new avenues for research. Numerous similar surveys have now been fielded in sub-Saharan Africa, and quite a few academic papers based on the data have been written.

- An important general insight based on the firm-level data is that there is substantial heterogeneity across firms, within countries, with respect to key economic variables. **Most** firms have not fared well during the last 10-15 years - but **some** have performed very well indeed.

- Understanding the causes, consequences and implications of heterogeneity in performance across firms is important. To this end, firm-level data must be available.

3.1 Söderbom& Teal: Motivation and theoretical underpinnings

Three prominent issues in policy discussions of the problems facing firms in developing countries:

1. Firms in developing countries lack the technical capacity to perform well. “Without an increase in proficiency, the responsiveness of output to even the best designed structural adjustment program is likely to be limited. Prices are one-half of a scissor, the other being technical skill”, Pack (1993, p. 1).

2. Technology differences explain factor choices across firms of differing size (this argument is usually attributed to work done by Howard Pack in the

1970s & 80s). Small firms have come to be identified with more labour intensive technologies - hence promoting small-scale enterprises is seen as a means of creating jobs.

3. Factor prices differ by firm size, due to market failures. This explains why larger firms are more capital intensive than smaller ones - larger firms face a lower cost of capital and a higher cost of labour.

Söderbom & Teal investigate the relative importance of the possible reasons for poor performance in African manufactures: lack of skills and the extent of scale, technical and allocative inefficiency.

- Data: A seven year panel of plant-level data from Ghana's manufacturing sector. The existence of a panel means we can estimate a production function controlling for time-invariant unobserved skills. Data on the human capital in the firms exist, hence can investigate whether skills play an important role in the efficiency with which firms perform.

3.1.1 Technology and factor choice

The firm's demand for capital (K) and labour (L) depends on the firm's technology and factor prices.

Factor intensities (K/L) vary substantially across firms. Two possible mechanisms:

- Technology is homothetic and the relative cost of capital decreases with size. In this case large firms choose more capital per employee than small ones because capital is relatively cheaper. Why would factor prices vary across firms? In a perfect market they would equalize, ruling out arbitrage, yes? Indeed. Varying factor prices in the cross-section is indicative of market failures - e.g. perhaps limited access to credit for small firms drives up the shadow cost of capital.

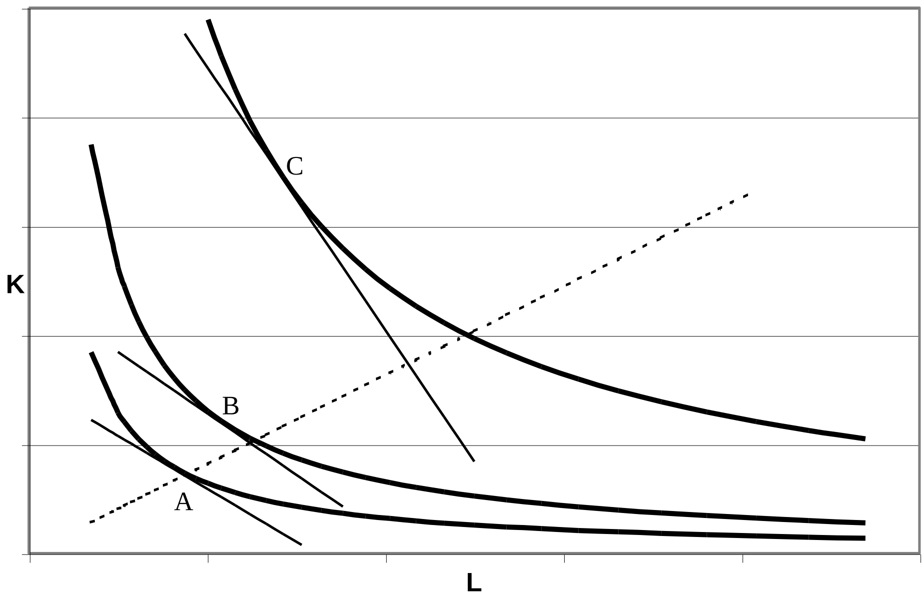
- Technology is non-homothetic and the relative factor prices are constant across firms. Large firms have higher capital-labour ratios because of the nature of the technology, it's got nothing to do necessarily with varying prices or market failures.

[Figure: nonhomothetic technology; varying relative factor prices]

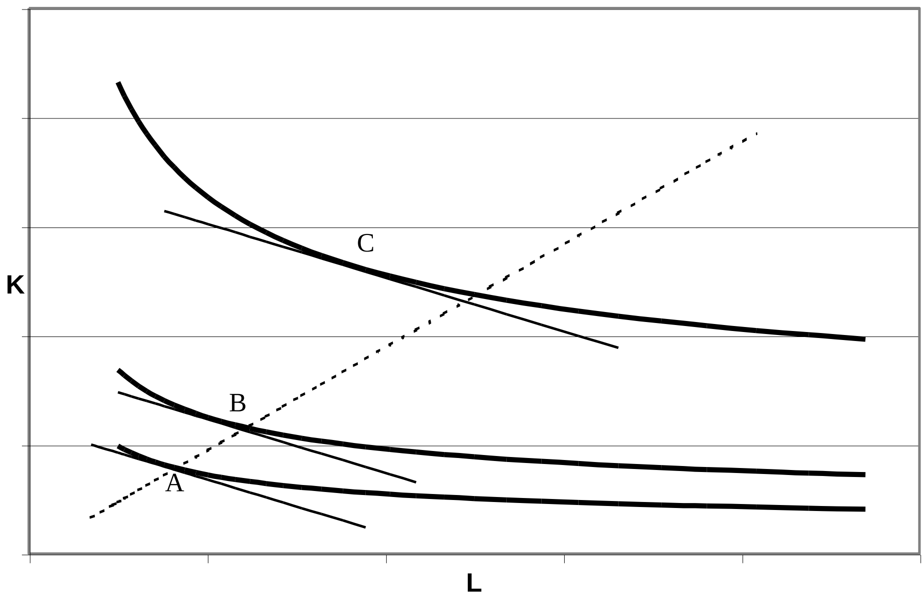
FIGURE 1

TECHNOLOGY AND FACTOR PRICES

A. Homothetic Technology, Variable Factor Price Ratio



B. Non-Homothetic Technology, Constant Factor Price Ratio



3.1.2 Total Factor Productivity

Consider the simple production function again:

$$y_i = \alpha l_i + \beta k_i + a_i.$$

What kinds of question might be interesting to ask about a_i?

- What's the average level of TFP (first moment)? Does it vary with observable characteristics (e.g. ownership, location)?

- How much variation is there in TFP across firms at a point in time (second moment)?

- How do the first & second moments develop over time, say in response to policy reforms or technological innovations?

In any case, TFP is unobserved and so has to be estimated.

But hard to distinguish genuine variation in TFP from noise in the dependent variable. Some possible methods:

- **Data envelopment analysis** (DEA) - no functional form assumptions about the relationship between inputs and output which is nice; however the computed inefficiency scores are very sensitive to measurement errors, either in output or the input variables. Basically, deviations from the frontiers arise only because of inefficiency; random noise doesn't play a role by assumption. Might be fine if you have very good data on a homogenous industry - but not if you're analyzing noisy survey data on firms with different inputs & outputs.

- **Stochastic frontier analysis** - accommodates statistical noise in the dependent variable by means of introducing a residual, while treating inefficiency as a random parameter. Parametric assumptions enable you to distinguish between noise and inefficiency but these assumptions are more or less arbitrary. Another unattractive feature is that the inefficiency term typically is assumed to be uncorrelated with the explanatory variables in the frontier production function. If the inefficiency terms are in fact correlated with firm attributes, the estimated parameters and the inefficiency scores from such models will be biased.'

- If **panel data** are available, and inefficiency is approximately constant over time, then we can model inefficiency as a time invariant firm specific effect. This is the route taken in the paper. No assumptions about the distribution of inefficiency are needed, and the inefficiency can be correlated with the arguments of the production function.

- The production function is thus written as

$$Y_{it} = A_{it}F\left(Z_{it}\right)U_{it}e^{\varepsilon_{it}},$$

where Y_{it} is output, $A_{it}F\left(Z_{it}\right)$ is the output frontier (maximum output attainable with technology A_{it} and input vector Z_{it}), $F\left(.\right)$ denotes the production function, U_{it} is an index of inefficiency (1=fully efficient; less than $1 =$ inefficient), and ε_{it} is time varying unobservables determining recorded output (could be measurement errors in output or inputs; could also be unobserved genuine economic factors such as demand shocks).

3.1.3 Allocative Efficiency

- Allocative inefficiency is defined in the paper as a situation arising when as a result of price differentials firms of differing size select different factor combinations. Notice that allocative inefficiency defined like this is not the result of optimization errors made by the firm. Such differentials may be due, for example, to **non-competitive factor markets** or **differential taxation on firms of differing size**.

- Why this definition of allocative inefficiency? Because we think of the counterfactual as the factor choices that would be made if there were no price differentials. Actual choices are different and there will be an opportunity cost associated with the difference.

3.2 The production function

- Need a functional form that is sufficiently flexible to allow for non-homotheticity. Translog:

$$\ln F_{it} = \sum_j \beta_j \ln X_{jit} + 1/2 \sum_k \sum_m \beta_{km} \ln X_{kit} \ln X_{mit}, \qquad \beta_{rs} = \beta_{sr},$$

  where X_j is the jth input in the production function, $j = 1, 2, ..., J$ and β denotes parameters to be estimated.

- Testable hypotheses:

$$\begin{aligned}
\text{Homotheticity} &: \quad \sum_k \beta_{km} = 0, \qquad m = 1, 2, ...J \\
\text{Constant returns to scale} &: \quad \left\{ \begin{array}{c} \text{Homotheticity} \\ \sum_k \beta_k = 1 \end{array} \right\} \\
\text{Cobb-Douglas} &: \quad \beta_{km} = 0, \quad k = 1, 2, ...J; m = 1, 2, ..., J
\end{aligned}$$

- Empirical specification is augmented with firm-level averages of employees' years of education, tenure, age and age squared. Note similar to the Mincerian earnings function.

- Empirical specification:

$$\ln F_{it} = \sum_j \beta_j \ln X_{jit} + 1/2 \sum_k \sum_m \beta_{km} \ln X_{kit} \ln X_{mit} + \alpha h_{it} + \mu_i + \delta_t + \varepsilon_{it},$$

  where h is the vector of human capital variables, $\mu_i = -\log U_i$ measures inverse inefficiency (think: efficiency), and δ_t is a time effect common to all firms.

- Main estimation problem: inputs (labour, capital, intermediate inputs) are potentially endogenous, because managers may observe error term (reflecting, say, demand shocks while the econometrician doesn't).

- To deal with this we use instrumental variables, using lagged values of the explanatory variables as instruments. We take first differences to wipe out the firm effects. In the case of highly persistent data, lagged variables in levels are likely to be weak instruments for contemporaneous differences. We therefore follow Blundell and Bond and combine the differenced equation with a levels equation to form a system generalised method of moments (GMM) estimator.

- Thus, we use lagged levels as instruments for contemporaneous differences and lagged differences as instruments for contemporaneous levels.

3.3 Data

- Panel data on manufacturing firms in Ghana, collected in face-to-face interviews. 1991 to 1997. Estimation sample consists of 676 observations - thus small sample.

- A sample of workers and apprentices was chosen from each firm. It is therefore possible to use the responses from workers in the firm to create firm-level averages of worker characteristics.

[Table 1: descriptive statistics].

Table 1
Summary statistics

	[1] Micro	[2] Small	[3] Medium	[4] Large	[5] All
Conditional means[a]					
Log output per employee	13.75 (0.12)	13.91 (0.06)	14.08 (0.07)	14.68 (0.10)	14.08 (0.04)
Log value-added per employee	12.48 (0.15)	12.67 (0.07)	12.92 (0.09)	13.68 (0.12)	12.90 (0.05)
Log capital per employee	12.39 (0.19)	12.48 (0.09)	13.73 (0.12)	14.72 (0.16)	13.22 (0.06)
Ratio of value-added to output	0.34 (0.02)	0.36 (0.01)	0.38 (0.01)	0.39 (0.02)	0.37 (0.01)
Average education in years	9.58 (0.27)	9.81 (0.13)	10.28 (0.16)	11.27 (0.22)	10.18 (0.09)
Average age in years	26.52 (0.78)	28.69 (0.38)	33.43 (0.47)	35.55 (0.64)	31.02 (0.26)
Average tenure in years	3.38 (0.48)	4.22 (0.24)	6.90 (0.29)	7.21 (0.39)	5.42 (0.16)
Log average labour cost	11.94 (0.10)	12.25 (0.05)	12.67 (0.06)	13.05 (0.08)	12.48 (0.03)
Unconditional means					
Firm age (years)	15.17	14.45	19.28	22.87	17.39
Any foreign ownership (proportion)	0.14	0.07	0.23	0.59	0.22
Any state ownership (proportion)	0.00	0.02	0.09	0.06	0.05
Industry (proportions)					
Food and bakery	0.32	0.22	0.18	0.25	0.22
Wood	0.00	0.03	0.06	0.29	0.08
Furniture	0.13	0.28	0.27	0.20	0.25
Textiles and garments	0.33	0.16	0.20	0.00	0.16
Metal and machinery	0.22	0.31	0.28	0.25	0.28
Location (proportions)					
Accra (capital city)	0.32	0.60	0.63	0.55	0.57
Cape	0.06	0.02	0.03	0.03	0.03
Kumasi	0.49	0.35	0.28	0.13	0.30
Takoradi	0.14	0.03	0.06	0.29	0.10
Number of observations	72	292	193	119	676
Number of firms	15	61	40	27	143

The size of the firm is its total number of employees when first observed in the sample, where a micro firm has less than 6 employees, a small firm has from 6 to 29, a medium firm has from 30 to 99, while a large firm has 100, or more, employees. The figures in ( ) are standard errors. All monetary variables have been deflated using firm-level price indices.

[a] The numbers reported in Columns [1]–[4] are predictions based on OLS regressions in which the regressors are sector, time and size dummies. The predicted values are calculated using sample means of the sector and time dummies, i.e. of the form $\hat{y}_{\mathrm{size}\_i}=\alpha\cdot\mathrm{size}\_i+\hat{\beta}\cdot\bar{x}$, where size_$i$ indicates the ith size category, $\bar{x}$ is the vector of mean values of the sector and time dummies, and $\hat{\alpha}$ and $\hat{\beta}$ are estimated coefficients. The numbers reported in the fifth column are predictions based on sample means of all regressors, hence they are effectively (unconditional) sample means.

those with from 6 to 30, medium those with from 31 to 99, and large those with 100, or more, employees. The upper panel of the table (under the heading ‘Conditional means’) shows mean values for the monetary and human capital variables, purged of sectoral and time effects as explained in the notes to the table.[11] The fact, shown in Table 1, that the capital–labour ratio differs substantially across firms of differing size when the data are purged of sectoral effects is important for establishing that it is not differences in technology related to

[11] Because the panel is unbalanced, the sample composition is not constant over time. To ensure that the summary statistics are not driven by changes in the sample composition during the course of the surveys, we purge the variables of time effects.

3.4 Results

OLS and fixed effects results are shown in Table 2. Brief summary:

- OLS:
  - No strong evidence of non-homotheticity or variable returns to scale.
  - The Cobb-Douglas specification can easily be accepted given the translog functional form with the value-added specification, but not with the output model.
  - The human capital coefficients all have the anticipated signs, however only the age effect, which is a quadratic, is significant.

- Fixed effects:
  - Unreasonable results - e.g. capital coefficient negative!?
  - Could be due to measurement errors in inputs - the bias gets worse if you difference the data. IV methods could solve/mitigate this problem.

[Show Table 2 here]

Table 2
OLS and within estimates of production function parameters

	Dependent variable: log value-added				Dependent variable: log output			
	Translog		Cobb–Douglas		Translog		Cobb–Douglas	
	[1] OLS	[2] Within	[3] OLS	[4] Within	[5] OLS	[6] Within	[7] OLS	[8] Within
Marginal effects[a]								
Log employment	0.84 (8.79)**	0.30 (1.31)	0.89 (9.74)**	0.34 (1.59)	0.10 (3.33)**	0.03 (0.46)	0.14 (3.78)**	0.08 (1.18)
Log capital	0.20 (3.57)**	– 0.25 (0.65)	0.18 (3.43)**	– 0.22 (0.59)	0.02 (1.68)$^+$	– 0.06 (0.46)	0.03 (2.12)*	– 0.06 (0.53)
Log raw materials					0.72 (37.49)**	0.69 (22.78)**	0.71 (29.12)**	0.65 (17.12)**
Log indirect costs					0.15 (8.07)**	0.11 (4.51)**	0.12 (5.19)**	0.10 (3.47)**
Human capital coefficients								
Education	0.04 (1.63)	0.02 (0.57)	0.04 (1.59)	0.02 (0.53)	0.01 (2.53)*	0.00 (0.35)	0.01 (2.11)*	0.00 (0.35)
Age	0.13 (2.19)*	0.21 (3.08)**	0.13 (2.04)*	0.20 (3.09)**	0.05 (3.04)**	0.07 (2.59)**	0.04 (2.53)*	0.08 (3.01)**
Age^2/100	– 0.20 (2.26)*	– 0.33 (3.46)**	– 0.19 (2.09)*	– 0.33 (3.49)**	– 0.07 (2.96)**	– 0.11 (2.82)**	– 0.06 (2.58)**	– 0.13 (3.33)**
Tenure	0.03 (1.70)$^+$	0.05 (2.46)*	0.03 (1.43)	0.05 (2.56)*	0.004 (1.08)	0.01 (2.55)*	0.01 (1.35)	0.02 (3.47)**
Diagnostics and tests								
R^2	0.74	0.10	0.74	0.09	0.98	0.82	0.97	0.80
Quasi-concavity (proportion)	1.00	0.61			0.44	0.04		
Monotonicity (proportion)	1.00	0.00			0.69	0.04		
Homotheticity[b] (p-value)	0.30	0.64			0.25	0.68		
Homotheticity and CRS[c] (p-value)	0.36	0.21	0.39	0.04	0.31	0.49	0.86	0.06
Cobb–Douglas (p-value)[d]	0.49	0.82			0.00	0.00		

Time dummies are included in all regressions. The OLS regressions include controls for the age of the firm, industry, ownership structure and location. The numbers in ( ) are t-statistics based on standard errors robust to heteroskedasticity. Significance at the 1%, 5% and 10% level is indicated by *, ** and $^+$, respectively.

[a] For the translog specification, the marginal effects is a function of the inputs, and have therefore been evaluated at the sample means. For the Cobb–Douglas specification, the marginal effects are equal to the estimated coefficients.

[b] H$_0$: $\sum_k \beta_{km} = 0, \quad m = 1, 2, \ldots, J$ (see Eq. (3)).

[c] For translog specifications, H$_0$: $\sum_k \beta_{km} = 0, \quad m = 1, 2, \ldots, J$, and $\sum_k \beta_k = 1$. For Cobb–Douglas specifications, H$_0$: $\sum_k \beta_k = 1$ (see Eq. [3]).

[d] H$_0$: $\beta_{km} = 0$, $k = 1,2,\ldots,J$, $m = 1,2,\ldots,J$ (see Eq. (3)).

System GMM results - supposedly robust to endogeneity problems - are shown in Table 3. Summary:

- Cobb-Douglas results for the value-added specification are reported in Column [2]. The labour coefficient is 0.73, and the capital coefficient is 0.31; both are significant at 1% level.

- No evidence of variable returns to scale or non-homotheticity. That is, a simple constant returns to scale Cobb-Douglas function seems to perform as well as more general specifications.

- Among the human capital variables only age has a significant impact on productivity.

Table 3
System GMM estimates of production function parameters

	Dependent variable: log value-added[a]		Dependent variable: log output[b]	
	[1] Translog	[2] Cobb–Douglas	[3] Translog	[4] Cobb–Douglas
Marginal effects[1]				
Log employment	0.88 (4.20)**	0.73 (3.25)**	0.10 (0.93)	0.17 (2.37)*
Log capital	0.25 (2.49)*	0.31 (3.58)**	0.08 (1.61)	0.09 (2.06)*
Log raw materials			0.68 (12.58)**	0.68 (14.02)**
Log indirect costs			0.13 (2.35)*	0.06 (1.14)
Human capital coefficients				
Education	– 0.01 (0.82)	– 0.003 (0.07)	0.01 (0.65)	0.006 (0.35)
Age	0.24 (2.40)*	0.26 (3.11)**	0.04 (1.42)	0.07 (2.61)**
$Age^2/100$	– 0.38 (2.60)**	– 0.41 (3.35)**	– 0.08 (2.07)*	– 0.11 (2.88)**
Tenure	0.04 (0.76)	0.05 (1.25)	0.01 (1.48)	0.02 (1.89)$^+$
Diagnostics and tests				
Quasi-concavity (proportion)	0.52		0.44	
Monotonicity (proportion)	0.87		0.59	
Homotheticity (*p*-value)[2]	0.63		0.46	
Constant returns to scale (*p*-value)[2]	0.74	0.85	0.59	0.92
Cobb–Douglas (*p*-value)[2]	0.66		0.37	
m1 (*p*-value)[3]	0.00	0.00	0.01	0.00
m2 (*p*-value)[4]	1.00	0.93	0.81	0.15
Sargan–Hansen (*p*-value)[5]	0.57	0.79	0.42	0.39

Time dummies are included in all regressions. The numbers in ( ) are *t*-statistics. Significance at the 1%, 5% and 10% level is indicated by *, ** and $^+$, respectively. Hypothesis tests are based on robust, finite sample corrected standard errors (see footnote 13) calculated using the method proposed by Windmeijer (2000).

[a] The instrument set for the differenced equation consists of the levels of employment, physical and human capital, in periods $t-2$ and $t-3$. The instrument set for the levels equation consists of employment, physical and human capital, differenced, in period $t-1$, a constant and year dummies.

[b] The instrument set for the differenced equation consists of the levels of employment, raw material, indirect costs and physical and human capital, in periods $\mathrm{t}-2$ and $\mathrm{t}-3$. The instrument set for the levels equation consists of employment, raw material, indirect costs and physical and human capital, differenced, in period $\mathrm{t}-1$, a constant and year dummies.

[1] For the translog specification, the marginal effects is a function of the inputs, and have therefore been evaluated at the sample means. For the Cobb–Douglas specification, the marginal effects are equal to the estimated coefficients.

[2] See Table 2.

[3] Tests the null hypothesis that the differenced residuals in periods t and $t-1$ are uncorrelated.

[4] Tests the null hypothesis that the differenced residuals in periods t and $t-2$ are uncorrelated.

[5] Tests for the validity of the overidentifying restrictions.

Column [3] in Table 3 reports system GMM estimates of the output translog production function. Like for all other models previously reported, there is no evidence for non-homotheticity or variable returns to scale. Further, in contrast to the OLS and within models reported in Table 2 for the output specification, we can now comfortably accept the Cobb–Douglas specification, reported in Column [4], as a result of using instrumental variable techniques. We consequently focus on the results in Column [4]. The estimated

3.5 Inefficiency

- Based on the output production function in Column [4], Table 3, we predict the firm fixed effects ($\hat{\mu}_i$).

- In Table 4 we regress the fixed effects estimates on time invariant variables, to see if we can detect any systematic variation in efficiency across sectors, ownership structures, locations and other firm characteristics. Some differences across sectors, but neither firm age nor ownership are associated with significant efficiency differentials. The model explains only 16 per cent of the variation in the fixed effects.

- To get a feel for the 'average' level of inefficiency, and the dispersion of inefficiency, we compute an inefficiency index using the formula

$$te_i) \, \mathsf{exp} \left(- \left(\hat{\mu}_{\mathsf{max}} - \hat{\mu}_i \right) \right),$$

where $\hat{\mu}_{\text{max}}$ is the sample maximum of the firm effects. Figure 1 shows the distribution of this efficiency index. What do we learn from this?

[Figure 1 here]

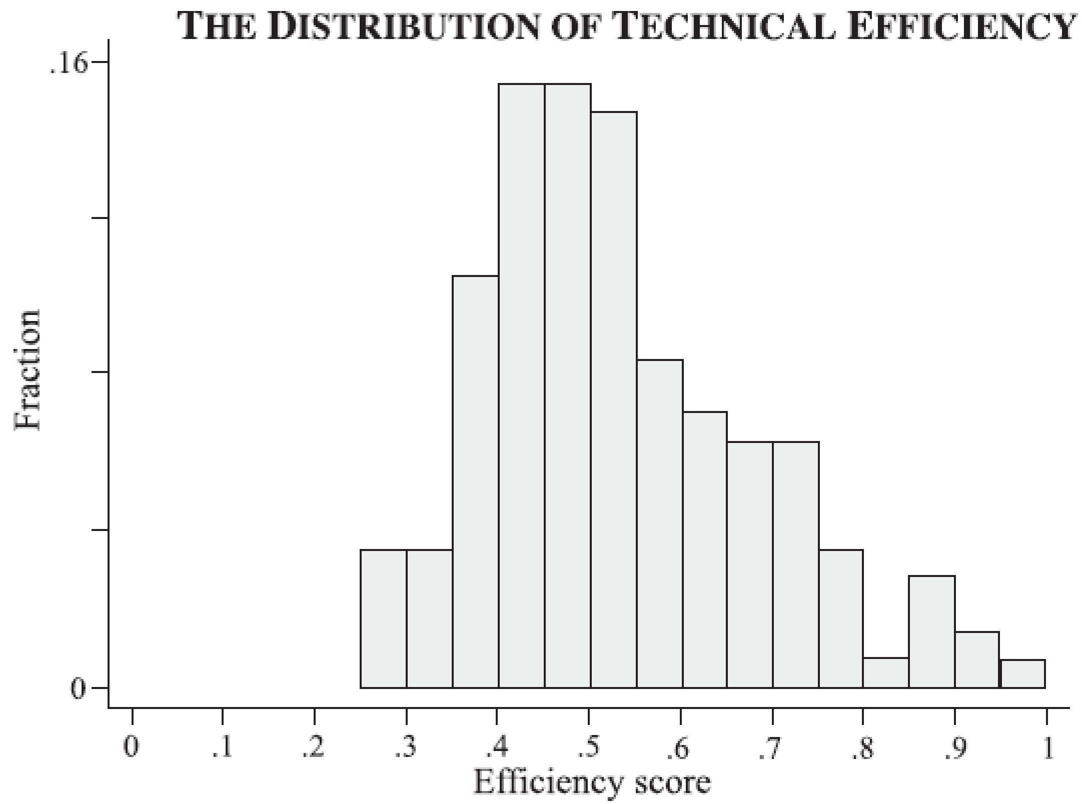


Fig. 1. The distribution of technical efficiency. Note: Technical efficiency is defined as $te_i = e^{-(\hat{\mu}_{\max} - \hat{\mu}_i)}$, where $\hat{\mu}_{\max}$ is the sample maximum of the estimated fixed effects, and $\hat{\mu}_i$ is the estimated fixed effect for firm i. This is interpretable as a measure of the dispersion across firms of productivity, conditional on inputs and human capital. The estimated mean of technical efficiency is 0.53 and the estimated standard deviation is 0.15. The fixed effects estimates are based on the regression in Column [4], Table 3.

3.6 Allocative inefficiency

3.6.1 Capital intensity and firm size

To investigate the relation between capital intensity and firm size, we consider a partial linear model of the form

$$\ln k_{it} = f\left(\ln L_{it}\right) + \text{time \& sector controls} + \text{residual},$$

where k_{it} is a the capital-labour ratio adjusted to take into account differences in labour quality across firms (using the human capital data and the estimated coefficients on these variables).

- To estimate the function $f(.)$ we use a semiparametric approach, which is flexible. Figure 2 shows the estimated function and pointwise 95 per cent confidence bands, obtained through bootstrapping.

- The pattern is non-linear. The positive correlation between size and capital intensity is strongest for firms between ten and 90 employees, outside this range the regression function is relatively flat and the confidence bands wide.

- Within the (10, 90) range, the average slope of the regression line is about 0.8, indicating that a one per cent increase in the labour force is associated with a 0.8 per cent increase in the capital labour ratio.

[Results in Figure 2]

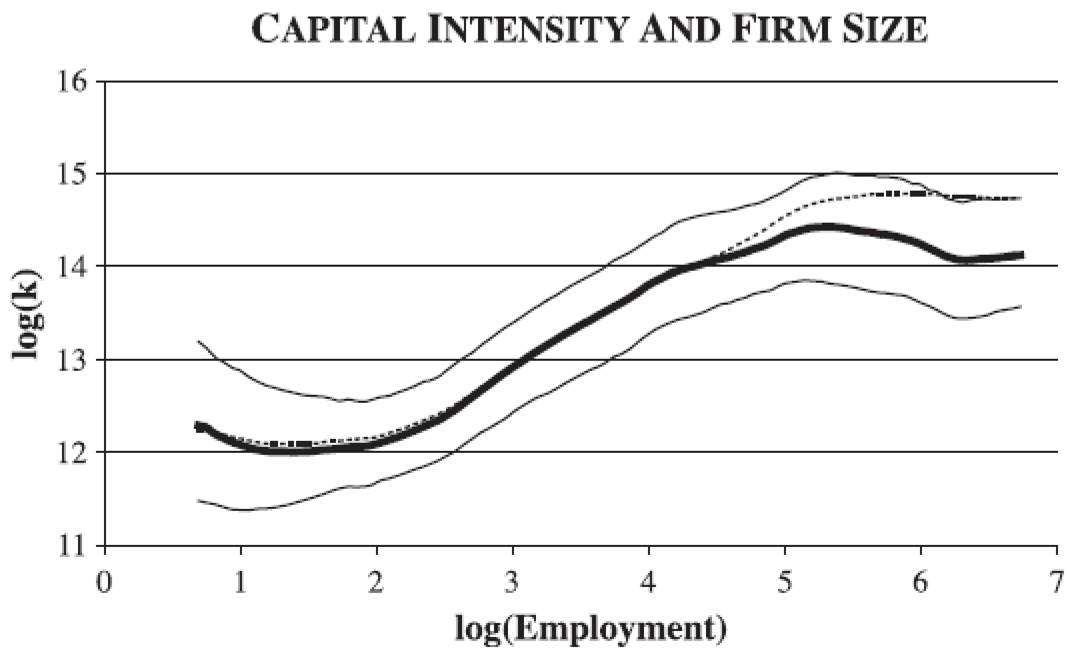
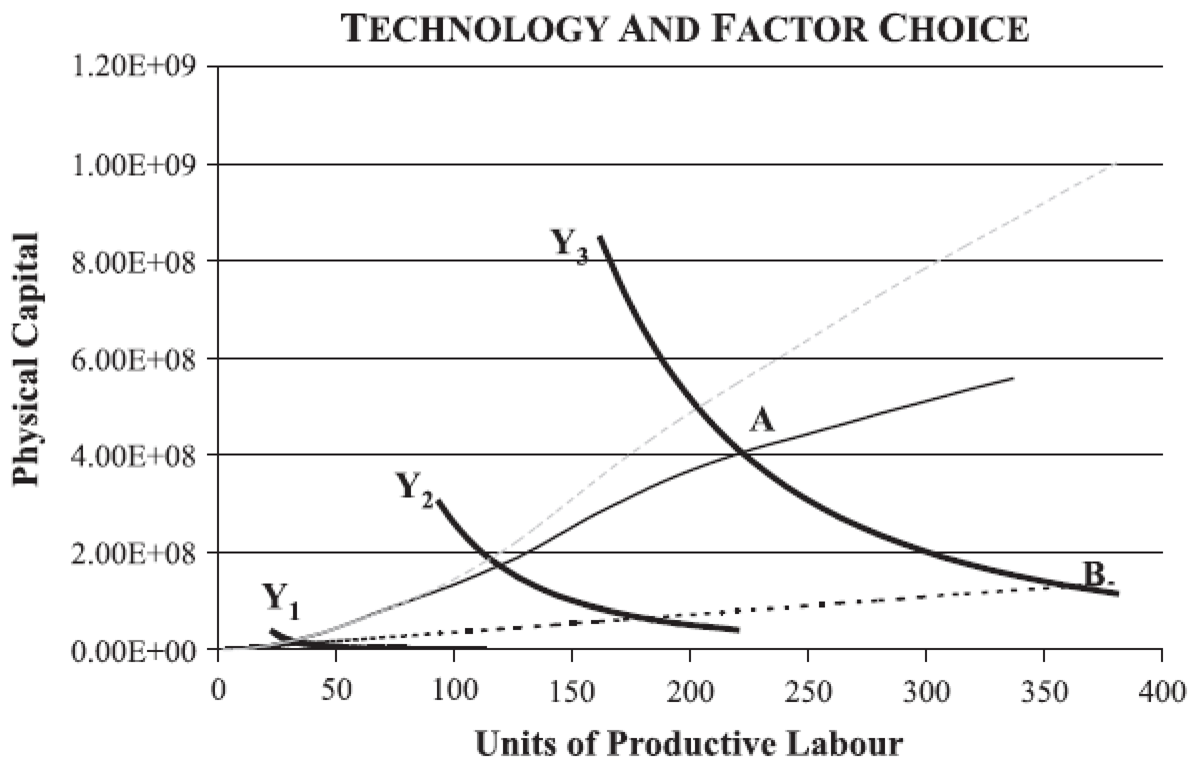


Fig. 2. Capital intensity and firm size. Note: The solid line shows the regression line of ln(k) on ln L. The kernel is Epanechnikov and the bandwidth is equal to 1.20. The thin lines indicate pointwise 95% confidence bands, calculated from 800 bootstrapped replications. To take the panel nature of the data into account, we bootstrapped from the firms rather than from the observations, which is a similar procedure to that used by Deaton (1997, pp. 216–218) for clustered data. The dashed line shows the regression line of ln(k) on ln L when k is not adjusted for worker quality heterogeneity as explained in the text.

- Now use the predicted values of capital intensity, and calculate the implied capital stock values for given levels of labour. This yields the expansion path of factor combinations. We plot this in Figure 3 along with isoquants based on a Cobb-Douglas value-added production function with a capital coefficient of 0.3 and a labour coefficient of 0.7, thus very similar to the regression reported in Column [2], Table 3.

- Recall: no evidence that the technology is non-homothetic - so it is not differences in technology which explain differences in factor choice across firms of differing size.

[Figure 3 here]



Capital-labour ratio at **A**: 1.83 million Cedis.
Capital-labour ratio at **B**: 0.36 million Cedis.

Fig. 3. Technology and factor choice. Note: The underlying production function is $\ln Y = \ln A + 0.3 \ln K + 0.7 \ln L$, where A is a constant. Y_1, Y_2 and Y_3 depict three isoquants under this technology, with $Y_2 = 5Y_1$ and $Y_3 = 10Y_1$. The solid line starting in the origin is the empirical expansion path, derived from the nonparametric regression shown in Fig. 2. The shaded line is the expansion path following from the capital intensity regression not adjusted for labour quality. The dashed line starting in the origin is a hypothetical expansion path for relative factor prices constant at the level observed for a firm with 18 employees.

- The figure implies that firms with 200 employees have up to 5 times higher relative labour to capital costs than firms with 21 employees. These differences in relative factor prices imply a substantial dispersion of factor choices.

- If relative prices differ so much across firms of differing size - is this because of differing labour costs, differing capital costs, or a combination?

- Capital costs are unobserved (basically since capital is purchased, rather than rented).

- But labour costs are observed. So we can run a regression modeling average wage as a function of firm size:

$$\ln w_{it} = \gamma h_{it} + \theta_1 \ln L_{it} + \theta_2 (\ln L_{it})^2 + \tau_t + \omega_i + \upsilon_{it}.$$

- In our preferred specification (which excludes the squared size term since insignificant) we obtain $\hat{\theta}_1 = 0.15$.

- Suggests that firms of differing size face different costs for a given amount of productive labour input. The point estimate of 0.15 implies that labour costs will rise by 40 per cent as firms expand from small to large i.e. from 21 to 200 employees.

- This implies that most of the factor intensity differential documented in the previous sub-section is due to differing capital costs! (Recall: relative factor prices differs by a factor of 5!)

- Infer the cost of capital from the formula

$$r = (0.3/0.7) \times \hat{w} \times \left(\widehat{K/L}\right)^{-1},$$

where we put $\hat{}$ to indicate these quantities are predicted from the models discussed above.

- Figure 4 shows how r varies with firm size. After an initial increase for firms with less than seven employees, the capital price falls sharply with size. Micro firms (firms with at most five employees) face average capital prices between 0.34 and 0.52, while large firms (with more than 99 employees) face prices between 0.09 and 0.14.

- If factor prices vary due to policy or labour market distortions, which are removable, then substantial cost reductions are possible. To illustrate, suppose the large firm output were produced using small firm factor proportions - then the amount of labour would be 64 per cent higher while the capital required would be about a third that actually used. If large firms faced the labour costs of small firms our findings imply that their unit costs would fall by 20-25 per cent.

3.7 Summing up

- A very simple functional form, the Cobb-Douglas, adequately represents the production technology. The implication of the acceptance of the Cobb-Douglas functional form is that differences if factor choices over firms of differing size is not a reflection of differences in technology.

- The measures of human capital appear not to be quantitatively very important in determining productivity.

- Large firms facing higher relative labour costs than smaller firms use a much more capital intensive technology and operate with costs 20-25 per cent higher than those which would occur if factor prices differentials across firms of differing sizes could be eliminated.

4 Agricultural Productivity and Farm Size

Reference: Lamb (2003).

- Puzzle in empirical work on developing country agriculture: the "Inverse Productivity" (IP) relationship. Basic model:

$$\ln Y_i = \alpha + X_i\beta + \gamma \ln A_i + u_i,$$

  where Y is output (or profits), A is total area farmed, X is a vector of "control variables" and u is a residual.

- The puzzle: $\gamma < 1$, i.e. a 1% increase in land area yields a smaller than 1% increase in output (profits).

- Similar relationship is observed for labour demand: labour demand rises less quickly than area farmed.

- Why this finding?

- Possible explanation #1: Small farmers are more efficient than large farmers. If this is true, total output might increase as a result of dividing up large farms into smaller plots.

- But intuition suggests this is not correct - if anything, you would think there are increasing returns to scale amongst farmers.

- Possible explanation #2: Small farmers have land that is of higher quality (on average) than that used by large farmers. If this is true, land redistribution from large to small landholders will not raise agricultural output and rural incomes.

- Possible explanation #3: Market failures. For example, small-scale farmers can't optimally divide their time between self-employment at the farm and wage employment. So they end up working at their own farm all the time (over–allocate labour), driving the marginal production of own-farm labour below the market wage rate. As a result, small farms get a lot of owner input.

- Common theme in explanations 2-3: The IP result is not genuine. Instead, it is caused by a specification error - there is an **omitted variable** in the model.

- Possible explanation #4: Poor data. Measurement errors in area.

- Purpose of this paper is to test competing explanations of the IP relationship using data on Indian farms.

[Summary statistics in Table 1]

Survey panel data on rural households in India, 1975-1985. Sample consists of 1,060 households drawn from 8 villages. Sample means in Table 1:

Table 1
Sample means

Log total cropped area	2.18 acres
Log real household profits	7.51 rupees
Log hours of male labor	6.50 h
Log hours of female labor	6.54 h
Average value of land	2607 rupees/acre
Share of irrigated land	0.16%
Share, type 1 land	0.09%
Share, type 2 land	0.40%
Share, type 3 land	0.21%
Share, type 5 land	0.19%
Share, type 6 land	0.02%
Real wage, male, period 1	0.80 rupees/h
Real wage, male, period 2	0.83 rupees/h
Real wage, female, period 1	0.46 rupees/h
Real wage, female, period 2	0.51 rupees/h
Real fertilizer price	2.83 rupees/kg
Real price of sorghum	1.70 rupees/kg
Real price of fodder	27.3 rupees/quintal

1 US $ ~= 9 Indian Rupees 1975-85.
1 acre ~= 4,000 square meters.

- Small plots.
- Low profits.
- Low wages.

- Econometric model:

$$\ln Y_{it} = \alpha_i + X_{it}\beta + \gamma \ln A_{it} + u_{it}.$$

  Notice i, t subscripts on explanatory variables (reflecting panel data). Also notice that the intercept varies across households but is constant over time (you see this because α has an i-subscript but no t-subscript). The author will thus use a **fixed effects** estimator, which you will recall is analogous to including separate dummy variables for each households in the model.

- The author will also consider results from a **random effects** estimator. I will focus mostly on the fixed effects results.

- [Regression results in Table 2]

Table 2
The inverse productivity relationship in ICRISAT data[a] dependent variable: log of household profits

	(1) Random effects	(2) Fixed effects	(3) Random effects[b]	(4) Fixed effects[b]
Log total cropped area[c]	0.89***	0.62***	0.97	0.71***
	(2.57)	(4.53)	(0.72)	(3.48)
Monsoon onset	− 0.01***	− 0.01**	− 0.01***	− 0.01*
	(− 2.95)	(− 2.05)	(− 3.96)	(− 1.83)
Monsoon end	− 0.001	− 0.000	0.00	0.00
	(− 0.67)	(− 0.12)	(0.17)	(0.45)
Frequency of days with rain	− 0.79	− 0.87	− 0.81	− 0.61
	(− 1.59)	(− 1.54)	(− 1.62)	(− 1.06)
Total rainfall	0.000	0.000	0.00	− 0.00
	(1.34)	(0.26)	(1.45)	(− 0.48)
Real fertilizer price	0.22**	0.24	0.32***	0.35**
	(2.24)	(1.44)	(3.05)	(2.02)
Real sorghum price	− 0.04	− 0.009	0.05	0.00
	(− 0.30)	(− 0.04)	(0.04)	(0.01)
Real fodder price	− 0.004	− 0.01**	− 0.01***	− 0.02***
	(− 1.04)	(− 2.00)	(− 3.34)	(− 3.05)
Real wage, male, period 1	− 0.82*	− 0.69	− 0.76	− 1.25
	(− 1.65)	(− 0.92)	(− 1.54)	(− 1.59)
Real wage, male, period 2	1.33***	0.67	1.52***	0.40
	(2.65)	(0.74)	(2.72)	(0.64)
Real wage, female, period 1	0.55	0.72	0.59	1.77*
	(0.98)	(0.78)	(1.07)	(1.75)
Real wage, female, period 2	− 0.06	0.91	− 1.33*	0.45
	(− 0.10)	(1.03)	(− 1.77)	(0.63)
Share of irrigated land	***	***	1.14***	0.71**
			(6.61)	(2.31)
Average value of cropland	***	***	19.43***	18.22***
			(5.78)	(3.54)
Test for joint significance of wages and prices (*p*-value)	4.82	5.28	39.11	5.49
	(0.00)	(0.00)	(0.00)	(0.00)
Test for joint significance of rainfall (*p*-value)	3.34	3.50	16.99	3.99
	(0.01)	(0.01)	(0.00)	(0.00)
Test for joint significance of land quality (*p*-value)			1.92	3.91
			(0.00)	(0.00)

- Col. (2): Coef. on log Area less than 1
- Significantly different from 1. How is this test done?
- Thus, results suggest IP.
- Now probe the data further to see if any of the alternative explanations might be true (land quality, market imperfections, measurement errors)

Table 2

The inverse productivity relationship in ICRISAT data[a] dependent variable: log of household profits

	(1) Random effects	(2) Fixed effects	(3) Random effects[b]	(4) Fixed effects[b]
Log total cropped area[c]	0.89***	0.62***	0.97	0.71***
	(2.57)	(4.53)	(0.72)	(3.48)
Monsoon onset	− 0.01***	− 0.01**	− 0.01***	− 0.01*
	(− 2.95)	(− 2.05)	(− 3.96)	(− 1.83)
Monsoon end	− 0.001	− 0.000	0.00	0.00
	(− 0.67)	(− 0.12)	(0.17)	(0.45)
Frequency of days with rain	− 0.79	− 0.87	− 0.81	− 0.61
	(− 1.59)	(− 1.54)	(− 1.62)	(− 1.06)
Total rainfall	0.000	0.000	0.00	− 0.00
	(1.34)	(0.26)	(1.45)	(− 0.48)
Real fertilizer price	0.22**	0.24	0.32***	0.35**
	(2.24)	(1.44)	(3.05)	(2.02)
Real sorghum price	− 0.04	− 0.009	0.05	0.00
	(− 0.30)	(− 0.04)	(0.04)	(0.01)
Real fodder price	− 0.004	− 0.01**	− 0.01***	− 0.02***
	(− 1.04)	(− 2.00)	(− 3.34)	(− 3.05)
Real wage, male, period 1	− 0.82*	− 0.69	− 0.76	− 1.25
	(− 1.65)	(− 0.92)	(− 1.54)	(− 1.59)
Real wage, male, period 2	1.33***	0.67	1.52***	0.40
	(2.65)	(0.74)	(2.72)	(0.64)
Real wage, female, period 1	0.55	0.72	0.59	1.77*
	(0.98)	(0.78)	(1.07)	(1.75)
Real wage, female, period 2	− 0.06	0.91	− 1.33*	0.45
	(− 0.10)	(1.03)	(− 1.77)	(0.63)
Share of irrigated land	***	***	1.14***	0.71**
			(6.61)	(2.31)
Average value of cropland	***	***	19.43***	18.22***
			(5.78)	(3.54)
Test for joint significance of wages and prices (*p*-value)	4.82	5.28	39.11	5.49
	(0.00)	(0.00)	(0.00)	(0.00)
Test for joint significance of rainfall (*p*-value)	3.34	3.50	16.99	3.99
	(0.01)	(0.01)	(0.00)	(0.00)
Test for joint significance of land quality (*p*-value)			1.92	3.91
			(0.00)	(0.00)

- Col. (4): Controlling for **land quality** (e.g. soil quality – not shown in table but included in regression), **increases** the coefficient on area.
- Indicates that small farms use higher-quality land, perhaps because high-quality land is subdivided more often.
- Why this conclusion? Go back to formula for omitted variable bias in the OLS estimator.
- But coefficient is still less than 1, thus IP still seems to hold.

Table 3
Labor demand and the inverse relationship[a] (household fixed effects, White standard errors) dependent variable: log total hours, by gender

	Fixed effects		Random effects	
	(1) Male	(2) Female	(3) Male	(4) Female
Log total cropped area[b]	0.81***	0.80***	0.92***	0.86***
	(−4.00)	(−3.68)	(−4.08)	(−5.03)
Monsoon onset	0.01***	−0.00	0.01***	0.00
	(5.31)	(−1.19)	(7.11)	(0.25)
Monsoon end	0.00	−0.00	0.00	−0.00*
	(0.91)	(−0.63)	(0.97)	(−1.63)
Frequency of days with rainfall	1.03***	0.44	0.81***	−0.06
	(5.09)	(1.75)	(3.84)	(−0.22)
Total rainfall	−0.06***	0.02	−0.00***	0.03*
	(−4.68)	(1.00)	(−5.85)	(1.64)
Real price of fertilizer	−0.09	−0.01	−0.03	0.23***
	(−1.48)	(−0.20)	(−0.60)	(3.53)
Real price of sorghum	0.30***	0.26***	0.26***	0.02
	(4.06)	(2.67)	(4.33)	(0.28)
Real price of fodder	−0.00	−0.00	−0.01***	−0.01***
	(−0.16)	(−1.19)	(−4.37)	(−3.69)
Real wage, male, period 1	−0.17	0.05	−0.58***	−1.10***
	(−0.64)	(0.14)	(−2.69)	(−3.63)
Real wage, male, period 2	−0.32	0.16	−0.71***	0.62*
	(−1.10)	(0.41)	(−2.94)	(1.86)
Real wage, female, period 1	0.94***	0.25	2.03***	1.88***
	(2.62)	(0.51)	(8.00)	(5.19)
Real wage, female, period 2	−0.04	−0.41	0.14	−1.13**
	(−0.12)	(−0.81)	(0.42)	(−2.51)
Share of irrigated land	1.12***	1.09***	1.38***	1.34***
	(8.14)	(6.59)	(17.72)	(12.10)
Average value of cropland	0.89	5.58**	5.39***	9.23***
	(0.46)	(2.19)	(3.53)	(4.27)
Test for joint significance of rainfall (p-value)	11.47	3.98	68.79	7.27
	(0.00)	(0.00)	(0.00)	(0.12)
Test for joint significance of prices/wages (p-value)	4.00	2.04	99.00	59.11
	(0.00)	(0.04)	(0.00)	(0.00)
Test for joint significance of land quality (p-value)	10.54	12.61	564.4	321.7
	(0.00)	(0.00)	(0.00)	(0.00)

- Labour demand: dependent variable is now log total hours worked.
- Cols. (1)-(2): Coefficient on log area less than 1 & signficiantly different from 1.
- Suggests a 1% increase in area leads to less than 1% increase in labour input.

4.1 The inverse relationship and imperfect markets

- Taking stock: land quality explain to some extent - but not fully - the IP relationship in profits (since coefficient on log area increases as a result of controlling for land quality - make sure you understand the logic here). Also, clear IP relationship in labour demand.

- Lamb now investigates if there is any evidence that small farmers **over-allocate** labour to their own farms. This would not be optimal - marginal product of labour is lower than the going wage rate - but would result in a lot of output at the farm, hence high productivity. To test for this, he adds to the model variables measuring **labour market characteristics** (e.g. unemployment rates). The idea is that if unemployment is high, small farmers will over-allocate labour to their own farms, and get high average productivity as a result.

- Hence we expect the coefficient on log area to **increase** as a result of controlling for unemployment.

- However, this generalization of the model does **not** change the main result, i.e. it is still the case that the coefficient on log area is less than 1 and significantly different from 1 (in the fixed effects specifications).

4.2 Measurement errors

Suppose that log area is measured with error, so that

$$\ln A_{it}^{obs} = \ln A_{it}^{true} + \eta_{it},$$

where $\ln A_{it}^{obs}$ is observed (measured) area, $\ln A_{it}^{true}$ is true, unobserved area, and η_{it} is a measurement error (assumed uncorrelated with true area) The correct specification is clearly

$$\ln Y_{it} = \alpha_i + X_{it}\beta + \gamma \ln A_{it}^{true} + u_{it},$$

but unfortunately this cannot be estimated as $\ln A_{it}^{true}$ does not exist in the data.

If we estimate

$$\ln Y_{it} = \alpha_i + X_{it}\beta + \gamma \ln A_{it}^{obs} + e_{it},$$

it must be that the residual in this equation contains the measurement error:

$$e_{it} = u_{it} - \gamma \eta_{it}.$$

This will lead to **downward bias** in the estimate of γ as ln A_{it}^{obs} will be negatively correlated with $e_{it} = u_{it} - \gamma \eta_{it}$.

- Intuition: The correlation is **negative**, since a **high** value of η_{it} simultaneously leads to a **low** value of e_{it} and a **high** value of ln A_{it}^{obs}. The bias will be more severe the more important are the measurement errors, and in the extreme case where measurement errors are completely dominating the data the estimated coefficient γ goes to zero, regardless of the true value of the parameter.

- The intuition is straightforward: you can't explain anything (i.e. you will get a very low or zero coefficient) if your explanatory variable is basically garbage (contains little true information).

- To investigate if this is indeed what is going on in the data, the author uses an **instrumental variable** approach. The proposed instruments now need to be:

  - Uncorrelated with the residual e_{it}, i.e. uncorrelated with the measurement error η_{it} (in which case instruments are **valid**).

  - Correlated with log area (in which the instruments are **informative**, or **relevant**).

- Author uses as instruments dummy variables for **sharecropping** and **double cropping** by the household. The idea is that such activities should be associated with larger farms, and should not be related to the measurement error in farm size.

[Results in Table 5]

Estimated coefficient now equal to 1.00. No IP!

Table 5

Instrumental variables estimates of profit and labor demand equations[a] household fixed effects

	(1) Log total area	(2) Profits	(3) Male	(4) Fem
Log total cropped area[b]		1.00	0.83*	1.12
		(−0.01)	(−1.77)	(0.86)
Monsoon onset	−0.01	−0.02	−0.01*	−0.02*
	(−1.32)	(−1.06)	(−1.85)	(−1.82)
Monsoon end	−0.00	0.00	−0.00	0.00
	(−1.30)	(0.13)	(−0.67)	(1.53)
Frequency of days with rainfall	−0.60	9.92***	1.51*	2.70**
	(−0.65)	(4.33)	(1.73)	(2.13)
Total rainfall	0.00*	−0.00*	0.00**	0.00*
	(1.82)	(−1.71)	(2.16)	(1.88)
Real price of fertilizer	0.04*	−1.24**	0.50**	0.50
	(0.18)	(−2.00)	(2.10)	(1.46)
Real price of sorghum	−0.27	3.59***	−0.61**	0.03
	(−0.88)	(4.60)	(−2.07)	(0.06)
Real price of fodder	0.01	0.06	0.04**	0.06**
	(0.40)	(1.37)	(2.38)	(2.59)
Real wage, male, period 1	−0.47	−2.22	−0.48	0.87
	(−0.55)	(−1.05)	(−0.60)	(0.74)
Real wage, male, period 2	1.00*	1.45*	2.02***	1.03
	(1.62)	(0.92)	(3.37)	(1.18)
Real wage, female, period 1	1.04	5.29***	2.68***	−0.16
	(1.32)	(2.69)	(3.57)	(−0.14)
Real wage, female, period 2	−1.18	−6.24	−4.97***	−5.46**
	(−0.66)	(−1.39)	(−2.90)	(−2.20)
Unemployment rate, male, period 1	−0.18*	−16.10***	−2.30	−5.81**
	(−0.09)	(−3.21)	(1.20)	(−2.09)
Unemployment rate, male, period 2	2.18*	3.74	4.72***	2.80
	(1.74)	(1.17)	(3.87)	(1.58)
Unemployment rate, female, period 1	0.28	3.63*	0.02	−0.73
	(0.35)	(1.82)	(0.03)	(−0.66)
Unemployment rate, female, period 2	−2.01**	−3.66	−1.29	1.78
	(−2.10)	(−1.52)	(−1.40)	(1.33)
Village share of land sharecropped/leased in	0.42	−22.67***	−1.14	−4.99
	(0.14)	(−2.94)	(−0.39)	(−1.17)
Share of irrigated land	−0.70***	0.53	1.14***	1.53
	(−5.53)	(1.54)	(8.29)	(7.64)
Average value of cropland	−4.13	11.42*	2.38	0.78
	(−1.54)	(1.67)	(0.91)	(0.21)
Dummy for household sharecropping in	0.48***			
	(9.09)			
Dummy for double-cropping	0.25***			
	(3.88)			
Test for joint significance of rainfall (*p*-value)	2.30	24.97	8.93	9.22
	(0.06)	(0.00)	(0.06)	(0.06)
Test for joint significance of prices/wages (*p*-value)	4.07	38.04	27.44	13.06
	(0.00)	(0.00)	(0.00)	(0.07)

Instruments informative.

Table 5 (*continued*)

	(1) Log total area	(2) Profits	(3) Male	(4) Female
Test for joint significance unemployment rates and village sharecropping (*p*-value)	4.44	13.83	20.97	22.83
	(0.00)	(0.02)	(0.00)	(0.00)
Test for joint significance of land quality (*p*-value)	14.75	9.26	80.83	66.29
	(0.00)	(0.23)	(0.00)	(0.00)
P-value, test for over-identifying restrictions		0.94	0.65	0.05*
Hausman test for difference between IV and FE estimate		1.54	***	2.07
		(0.12)		(0.05)

[a] Includes variables for the share of different soil types not reported in tables; values in parentheses are *t*-statistics. Standard errors corrected for heteroskedasticity using Huber/White correction in fixed effects estimates.

[b] Reported *t*-statistics is for a test of the null hypothesis that $\gamma=1$.

[*] Significant at the 10% level.

[**] Significant at the 5% level.

[***] Significant at the 1% level.

4.3 Conclusions

- If you do not take into account land quality differences across farms of differing size, and measurement errors in farm size, you will obtain a spurious result suggesting that productivity falls with farm size.

- Once you take these mechanisms into account, it is clear that productivity does **not** vary with farm size.

- Farm policies concerned with aggregate productivity: no reason to encourage the formation of small farms.

5 Productivity Growth Across Sectors: Agriculture vs. Manufacturing

In most poor countries, the agricultural sector is dominant in terms of employment and output. The manufacturing sector, although typically much smaller than the agricultural sector, is often perceived to be 'special':

- Leading edge of "modernization" (creates skilled jobs & technological spillover effects). Agriculture, in contrast, is often considered "stagnant" & "traditional" (e.g. Lewis model - surplus labour due to low marginal product of labour)

- Historically, we know that manufacturing exports has been a key factor in the rapid development of the Asian 'tigers'

Because manufacturing is "modern" and agriculture "traditional & stagnant", productivity growth is often assumed to be higher in the manufacturing sector. Policies often favour the manufacturing sector.

The paper by Martin and Mitra (2001) investigates empirically how the growth in total factor productivity (TFP) - or just productivity - differs between manufacturing and agriculture. This analysis sheds light on whether the assumption that manufacturing is the 'engine of growth' is actually correct.

5.1 Data

- Basic production function approach.

- **Panel data** on output and inputs in the manufacturing and agricultural sectors over the 1967-92 period. That is, **sector-level** data (as distinct from firm-level or farm-level data).

- Manufacturing: 38 countries (23 developing countries); Agriculture: 49 countries (32 developing countries).

5.2 Estimating productivity growth rates

The authors use two specifications of the production function.

- The **first** specification is Cobb-Douglas:

$$\ln Y_t = \ln A_0 + r \times t + \alpha \ln L_t + \beta \ln K_t + \varepsilon_t,$$

where Y is value-added, L is labour, K is physical capital, t denotes time, A_0 is initial productivity (TFP) level, ε_t is a residual (omitted in the authors' exposition), and r, α, β are parameters to be estimated. Productivity at time $t > 0$ is equal to $\ln A_0 + r \times t$, hence it follows that r is a measure of the annual growth rate in productivity:

$$\begin{aligned} \ln A_{t+1} - \ln A_t &= \ln A_0 + r \times (t+1) - [\ln A_0 + r \times t] \\ \ln A_{t+1} - \ln A_t &= r. \end{aligned}$$

The parameters $\ln A_0$ and r differ across countries, but α, β are assumed to be constant across countries. The authors also assume constant returns to scale - i.e. $\alpha + \beta = 1$.

- The **second** specification is a generalization of the Cobb-Douglas model - the Translog model:

$$\begin{aligned}
\ln Y_t &= \ln A_0 + r \times t + \alpha \ln L_t + \beta \ln K_t \\
&\quad + \gamma \frac{(\ln L_t)^2}{2} + \delta \frac{(\ln K_t)^2}{2} + \mu (\ln L \times \ln K) + \varepsilon_t.
\end{aligned}$$

The interpretation of the coefficient r is the same as for the Cobb-Douglas model - i.e. as a measure of the percentage growth in productivity per year.

- Notice that the Cobb-Douglas model is a special case of the Translog model, which results if $\gamma = \delta = \mu = 0$.

- Model is estimated with OLS, with country dummies added (hence country "fixed effects" model).

- For manufacturing, the models are as described above. For agriculture, land is added as a factor - so for agriculture the Cobb-Douglas model looks like this:

$$\ln Y_t = \ln A_0 + r \times t + \alpha \ln L_t + \beta \ln K_t + \psi \ln M_t + \varepsilon_t,$$

  where M_t is total arable land plus land under permanent crops.

- Results are shown in Table 1 (developing countries) and Table 2 (developed countries).

[Table 1 & 2 here]

TABLE 1

TFP GROWTH (in % per Year) IN MANUFACTURING AND AGRICULTURE IN LOW- AND MIDDLE-INCOME COUNTRIES

DEVELOPING COUNTRIES	TL-CRS Man	TL-CRS Agri	CD-CRS Man	CD-CRS Agri	SHARES Man	SHARES Agri
Low-income countries:						
Egypt	4.15	1.86	2.63	1.23	3.68	1.47
Honduras	...	1.28	...	1.60	...	2.51
India	−.20	1.90	−.33	1.52	−.39	2.29
Kenya	1.50	2.36	.68	1.69	.61	2.22
Sri Lanka	−2.00	2.38	−2.00	1.94	−.11	1.77
Madagascar	...	−.06	...	−.18	...	1.20
Malawi	...	.68	...	.30	...	.73
Pakistan	2.33	2.30	1.40	1.7	1.85	3.05
Tanzania	...	5.67	...	5.22	...	6.19
Zimbabwe	−.20	−.36	−1.04	−.67	−1.49	−1.51
Low-income average	.93	1.80	.22	1.44	.69	1.99
Middle-income countries:						
Chile	2.36	2.73	2.83	2.70	3.38	3.45
Colombia	1.83	2.97	1.52	2.90	2.88	4.20
Costa Rica	1.21	−2.82	.54	−3.03	2.90	−1.55
Czechoslovakia	...	.84	...	.99	...	.98
Dominican Republic	...	2.49	...	2.89	...	3.61
Greece	1.17	2.91	−.59	3.1	1.34	3.49
Indonesia	3.79	2.94	5.63	2.74	5.12	4.3
Iran	1.77	...	5.11	...	6.75	...
Jamaica	−1.39	.93	−1.4	.48	−1.36	1.65
Korea	5.87	2.89	4.01	3.42	4.92	3.18
Malta	...	5.67	...	5.89	...	4.73
Mauritius	−1.14	−.12	−1.98	.06	−2.52	.58
Morocco	...	1.02	...	1.31	...	2.81
Peru	1.87	1.99	2.42	2.06	.53	3.34
Philippines	1.80	1.64	.95	1.57	1.22	2.07
Poland	...	1.30	...	1.35	...	2.84
El Salvador	−2.70	1.43	−2.10	1.05	−1.44	3.55
Trinidad and Tobago	−1.35	−2.92	−2.34	−2.52	−.70	−.74
Tunisia	.51	2.93	.33	2.99	1.56	5.23
Turkey	2.65	3.37	2.06	3.46	1.82	4.97
Uruguay	...	1.58	...	2.10	...	4.73
Venezuela	−3.83	2.61	−4.03	2.95	−10.39	3.83
South Africa	1.28	2.76	−.10	3.40	.43	2.75
Middle income average	.92	1.78	.76	1.9	.97	2.91
Developing country average	.92	1.79	.62	1.76	.9	2.62
Overall average	1.86	2.34	1.13	2.31	1.65	2.91
Intersection sample average	1.74	2.29	.93	2.26	1.41	2.69

- Higher growth in agriculture:
  - Man: 1.1 – 1.9%
  - Agri: 2.3 – 2.9%

- Slower growth in developing countries
  - Man: 0.6 - 0.9%
  - Agri: 1.8 – 2.6%

- Very low growth in low-income countries.
- Results from t-tests indicate we can **reject** H0: productivity growth is the same in both sectors.
- Results for developed countries shown in Table 2 (you may skip).

5.3 Testing for convergence in productivity levels & growth rates

Define the productivity gap between the U.S. and country i as follows:

$$D_{it} = \ln A_{Rt} - \ln A_{it},$$

where $\ln A_{Rt}$ is the productivity level in the U.S (subscript R for "Reference country"). and $\ln A_{it}$ is the productivity level in country i (productivity levels are estimated separately for manufacturing & agriculture, of course). Productivity is calculated from the Translog regressions as

$$\ln A_{it} = \ln A_0 + r \times t + \varepsilon_t,$$

(the expression in their eq. 4 looks different but is equivalent).

Productivity in country i is considered to be **converging** towards a constant (long-run) gap D_i if $\rho < 1$ in the following equation:

$$D_{it} = a_i + \rho D_{i,t-1} + \epsilon_{it},$$

where a_i is a country specific intercept (captured by country dummies). You should verify that, if $\rho < 1$, it must be that the long-run expected value of the productivity gap is equal to $a_i / (1 - \rho)$.

An important implication of convergence in this context is that all countries are expected to have *the same steady-state growth rates.* Why is this true? Why is this interesting?

Estimation results are shown in Table 5

[Table 5 here]

TABLE 5

TESTS OF CONVERGENCE

	Sector	
	Manufacturing	Agriculture
Test 1:		
T (time trend)	.009	−.005
t ratio	8.128	−7.970
$\hat{\rho}$	.691	.819
$t(\rho = 1)$	−15.550	−10.588
$\bar{\rho}$ (adjusted)	.76	.90
Test 2:		
$\hat{\rho}$ with time dummy	.667	.864
$t(\rho = 1)$	−15.857	−9.786
ρ_{FGLS}	.687	.862
$t(\rho = 1)$	−15.095	−9.857
Individual country-level convergence tests:		
Mean $\hat{\rho}$	.807	.678
SD	.164	.219
Total number of countries (excluding United States)	37	48
Number of countries rejecting no convergence null hypothesis	6	28
Number of countries with higher TFP growth than United States	8	30

NOTE.—FGLS = Feasible generalized least squares.

- The estimate of rho is <1 & significantly different from 1 for both sectors (if rho = 1, then no convergence).

- Implication: in the steady state, all countries will have the same productivity growth rates – compare with macro lecture.

- What about productivity **levels**? Results ”T (time trend)”indicate that the productivity gap is **increasing** over time in manufacturing, and **decreasing** over time in agriculture.

5.4 Conclusions

- Findings **weaken** the case for policies that discriminate against the agricultural sector, in favour of the supposedly more dynamic manufacturing sector.

- Having a large agricultural sector may even be an advantage in terms of growth performance.

6 Firm Growth

Reference: Sleuwaegen and Goedhuys (2002).

- A better understanding of the relationships between growth and certain firm characteristics is important, since it can offer guidance as to what type of firms are likely to be relatively successful and good at creating jobs in the future. Identification of such firms would clearly be informative to policy makers.

- The relationship between firm size and growth is of particular interest in the poor countries, since most firms in such countries are very small. How realistic is it to hope that some of these firms will grow and become successful large firms in the future?

- The relation between firm age and growth is also important. For example, if young firms grow quickly, policy measures aimed at encouraging entry may have significant growth effects in the short and medium term.

- Popular framework: the Jovanovic (1982) learning model, in which firm growth depends on firm age and firm size. In the model individual managers are initially uncertain about their own abilities, but can assess their capability by observing how they perform. Over time efficient firms grow and survive, while inefficient firms decline and fail. This process generates a correlation between survival rates and firm size and age.

- In reality the growth process is complex. Sleuwaegen and Goedhuys (2002) consider the role played by terms non-linear in size and age in the growth regressions:

$$\begin{aligned}\text{Growth} &= a_0 + a_1 \log \text{Size} + a_2 [\log \text{Size}]^2 + a_3 \log Age \\ &\quad + a_4 [\log Age]^2 + a_5 \log \text{Size} \times \log Age + \text{controls}\end{aligned}$$

- Strong evidence of a positive interaction effect between age and size on growth: the relationship between age and growth is less negative (or more positive) for large firms than for small firms, and that the relationship between size and growth is less negative (or more positive) for old firms than for young ones.

- To facilitate interpretation of the results, see Figure 1, which shows predicted size as a function of initial size and age, based on the estimates reported in Sleuwaegen and Goedhuys (2002), Table 3, column 1.

[Table 3 + graph here]

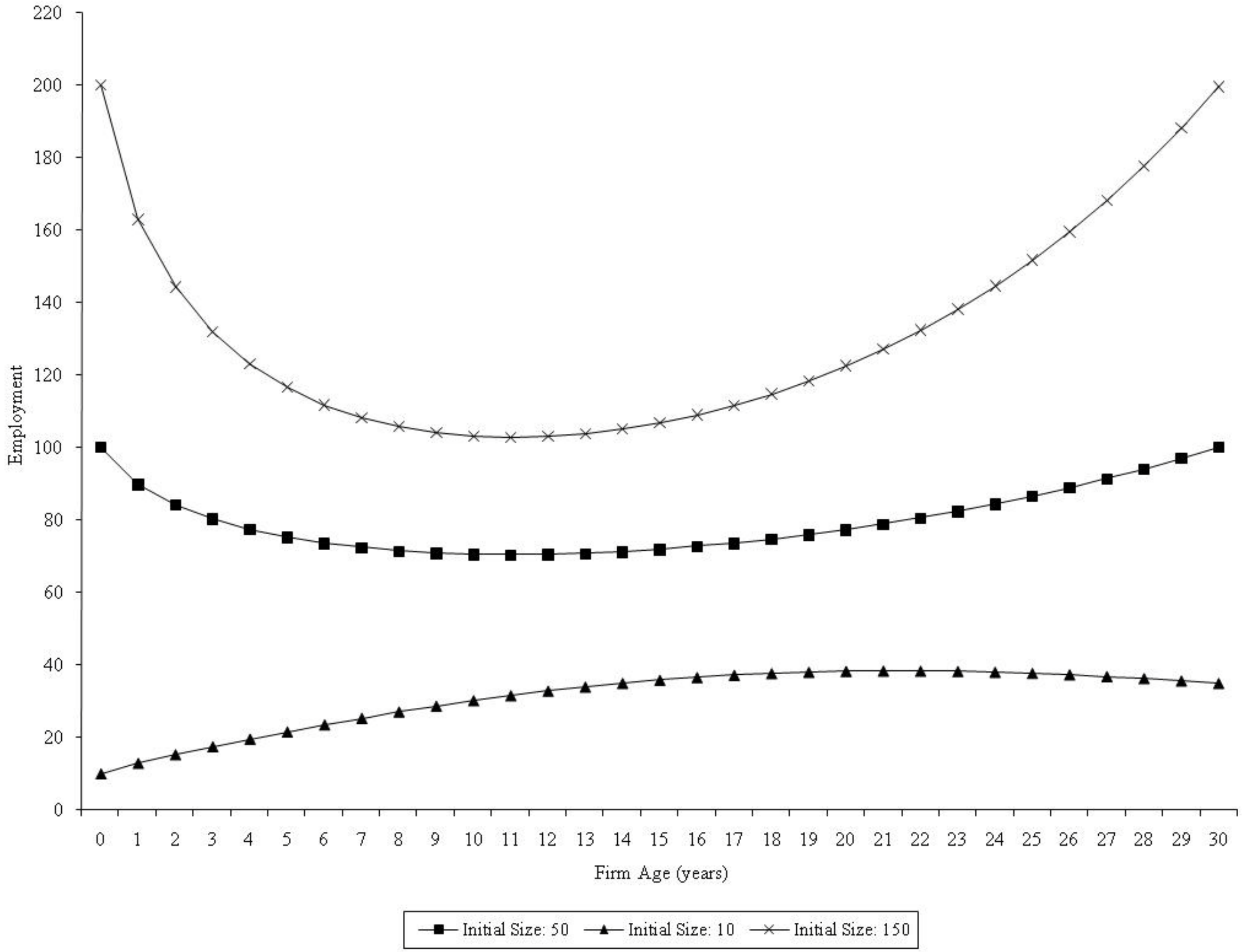
Table 3
Regression results for employment and sales growth over the period startup-1994 and 1989–1994

	Employment growth		Sales growth	
	Startup-1994	1989–1994	Startup-1994	1989–1994
AGE	−0.157***	−0.146***	−0.116*	−0.116**
	(0.050)	(0.050)	(0.062)	(0.049)
SIZE	−0.199*	−0.185**	−0.226***	−0.233***
	(0.107)	(0.073)	(0.057)	(0.053)
SIZE^2	0.006	0.007	0.004	0.009**
	(0.006)	(0.006)	(0.004)	(0.004)
AGE∗SIZE	0.041**	0.031***	0.038***	0.027***
	(0.019)	(0.010)	(0.012)	(0.010)
CAPITAL	0.183*	0.185**	0.020**	0.020**
INTENSITY	(0.108)	(0.081)	(0.009)	(0.009)
FORMAL	0.258	0.300**	0.534*	0.659***
	(0.220)	(0.123)	(0.276)	(0.229)
NABIDJAN	−0.075	−0.101**	−0.218***	−0.182***
	(0.065)	(0.046)	(0.075)	(0.064)
INEFF	−0.002	−0.006	−0.008	−0.010*
	(0.002)	(0.004)	(0.005)	(0.006)
AFRICAN	0.078	0.136*	0.022	0.051
	(0.067)	(0.082)	(0.084)	(0.089)
EUROPEAN	0.036	0.038	−0.072	−0.013
	(0.064)	(0.046)	(0.072)	(0.063)
ASIAN	−0.063	−0.049	−0.060	−0.019
	(0.071)	(0.050)	(0.102)	(0.079)
TEXTILES	0.133*	0.086	−0.028	−0.031
	(0.069)	(0.056)	(0.076)	(0.063)
WOOD	0.049	0.045	0.105*	0.101*
	(0.047)	(0.046)	(0.060)	(0.057)
METAL	0.076*	0.066*	0.056	−0.017
	(0.040)	(0.035)	(0.077)	(0.054)
Constant	0.374***	0.329***	0.182	0.070
	(0.131)	(0.122)	(0.129)	(0.112)
N	107	129	66	107
R-Adj.	0.372	0.397	0.448	0.357

Standard errors (in parentheses) are estimated using White's consistent estimator (White, 1980).

(From Sleuwaegen and Goedhuys, 2002)

Illustration of implied relationship between size, age and growth (based on col. 1, T3):



- Caveat: The relationship between size and growth could in fact be spurious. The problem arises whenever there are transitory fluctuations in size or whenever there are transitory measurement errors in observed size. The resulting bias in the estimated relationship between initial size and growth is negative, hence failure to address this problem can produce a picture of the growth of small firms that is too good. In view of this, the conclusion drawn by Sleuwaegen and Goedhuys (2002) that the "...the results go against Gibrat's law of random growth behaviour" may be premature.