Journal of International Development: Vol. 8, No. 3, 445–466 (1996)

LEARNING BY DOING IN SUB-SAHARAN AFRICA: EVIDENCE FROM GHANA

PATRICIA JONES AND ABIGAIL BARR
Institute of Economics and Statistics, University of Oxford, Oxford, UK

Abstract:There has been interest in the implications of learning by doing, and in particular in the possibility that learning by doing may be slower in less developed countries and in industries which use simpler technologies. This paper uses firm-level data from Ghana to estimate learning-by-doing effects and generates three main findings. First, the learning curve, though present, is flatter in Ghana than in developed countries. Second, any industry-wide spillovers are small and insignificant. Third, (contrary to the assumption of much theory) learning-by-doing effects are stronger at low levels of technology than at intermediate levels.

1 INTRODUCTION

In recent years there has been a renewed interest in the process of learning by doing (LBD) and the potential role it plays in generating economic growth. Numerous theoretical papers (Lucas, 1988; 1993; Romer, 1986; 1993; Stokey, 1988; 1991; Young, 1991; 1993, Chamley, 1993) have modelled the mechanisms by which differing rates of LBD can lead to sustained differences in productivity across countries. One of the assumptions underlying these models is that the production of some goods induces a higher rate of learning than others. This implies that a country's rate of human capital formation is indirectly determined by its product mix. Since lower income countries find it optimal in the short run to produce goods with lower learning potentials, the models predict that the growth rates of these countries will not converge toward those of high-income countries.

These new models provide a much sharper image of the development process. However, their validity has yet to be tested empirically using data from developing countries. In this paper, I employ micro data from Ghana to test several of the underlying assumptions of the endogenous growth models. Specifically, I test the following two hypotheses: (i) firms in low-income countries have flatter learning curves than firms in high-income countries; and (ii) firms that produce goods using basic technologies have lower rates of learning than firms that produce goods using advanced technologies.

This study contributes to the literature in several ways. First, it provides the only estimates to date on the relative size of the productivity effects of LBD in a

CCC 0954–1748/96/030445–22

less-developed country. It is shown that the rates of learning in Ghanaian manufacturing are substantially lower than those estimated for industries in high-income countries. Estimates of the learning curve in Ghana vary from 11 to 16 per cent which indicates that a doubling of cumulative output is accompanied by a 9 to 12 per cent rise in productivity. For high-income countries, the learning curves are much steeper. Empirical evidence suggests that productivity can rise by as much as 80 per cent with each doubling of output in industrial economies. Second, the study finds some evidence that the rates of learning in Ghanaian manufacturing vary with firm technology. The results suggest that productivity rises very rapidly at low levels of technology (i.e., hand tools) as firms age. Productivity also rises at high levels of technology (i.e. computers) in firms with greater worker experience. These results provide evidence that different rates of learning by doing are associated with different technologies, but that the relationship is non-monotonic.

2 A BRIEF OUTLINE OF THE LEARNING-BY-DOING LITERATURE AND HOW IT RELATES TO A DEVELOPING COUNTRY

The concept of learning by doing dates back to the theoretical work of Arrow (1962) who suggests that all 'learning is a product of experience' (p. 155). According to Arrow, the best economic variable to capture learning by doing is cumulative gross investment because firms use investment to buy (or produce) new machines. Arrow proposes that machines stimulate learning by altering the production environment. As firms acquire new machines, workers learn as they change their production behaviour and this, in turn, makes them more adept at adopting machinery the next time round. The crux of Arrow's model is that the firm's stock of knowledge rises with cumulative gross investment (independent of the date at which the investment occurred) which enables firms to introduce new technologies and continue to learn without bound. The model proposes that learning by doing is the force behind advances in technology and economic growth.

Essential to the model is the assumption that learning by doing creates externalities in production. That is, the productivity increases resulting from LBD affect not only the firm where the learning takes place but also other firms engaged in similar production processes. Because the spillover effects are not solely proprietary to the firm, investment in LBD falls short of the socially optimal level. This result has been used to support a wide range of interventionist policies in low-income countries. Several theoretical papers argue that the externalities generated by learning by doing are large enough to warrant either infant-industry protection (Bardhan, 1970; Clamhout and Wan, 1970; Succar, 1987) or protectionist trade policies (Stokey, 1991; Young, 1991).

Since Arrow's seminal paper the concept of learning by doing has changed in several ways. First, it is now generally assumed that learning rates vary depending upon the goods being produced. So, a country's rate of LBD is determined by the type of goods it produces. The idea that goods are valued according to their characteristics was first applied in this context by Krugman (1987) and then later adapted in the growth models of Stokey (1988) and Lucas (1988). According to Stokey, goods can be indexed by the number of characteristics they provide where higher-index goods are considered 'better' because they provide more characteristics than lower-indexed goods. For example, steak and gruel are two foods with many of the same characteristics (e.g., vitamins, calories, protein, etc.) but steak is strictly 'better' in the sense that it is much tastier than gruel. In Lucas's model, the 'better' goods are those produced by more advanced technologies because it is assumed that high-technology goods are associated with faster rates of learning by doing.

The growth models of Stokey (1988) and Lucas (1988) have played an important role in rekindling economists' interest in learning by doing and its role in the development process. The basic argument put forth by Stokey is that the set of goods a country produces changes as it develops with higher-indexed goods replacing lower-indexed goods. Higher indexed goods induce faster economywide learning which causes productivity to rise and economic growth to continue without bound. Productivity rises because LBD enables a country to produce not only its current set of goods more efficiently but also a new set of higher-quality goods. Stokey extends this idea in a later paper (1991) which develops a general equilibrium model in which the South (i.e., less developed countries) always finds it optimal to produce lower quality goods than the North (i.e., more developed countries) under a free trade regime.

Several strict assumptions underlie Stokey's model. First, she assumes that learning by doing spillovers exist among goods (i.e., the knowledge gained in one production process can be applied to another production process). Second, she assumes that knowledge reduces the cost of all characteristics (i.e., all goods), although the cost reduction is greater for higher-index goods. In addition, the rates of learning by doing are faster for higher-indexed goods. Lastly, she assumes that societal knowledge rises as a result of learning by doing (i.e., a country's stock of knowledge depends upon its cumulative production). This last assumption implies that history matters in the development process.

Lucas (1988) also develops a multi-good model in which learning rates vary across different goods. In his model countries choose to produce those goods in which they have a comparative advantage based on their initial skill endowment. Lucas (1988) states 'that comparative advantages that dictate a country's initial production mix will simply be intensified over time by human capital accumulation' (p. 41). Like Stokey, Lucas regards knowledge accumulated through learning by doing as an essential determinant of economic growth. They both predict that: (i) the learning curve in developing countries are flatter than the learning curve in developed countries; and (ii) the learning rates associated with high-tech goods are higher than those associated with basic technologies.

Another development in the learning by doing literature took place within the context of information theory rather than growth theory. Several learning models have been proposed that are based on the assumption that learning rates vary among individuals with different levels of schooling or work experience. Stiglitz (1987; 1989) was among the first to apply this idea in a production context, suggesting that experience improves the ability to learn how to perform a task. According to Stiglitz, individuals learn how to learn and become more efficient at learning as they gain experience—a process he calls 'learning by learning'. Jovanovic and Nyarko (1995) present a more sophisticated version of this idea by developing a Bayesian learning model in which on-the-job experience improves the ability to gather and decode new information. In their model, experience enables an individual to acquire information about the parameters of the functional form which underlies his/her production activity. An important assumption of the model is that an understanding of the functional form will increase an individual's efficiency in carrying out his/her production activity. The basic framework of the Jovanovic and Nyarko (1995) is explained most easily by an example.

Suppose we are interested in modelling the learning process of a manager who makes daily decisions about how fast to set the speed of a production line. His ideal target (say, y) is the speed at which productivity is maximized. It is assumed that the manager does not know the exact value of y. In the case of a production line, there may be disturbances (e.g., machine break downs, worker illness, etc.) which affect the daily value of y. These disturbances (call them w) are assumed to be random variables with mean zero and variance $\sigma^2{}_w$. Although the manager does not know the exact value of y, he does know its distribution, except its mean value (call it θ). Each day the manager makes some decision (say, z) as to what speed to set the production line. The manager must choose z before y is revealed to him. Obviously, the manager hopes that his choice of z will match his ideal target, y. It is assumed that output is given by $\phi[1-(y - z)^2]$, so that any mistakes he makes will result in a loss of output that takes the value $\phi(y - z)^2$, where ϕ is some activity-specific weight which reflects the productivity gains associated with getting the choice of x right. At the end of each day the manager views his mistakes (and successes) and thus acquires more information about θ. As he accumulates more information, his choice of z improves (i.e., z becomes closer to y). This is, greater experience leads to an upward-sloping learning curve.

If we assume that λ is the total number of production runs that the manager has administered, then his expectation of the production line's optimal speed, denoted by $E_\lambda(y)$, will depend on the amount of information he has acquired during the previous λ production runs. That is, the manager's optimal decision is:

$$z_\lambda = E_\lambda\,(y_\lambda) = E_\lambda(\theta). \tag{1}$$

If $x_\lambda = E_\lambda(\theta))^2$ is the posterior variance over θ after λ production runs, then Bayesian updating implies that x_λ will fall as λ increases. Then the expected productivity on production run λ is:

$$E_\lambda(q_\lambda) = \phi(1 - x_\lambda - \sigma^2{}_w), \tag{2}$$

where q is the level of productivity associated with speed x.

The inclusion of a weight ϕ implies that certain types of production activities lead to a larger variance in q than others. According to Jovanovic and Nyarko, the variance is bigger for more complex tasks and it is a monotonic function of experience. Essentially, this proposition is the same as the hypothesis proposed by Lucas (1988) that high-tech goods have higher rates of learning than low-tech goods. The complementarity between the two models should be clear: the Bayesian learning model explains *why* workers with the same level of experience may have varying levels of learning by doing, while the models of Lucas (1988) and Stokey (1988) provide a framework for explaining *how* these differential rates of learning affect growth. That is, the learning models provide an important microfoundation for the endogenous growth theory.

So, what is the relevance of all these models to developing countries? As stated above, theory predicts that the rates of learning in less-developed countries are likely to remain lower than the rates of learning in more-developed countries because poor countries have a comparative disadvantage in human capital. Low-income countries find it optimal (under competitive conditions) to specialize in products with lower learning potentials because they have lower stocks of human capital and are unable to transfer the learning they need from abroad. Therefore, the learning models predict that the productivity growth rates in less-developed countries will not converge toward the rates experienced by more-developed countries. Instead, less-developed countries will continue to have lower stocks of human capital than more-developed countries and, as a consequence, they will remain poorer.

Given the gloomy scenario predicted by endogenous growth theory, is there any policy advice that economists can offer low-income countries for raising their economywide rates of learning by doing? There is no clear answer because so little empirical evidence exists on the economic factors which affect learning rates in both poor and rich countries. However, one type of policy that continues to gain a great deal of attention in both academic and political circles is the implementation of short-term protectionist policies. It is often argued that infant-industry protection and protectionist trade provide an environment which enables low-income countries to increase their stocks of human capital by switching production from low-technology goods to high-technology goods. Unfortunately, the problem with such quick-fix solutions is that they produce only level effects rather than growth effects. As pointed out by Lucas (1993), 'this is a one-time stimulus to productivity, and thereafter the mix of goods produced in this closed system can change only slowly, as the consumption mix changes' (p. 270). An alternative solution is to urge countries to produce only those goods that are near their quality frontier. However, the selection of such goods is certainly a formidable (if not impossible) task.

One of the major aims of this paper is to present some preliminary evidence on the size of the gap in learning rates which occurs between low-income and high-income countries. The format of the paper is as follows. In Section 3 below I discuss the data employed for analysis and the estimation techniques used to test each of the two assumptions underlying the endogenous growth models. Section 4 presents the results on learning by doing in Ghana and compares these results with other studies. Finally, Section 5 offers some possibly policy recommendations based on the results revealed in this paper.

3 DATA AND EMPIRICAL SPECIFICATION

3.1 Data

The data used in this analysis are from a panel survey of 200 manufacturing firms in Ghana organized under the World Bank's 'Regional Programme for Enterprise Development' (RPED) during the summers of 1992–94. This survey is part of a nine country (Burundi, Cameroon, Côte d'Ivoire, Ghana, Kenya, Rwanda, Tanzania, Zambia and Zimbabwe) study of the manufacturing sector in Africa which was funded by several European governments and the Canadian government. The Ghana case study was financed by the British Overseas Administration (ODA) and conducted by a team combining staff from the Centre for the Study of African Economies at Oxford University and the University of Ghana at Legon.

The data collected are extremely rich for an industrial survey and provide numerous indicators of how firms in Ghana have performed in the structural adjustment (i.e., 1983–93) period. These data also include output levels (in Cedis) for each firm's initial year of production which enable cumulative output to be extrapolated for all the firms. The sample includes firms which operate in seven three-digit manufacturing sectors (food processing, garments, textiles, wood products, furniture, machinery and metal products), representing about 80 per cent of manufacturing firms. Each of these firms is located in one of Ghana's four major industrial areas (Accra, Kumasi, Takoradi and Cape Coast).

Although 200 firms were interviewed in each of the three years, the final sample for the study was reduced to 179 observations. Missing data are mainly the result of observations being dropped from the 1991 data. The 1991 date are used as instrument variables only. In addition, the final sample deleted all public sector enterprises and firms that had been founded during the previous year. The rationale for deleting the public enterprises was that several of these firms are under receivership or barely producing so it was assumed that little (if any) learning by doing is occurring in these firms. Young firms were deleted because there were no data available to use as lagged endogenous variables.

3.2 Estimation Approach

This paper estimates a learning curve for Ghanaian manufacturing using an augmented Cobb–Douglas production function. Such an approach is slightly unusual, although not completely unprecedented in the papers which estimate learning curves. The most common approach is to use progress functions for estimating the reduction in unit costs resulting from cumulated firm experience. Such studies have been carried out on the following US industries: semiconductors (Irwin and Klenow, 1994), chemicals (Lieberman, 1984), radar equipment (Preston and Keachie, 1964), and machinery (Hirsch, 1952). Examples of learning curves that were estimated within the context of a production function are more limited. Rapping (1965) used production functions to estimate the learning effect in ship building and, more recently, Bahk and Gort (1993) used a production framework to decompose learning by doing for a cross-section of US industries. One approach is really no better than the other: the choice of which to use depends on the quality and availability of data.

Theoretically, learning by doing can enter the production function in several ways. It can be included as a separate argument in the production function (as it is in this paper) or it can be modelled as a shift parameter.[1] Empirically, it is feasible to estimate the growth effects of learning by doing if there are appropriate variables (or proxies) which measure the change in the stock of firm knowledge. Since no such variables exist in the Ghana data, this study estimates learning by doing as a separate parameter in the production function. I assume the production function takes the following form

[1] Bahk and Gort (1993) estimate both cases in an innovative paper which decomposes learning by doing into organisational learning, manual task learning and capital learning.

$$Y = f(K,L,X) \tag{3}$$

where K is the effective capital input, L is the effective labour input and X is the stock of firm knowledge. It is further assumed that the stock of firm knowledge is a function of the amount of cumulated learning by doing. That is, $X = q(V)$ where q represents the efficiency in which production activities are carried out and V captures the level of firm-specific learning by doing.

The following functional form is used to estimate the learning curve which assumes that learning by doing enters the production function in a power form

$$Y = AK^{(1-\alpha)} L^{\alpha} V^{\gamma} \tag{4}$$

where A is a constant, $(1 - \alpha)$ is the elasticity with respect to capital, α is the elasticity with respect to labour and γ is the elasticity with respect to learning by doing. The interpretation of γ is straightforward: a one per cent rise in cumulated experience results in γ per cent increase in output. An alternative way of interpreting the parameter γ is to apply the simple formula $(2^{\gamma} - 1)$. This formula indicates that each doubling of cumulative firm output is associated with a γ per cent rise in productivity. Theoretically, the γ can also be interpreted as a returns to scale parameter, as well as the effect of learning by doing. That is, the elasticity of output with respect to total input is estimated by the sum of the exponents on capital, labour and the stock of firm knowledge. Although the aim of this study is not to measure the internal economies in production, it should be pointed out that estimates of internal economies of scale may be upwardly-biased if significant external economies are present.[2] To eliminate possible bias due to the unobservable effects of external economies, several specifications of the model include aggregate industry output.

In order to estimate these functional forms, a few algebraic manipulations need to be made to the production functions. First, I assume that the labour input is not homogenous but, instead, comprised of four categories of workers who are differentiated by their level of educational attainment. That is:

$$Y = AK^{(1-\alpha)} \left(L_0 + \sum_{i=1}^{3} (c_i + 1) L_i \right)^{\alpha} V^{\gamma} \tag{5}$$

where

$$L = \sum_{i=0}^{3} L_i . \tag{6}$$

In equation (5) L_i represents the total number of workers in a firm who have completed i level of education where L_0 is the number of workers with no formal schooling, L_i is the number with basic education, L_2 is the number with secondary schooling and L_3 is the number with tertiary education. The parameters c_i reflect differences in labour productivity between L_i and the base category L_0. Differentiation of equation (5) with respect to L_i reveals that $c_i + 1$ is simply the ratio of the average marginal products of the workers in educational category i relative to those in the base category, L_0.

An algebraic transformation of equation (5) can be carried out by defining $\lambda_i = L_i / L$, so that equation (5) becomes

[2] See Hall (1988).

$$Y = AK^{(1-\alpha)} L^{\alpha} \left(1 + \sum_{i=1}^{3}(c_i\lambda_i\right)^{\alpha} V^{\gamma}. \tag{7}$$

In equation (7) the variable λ_i represents the proportion of each firm's workforce whose highest educational achievement is category i which implies that $\Sigma\lambda_i = 1$. That is, the coefficient on each λ_i represents the productivity differential between educational category i and the base category λ_0. By definition, if c_i is greater than zero, then the workers in educational category λ_i are more productive than the workers in the base category λ_0.

By dividing both sides by L and taking logarithms, we get

$$Ln\left(\frac{Y}{L}\right) = \ln A + (1 - \alpha)Ln\left(\frac{K}{L}\right) + \alpha\ln\left(1 + \sum_{i=1}^{3} c_i\lambda_i\right) + \gamma\ln V. \tag{8}$$

However, if we make use of the Taylor series approximation that $\ln(1 + x) \simeq x$, then it is possible to rewrite equation (8) as

$$Ln\left(\frac{Y}{L}\right) = \ln A + (1 - \alpha)\ln\left(\frac{K}{L}\right) + \alpha\sum_{i=1}^{3} c_i\lambda_i + \gamma\ln. \tag{9}$$

Finally, if we relax the assumption of constant returns to scale by adding $\theta\ln L$, then we get the productivity equation used to estimate the impact of learning by doing.

$$\ln\left(\frac{Y}{L}\right) = \ln A + (1 - \alpha)\ln\left(\frac{K}{L}\right) + \Theta lnL + \alpha\sum_{i=1}^{3} c_i\lambda_i + \gamma\ln V. \tag{10}$$

As stated above, the estimates of γ represent the productivity effects associated with learning by doing. Implicit in this functional form is the assumption that learning by doing increases the productivity of all the inputs by the same amount. That is, the models estimate *organizational* learning which is distinct from learning that augments either labour or capital.

Before discussing the variables used for estimation, it is necessary to mention how to interpret the estimates of c_i. Notice that in equation (10) the coefficients on λ_i are αc_i rather than c_i which means that the c must be separated from this multiplicative term. Fortunately, the parameters estimates on c are straightforward to calculate because they are simply the coefficients on λ_i divided by one minus the coefficients on K/L. That is,

$$c_i = \frac{\alpha c_i}{[1 - (1 - \alpha)]}. \tag{11}$$

However, calculating the standard errors is a bit more cumbersome and requires adjusting the standard errors using the variance of αc_i, the variance of $\ln(K/L)$, and the covariances between the two variables.[3] These adjustments have been made to all the results reported in the tables.

In this study four variables are used to capture the productivity effects of learning by doing. These variables are: (i) cumulative firm output, denoted by V_1, (ii)

[3]See Jones (1994) for a more thorough discussion of how to estimate the standard errors.

cumulative firm output per labour input, denoted by V_2; (iii) average years of worker experience in current firm, denoted by Exp, and (4) the age of the firm, denoted by Firmage. The values of V_1 and V_2 had to be extrapolated because the RPED data do not contain production data for the entire history of each firm. However, the RPED data do contain several years of pre-survey data (i.e., data on production prior to 1991) from which growth rates were calculated and then used to extrapolate the values for any missing years of data. There were three questions in the RPED survey that enabled me to estimate the cumulative output of a firm. These questions asked: (i) the value of production during the year in which the firm was founded; (ii) the value of production for 1983; and (iii) the value of production for 1988.[4] These three years of data were used first to estimate the growth rates in output between various years and then to impute the values of output from the estimated growth rates.

The basic framework of equation (10) is based on a modified Cobb–Douglas production function whose residual includes the effect of numerous omitted variables. These effects are well-known in the literature on productivity and include such factors as technological capability, unionization, entrepreneurial ability and worker effort. Several additional variables are added to the final productivity equation in order to control for these effects. These variables include the entrepreneur's schooling, 6 industry dummies, the logarithm of average hours worked per week, a union dummy, the proportion of apprentices, 2 technology dummies, and the complexity of tasks in the firm. See the Appendix for a full definition of these variables and all others that are estimated by the productivity equation.

One problem encountered when estimating equation (10) was that capital, labour, and learning by doing are all endogenous to the models. Even though cumulative output refers to past output, it is very likely that there will be serial correlation in the unobserved firm level factors that make firms productive. The usual solution to such a problem is to use instrumental variables. However, the instrumental variable approach is not recommended for small samples when only poor instruments are available. 'Poor' instruments are those with a low R-squared in the first stage regression of the instrument on the endogenous variable. Nelson and Startz (1990a,b) have shown that IV estimates are more biased than OLS estimates when poor instruments are used. In this study, several first stage regressions are estimated to find an instrumental variable for cumulative output. None of the specifications produced 'good fits'. Moreover, the potential bias from using OLS may be small (Caballero and Lyons, 1990).

The method used in this study to control for possible firm quality differences was to add the value of firm output per worker in 1988 to the productivity equation. It is hypothesized that firms with higher productivity in 1988 will be closely correlated to firms with higher productivity in 1992 and 1993. This addition of output per worker in 1988 did not change the significance of either V_1 or V_2 nor their parameter estimates. In the final specification, all the endogenous variables are instrumented by lagging the values by one year.

[4] The years 1983 and 1988 are important years in the economic history of Ghana. Structural adjustment policies were first implemented in 1983 and carried on until 1988 when the last major policies were carried out.

Table 1. Survey of selected empirical studies on learning by doing.

Author(s) & Date	*Sector(s) (country)*	*Variable(s) used to proxy LBD*	*Data*	*Estimation technique(s)*	*Results*
T-Y. Chuang (1996)	Manufacturing (Taiwan)	(1) Cumulative aggregate industry output	One set of industry level data and one set of firm level data on 17 two-digit manufacturing industries in Taiwan for the period 1974–1990	OLS and IV estimation of a generalised Cobb–Douglas production function. Model discriminates between internal and external economies	(1) Trade-induced learning by doing accounts for about 40% of output growth (2) Returns to scale are significantly diminished when external learning effects are included in the model; returns to scale at the industry level fall from 1.25 to 0.60
B. Jovanovic and Y. Nyarko, (1995)	Various industries (US), including the medical profession (heart surgery), aviation (air traffic controllers), insurance, steel and electricity	(1) Cumulative number of activities completed	Various sources, including data on the success rates of operations, plane landings insurance sales, etc.	(1) Fitted learning curves where the curves represent the squared sum of the deviations between the expected success rate and the average success rate (2) Progress ratios defined as the ratio of actual to expected productivity	(1) Success rates increase with the number of experiments (2) Coefficients of variation decline as more experience is gained in the activity
D. A. Irwin and P. J. Klenow, (1994)	Semiconductor (US)	(1) Generations of dynamic random access memory (DRAM) chips	Quarterly data on average industry selling price and firm-level shipments	(1) The log of the market price is regressed on the log of the world cumulative output (2) Experience is regressed on cumulative firm output, the firm's contribution to industry output and the contribution of national output to world output	(1) The market price falls by 16 to 24 per cent with each doubling of firm output (2) The effects of within country production are greater than the effects of international spillovers

R. S. Jarmin, (1994)	Rayon (US)	(1) Cumulative firm output	Dataset covers the first 28 years (1911–1938) of rayon production	The log of firm prices are regressed on the log of firm output, the log of non-price demand determinants, time, and a strategic parameter for each firm	(1) Firms exhibit asymmetric rates of learning by doing and asymmetric learning spillovers
B. Bahk and M. Gort (1993)	Forty-one manufacturing industries (US)	(1) Cumulative firm output (2) Cumulative firm output per worker (3) Firm age	Annual data on firm-level shipments and capital were drawn from the Bureau of Economic Analysis. Labour data were obtained from the US Census	Cobb–Douglas framework which regresses the log of firm shipments on the log of physical capital, the log of labour, and the log of firm knowledge	(1) A 1% rise in cumulative output is associated with a 3% change in output (2) A 1% rise in cumulative output per worker leads to a 8% rise in output (3) One additional year of firm production is correlated to a 1% change in output
A. D. Foster and M. Rosenzweig (1993)	Rural labour markets (India, the Philippines, and Pakistan)	(1) Time	Household data	2SLS estimation in which employer ignorance is regressed on cumulative days in the labour market and other explanatory variables	Increased exposure in the labour market reduces employer ignorance; each additional year in the labour market reduces the differential between time rate and piece rate by 6.5%
R. J. Caballero and R. K. Lyons (1992)	Manufacturing (US)	(1) Cumulative industry output	Two-digit manufacturing data for the period 1959–84	OLS and IV estimation of a generalised Cobb–Douglas production function; model discriminates between internal and external economies	Economies of scale are much larger at the aggregate industry level than they are for any two-digit industry
R. J. Caballero and R. K. Lyons (1990)	Manufacturing (West Germany, France, the UK, and Belgium)	(1) Cumulative industry output	Two-digit manufacturing data for the period 1970–86	OLS and IV estimation of a generalised Cobb–Douglas production function; model discriminates between	All four countries exhibit strong increasing returns to scale at the aggregate industry level. Only 3 of the

Table 1. Survey of selected empirical studies on learning by doing.

Author(s) & Date	*Sector(s) (country)*	*Variable(s) used to proxy LBD*	*Data*	*Estimation technique(s)*	*Results*
				internal and external economies	13 industries displayed evidence of internal economies of scale
M. B. Lieberman (1984)	Chemicals (US)	(1) Cumulative industry output (2) Time (3) Cumulated industry capacity (4) Annual rate of industry output (5) Average scale of plane (6) The rate of new plant investment	Annual firm-level data on 37 chemical products. These data account for about 25% of cumulative industry output	The log of average market price of regressed on cumulative firm output. Other specifications include time, cumulated industry capacity, annual rate of output, and average plant scale, and the rate of new plant investment as explanatory variables	The average market price falls by approximately 28% with each doubling of cumulative firm output
D. Levhari and E. Sheshinski, (1973)	Diamonds (Israel)	(1)Average experience of all workers	Plant-level data on the diamond industry collected by the Central Bureau of Statistics in 1961	An augmented Cobb-Douglas function which regresses the log of firm output on the log of capital, the log of labour, and the log of the stock of firm knowledge	An increase in average experience of one year is associated with a raise in output of 9%
E. Sheshinski (1967	Manufacturing Industries in seven countries (the US, Great Britain, Canada, West Germany, Australia, and Norway)	(1) Cumulated gross output (2) Cumulated gross investment	Annual cross-sectional and time series data for the period 1950–60; the US data covers the period 1929–64	CES production function in which the log of output per worker (or output per capital unit) is regressed on the log of wages (or capital costs), the log of cumulative output for cumulative investment) and time	(1) The rise in productivity associated with an increase in gross cumulated output (or gross cumulated investment) varies from 14% to 18% across countries and time

					(2) Productivity rises by about 10 to 12% with each doubling of cumulated gross output (or gross investment) (3) Learning rates vary substantially across countries. The following learning coefficients were estimated for six industrial countries. (1) 0.85 for the US, (2) 0.20 for West Germany, (3) 0.58 for Canada, (4) 0.16 for Britain, (5) 0.24 for Norway, and (6) 0.48 for Australia
L. Rapping (1965)	Liberty War Ships (US)	(1) Cumulated gross output (2) Time	Time series and cross-sectional data from 15 new ships yards set up during either 1941 or 1942. Data cover the period 1941–45	Cobb–Douglas production function in which the log of annual output is regressed on the log of capital, the log of labour, the log of cumulated output, and time	(1) A doubling of cumulated output is associated with a 29% rise in output (2) Output rises by 23% per year due to a shift in the production function alone
L. Preston and E. C. Keachie, (1964)	Radar Equipment (US)	(1) Cumulative Quantity	Twenty-two observations on unit costs and the volume of output per production run for five types of radar equipment	The log of units costs are regressed on the log of lot size and the log of cumulated output (as well as other specifications)	A doubling of cumulative output is accompanied by a 53% fall in unit costs
W. Z. Hirsh, (1952)	Machine Tools (US)	(1) Cumulative Output	Monthly data from a large machine builder for the period 1946–50	Progress ratios defined as the decline in labour requirements associated with each doubling of cumulative output	Unit labour costs fall by about 19% with each doubling of cumulative output

Table 2. Estimates of the impact of learning by doing on firm productivity. Source: RPED Ghana data.

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	9.5138[1]	9.3121[1]	11.297[1]	10.730[1]	9.1631[1]	8.9054[1]
	(1.492)	(1.418)	(1.299)	(1.394)	(1.499)	(1.469)
ln(*K/L*)	0.0897	0.0919	0.0764	0.0848	0.0785	0.0774
	(0.063)	(0.062)	(0.064)	(0.064)	(0.064)	(0.063)
ln(FS)	−0.4593	−0.5825[2]	−0.3814	−0.4978	−0.4919	−0.5970[2]
	(0.292)	(0.288)	(0.272)	(0.282)	(0.274)	(0.276)
Basic	0.2966	0.2371	0.3403	0.4379	0.2900	0.2311
	(0.334)	(0.314)	(0.346)	(0.348)	(0.334)	(0.299)
Second	0.4146	0.3474	0.5862	0.6363	0.3488	0.2757
	(0.396)	(0.378)	(0.394)	(0.400)	(0.401)	(0.364)
Tertiary	0.7273	0.6551	0.8622	0.9317	0.7104	0.6279
	(0.538)	(0.545)	(0.470)	(0.527)	(0.510)	(0.470)
ln(V_1)	0.1199[2]				0.1215[2]	
	(0.051)				(0.055)	
ln(V_2)		0.1613[1]				0.1611[1]
		(0.054)				(0.052)
Exp			0.1090		0.1158	0.1208
			(0.072)		(0.066)	(0.065)
Exp^2			−0.0042		−0.0032	−0.0034
			(0.003)		(0.003)	(0.003)
Firmage				0.0410	0.0020	0.0131
				(0.026)	(0.029)	(0.026)
Firmage^2				−0.0011	−0.0005	−0.0005
				(0.001)	(0.005)	(0.000)
Food	−0.1362	−0.1572	−0.0555	0.0674	−0.0256	−0.0726
	(0.295)	(0.293)	(0.292)	(0.301)	(0.292)	(0.295)
Garments	−0.2935	−0.2043	−0.4172	−0.3206	0.1715	−0.1711
	(0.362)	(0.353)	(0.363)	(0.360)	(0.359)	(0.347)
Wood	0.0473	0.0359	0.0399	0.0334	0.1796	0.1772
	(0.388)	(0.380)	(0.408)	(0.393)	(0.415)	(0.408)
Furniture	−0.4032	−0.4105	−0.3980	−0.2506	−0.2901	−0.3527
	(0.377)	(0.367)	(0.357)	(0.348)	(0.338)	(0.338)
Metal	0.0907	0.1077	−0.0132	0.1718	0.0926	0.0547
	(0.316)	(0.314)	(0.327)	(0.320)	(0.321)	(0.317)
Machinery	−0.6170	−0.5384	−0.7810	−0.5703	−0.5117	−0.5275
	(0.443)	(0.444)	(0.448)	(0.455)	(0.455)	(0.448)
ln(hrs)	0.1026	0.0639	0.0938	0.2065	0.1132	0.0617
	(0.253)	(0.257)	(0.239)	(0.253)	(0.243)	(0.246)
Union	0.1052	0.0973	0.2108	0.2963	0.1045	0.0487
	(0.346)	(0.339)	(0.320)	(0.328)	(0.310)	(0.306)
Papp	−0.6892[2]	−0.7227[1]	−0.5311	−0.8306[1]	−0.4849	−0.4174
	(0.302)	(0.295)	(0.362)	(0.321)	(0.347)	(0.343)
Hand	0.2980	0.3422	0.1432	0.2333	0.3006	0.3156
	(0.231)	(0.231)	(0.220)	(0.224)	(0.223)	(0.220)
Comp	1.4373[1]	1.4374[1]	1.5195[1]	1.4377[1]	1.5639[2]	1.5672[2]
	(0.398)	(0.402)	(0.360)	(0.411)	(0.389)	(0.378)
Ntasks	0.1536[1]	0.1487[2]	0.1396[1]	0.1449[1]	0.1578[2]	0.1530[2]
	(0.050)	(0.050)	(0.048)	(0.048)	(0.048)	(0.048)
Y1988	−0.0003	−0.0004	0.0001	0.0002	−0.0003	−0.0003
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year93	−0.3798[2]	−0.3614[2]	−0.3082	−0.3537[2]	−0.3764	−0.3568
	(0.169)	(0.169)	(0.165)	(0.168)	(0.165)	(0.166)
Adj. *R*-squared	0.3594	0.3744	0.3480	0.3510	0.3700	0.3833
Number of Observations	179	179	179	179	179	179

Notes: All standard errors have been corrected for heteroskedasticity using White's (1978) procedure. Statistical significance at the 0.01, and 0.05 levels are indicated by [1] and [2], respectively. Other explanatory variables include 6 industry dummy variables.

4 ESTIMATION RESULTS

Table 1 presents a list of empirical studies on learning by doing and related types of learning. Previous empirical studies of learning by doing in the United States and other industrial economies reveal a wide range of learning effects across industries. For example, the cross-country study by Sheshinski (1967) reveals that the productivity gain associated with a doubling of cumulative output ranges from 12 per cent in the UK to over 80 per cent in the US. Rates of learning also vary significantly across industries. Estimates of the productivity effects of learning by doing in the US range from 19 per cent in machine tools (Hirsch, 1952) to 29 per cent in ship building (Rapping, 1965). As stated above, a primary objective of this study is to compare the estimated rates of learning in Ghanaian manufacturing with those obtained for industries in more developed countries.

Table 3. Estimates of the impact of learning by doing and industry learning spillovers on firm productivity. Source: RPED data.

	(1)	(2)	(3)	(4)
Constant	8.2429[1]	8.1828[1]	10.543[1]	10.771[1]
	(1.647)	(1.598)	(1.231)	(1.313)
ln(K/L)	0.1008	0.1023	0.0767	0.0855
	(0.062)	(0.061)	(0.063)	(0.064)
ln(FS)	−0.5616[2]	−0.6819[1]	−0.3902	−0.5418[2]
	(0.260)	(0.261)	(0.191)	(0.255)
Basic	0.2318	0.1904	0.2832	0.3172
	(0.330)	(0.315)	(0.341)	(0.332)
Second	0.3475	0.3000	0.5636	0.5763
	(0.373)	(0.359)	(0.382)	(0.368)
Tertiary	0.5130	0.4406	0.7852	0.8042
	(0.529)	(0.535)	(0.500)	(0.503)
ln(V_1)	0.1112[1]			
	(0.046)			
ln(IndV_1)	0.0408			
	(0.029)			
ln(V_2)		0.1481[1]		
		(0.049)		
ln(IndV_2)		0.0338		
		(0.030)		
Exp			0.0943	
			(0.73)	
Exp^2			−0.0037	
			(0.004)	
Indexp			0.1304	
			(0.083)	
Firmage				0.0336
				(0.027)
Firmage^2				−0.0009
				(0.000)
Indage				0.0211
				(0.039)
Adjusted R-squared	0.3640	0.3776	0.3528	0.3509
Number of Observations	179	179	179	179

Notes: All Observations have been corrected for heteroskedasticity using White's (1978) procedure. Statistical significance at the 0.01 and 0.05 per cent levels are indicated by [1] and [2], respectively. Other explanatory variables include 6 industry dummy variables, the logarithm of average hours worked per week, 2 technology dummies, the average number of worker tasks, proportion of apprentices in the workforce, average output per worker in 1988, and year dummy.

Table 4. Complementarity between learning by doing and firm technology. Source: RPED data.

	(1) Cumulative firm output, V_1	(2) Cumulative firm output, V_2	(3) Average worker experience, Exp	(4) firm age, (Firmage)
Constant	9.5148[1]	9.2619[1]	11.689[1]	10.762[1]
	(1.517)	(1.416)	(1.291)	(1.334)
ln(K/L)	0.0834	0.0872	0.0772	0.0859
	(0.063)	(0.061)	(0.063)	(0.063)
ln(FS)	−0.4629	−0.5795[2]	−0.2873	−0.4882
	(0.292)	(0.296)	(0.278)	(0.272)
Basic	0.3116	0.2563	0.3601	0.4737
	(0.331)	(0.311)	(0.352)	(0.355)
Second	0.3949	0.3280	0.6337	0.7345
	(0.396)	(0.381)	(0.423)	(0.416)
Tertiary	0.7979	0.6774	0.8576	1.0010[2]
	(0.530)	(0.540)	(0.485)	(0.507)
LBD	0.1322[1]	0.1818[1]	0.0530	0.0445
	(0.057)	(0.057)	(0.641)	(0.024)
lbd^2			−0.0024	−0.0015
			(0.004)	(0.001)
LBD*Hand	−0.0081	−0.0429	0.0900[2]	0.0500[1]
	(0.099)	(0.091)	(0.047)	(0.020)
LBD*Comp	0.4625	0.2864	0.1405	0.0539
	(0.360)	(0.202)	(0.085)	(0.053)
Hand	0.4223	1.0429[1]	−0.2301	−0.5124
	(1.744)	(1.585)	(0.316)	(0.349)
Comp	−8.3536	−4.8829	0.7326	0.3261
	(0.360)	(4.486)	(0.6863)	(1.219)
Adjusted R-squared	0.3561	0.3737	0.3539	0.3635
Number of Observations	179	179	179	179

Notes: All Observations have been corrected for heteroskedasticity using White's (1978) procedure. Statistical significance at the 0.01 and 0.05 per cent levels are indicated by [1] and [2], respectively. Other explanatory variables include 6 industry dummy variables, the logarithm of average hours worked per week, 2 technology dummies, the average number of worker tasks, proportion of apprentices in the workforce, average output per worker in 1988, and year dummy.

The variables used in the estimates are defined in Appendix and their mean values presented. The first variable listed is my dependent variable, the log of value added per unit of labour. Various productivity equations modelling this variable are given in Tables 2, 3 and 4. The learning curve for Ghanaian manufacturing is estimated using four proxies for learning by doing (i.e., V_1, V_2, exp, and firmage). In Table 2 the model estimated assumes that learning by doing enters the production function in power form. Both V_1 and V_2 are significant at the 5 per cent level which indicates that a firm's cumulative output has a significant impact on its productivity. The estimated coefficients on V_1 and V_2 are 0.1199 and 0.1613, respectively. By plugging these coefficients into the formula $2\gamma-1$, the learning curve is estimated to be 8.67 when LBD is proxied by cumulative firm output and 11.83 per cent when LBD is proxied by cumulative firm output per labour unit. That is, each doubling in cumulative firm output is accompanied by a rise in productivity by approximately 9 to 12 per cent.

With the exception of the studies by Bahk and Gort (1993) and Levari and Sheshinski (1964), all the learning curves estimated for the Ghanaian manufacturing industry are flatter than the curves estimated for industries in developed countries. The results obtained by Bahk and Gort suggest that output per worker rises by only 8 per cent with each doubling of cumulative output in US manufacturing. Their estimates of learning by doing are lower than those found for Ghanaian manufacturing and other industries in the US. It is difficult to determine why their results are so much lower than those estimated by previous studies, except that Bahk and Gort controlled for both labour quality and capital vintage in their production analysis. If learning by doing is strongly correlated with the vintage of capital (as Arrow assumes), then it is possible that the multicollinearity between the two variables reduced the size of the coefficient on LBD.

Given the potential measurements error that might be associated with imputed values of cumulative firm output, I choose to proxy LBD by average worker experience and firm age also. The size of the coefficient on exp is similar to that estimated for V_1 and V_2, although it is not significant. Firm age has no statistical significance either which may be interpreted as evidence that history is not an important factor in explaining productivity differentials across firms. One possible complication with using four proxies of learning by doing is that they may be capturing the same effects. For example, exp may be measuring the effect of both learning by doing, as well as better firm training or cohort effects reflecting differences in school quality. One way to determine whether the LBD proxies are measuring the same effects is to enter them all in the same productivity equation. Columns (5) and (6) present the estimation results when firm age and worker experience are entered simultaneously with V_1 and V_2 respectively. The results from column (5) and column (6) reveal that cumulative output is the best proxy for learning by doing since it remains significant in both specifications. From these results, it can be inferred that the two sets of variables are not substitutes for each other.

One important issue is whether there are industry level spillover effects in learning by doing. This is to say, does a firm's productivity increase if other firms in the industry have acquired learning by doing? Hall (1988) notes that omitting to control for these external LBD effects may bias upwards estimates of internal LBD. In Table 3, I investigate this by augmenting the productivity equations with average industry measures of LBD. Using the cumulative production, V_1 and V_2, the industry effects are insignificant, whilst the firm-specific measures remain significant at the 1 per cent level. Moreover, the size of the coefficients suggest that even if there are spillovers from industry LBD, these are small in comparison to those of firm LBD. The coefficients on the firm-specific LBD variables are somewhat smaller when industry-wide LBD measures are added, but the differences (and hence possible biases) are small.

The results presented in Table 4 provide additional evidence on how the returns to learning by doing may vary under different circumstances. Table 4 reports the estimated coefficients that are obtained by interacting one of the LBD variables with a dummy variable indicating the level of firm technology. Three dummy variables are defined to capture differences across firms in their state of employed technology. These variables are hand (equal to one if the firm uses *only* hand tools), power (equal to one if the firm uses any power machinery), and comp (equal to one if the firm has a computer). Theory predicts that the rates of learning should be faster in firms that employ more advanced technologies. Instead, the results in Table 4 suggest that there is only a complementarity between rates of learning and technology at very low levels of technology (i.e. hand tools). However, this result holds only when LBD is proxied by the variables exp and firmage; a result that makes intuitive sense. It is not surprising that firms with technologies based on traditional crafts (i.e., hand tools) have very fast rates of learning by doing. Moreover, micro-firms based on traditional technologies tend to die with the entrepreneur. This fact might explain the significance of the interactive term when learning by doing is proxied by firmage. As expected, the overall levels of productivity are much higher in firms using both power tools and computers.[5]

5 CONCLUSION

This paper examines learning by doing in the manufacturing sector of Ghana using a production function framework. The primary purpose of the study is to test two assumptions underlying the growth models which endogenize learning by doing. These two assumptions propose that: (i) the learning curve in a less-developed country will be lower than the learning curve in a more-developed country; and (ii) the rate of learning is higher in firms producing high-technology goods. The results from the production analysis provide some evidence that the learning curve in Ghana is lower than those estimated for industries in high-income countries. However, the results do not support the proposition that learning rates are higher in firms with more sophisticated technologies. In contrast, the results reveal that the interactive effect between LBD and technology is significant only at very low levels of technology. This finding contradicts the assumption made in the new growth theories that firms with high rates of learning by doing produce goods using sophisticated technologies. Nevertheless, the estimated rates of learning in Ghana are lower than those estimated for more developed countries which suggests that factors other than technology may be responsible for the slow rate of growth of Ghanaian productivity.

REFERENCES

Arrow, K. (1962). 'The economic implications of learning by doing', *Review of Economic Studies*, **29**, 155–73.

Bahk, B-H. and Gort, M. (1993). 'Decomposing learning by doing in new plants'. *Journal of Political Economy*, **101**(4), 561–583.

Bardhan, P. K. (1970). *Economic Growth, Development and Foreign Trade*. New York: Wiley–Interscience.

Caballero, R. J. and Lyons, R. K. (1992). 'External effects in US procyclical productivity', *Journal of Monetary Economics*, **29**, 209–225.

Caballero, R. J. and Lyons, R. K. (1990). 'Internal versus external economies in European industry', *European Economic Review*, **34**, 805–826.

[5]A full set of the results of the models including the interactive terms can be obtained on request from the author.

Chamley, C. (1993). 'Externalities and dynamics in models of learning or doing'. *International Economic Review*, **34**(3), 583–609.

Chuang, Y.-C. (1996). 'Identifying the sources of growth in Taiwan's manufacturing industry', *Journal of Development Studies, 32*(3), 445–463.

Clemhout, S. and Yan, H. Y. Jr., (1970): 'Learning by doing and infant industry protection', *Review of Economic Studies*, **37**, 33–56.

Foster, A. D. and Rosensweig, M. R. (1993). 'Information, learning and wage rates in low-income rural areas, *Journal of Human Resources*, **28**(4), 759–790.

Hall, R. E. (1988): 'The relationship between price and marginal cost in US industry,' *Journal of Political Economy*, **96**(3), 921-947.

Hirsh, W. Z. (1952). 'Manufacturing progress functions', *Review of Economics and Statistics*, **34**, 143–55.

Irwin, D. A. and Klenow P. J. (1994). 'Learning by doing spillovers in the semiconductor industry', *Journal of Political Economy*, **102**(6), 1201–1227.

Jarmin, R. S. (1994). 'Learning by doing and competition in the early rayon industry', *Rand Journal*, **25**(3), 441–455.

Jones, P. (1994). 'Are manufacturing workers really worth their pay', *Centre for the Study of African Economies Working Paper WPS/94–12.*

Joyanovic, B. and Nyarko, Y. (1995). 'A Bayesian learning model fitted to a variety of empirical learning curves', *Bookings Papers: Microeconomics*, 247–305.

Krugman, P. R. (1987). 'The narrow moving band, the Dutch disease, and the consequences of Mrs. Thatcher: notes on trade in the presence of scale economies', *Journal of Development Economics*, **27**, 41–55.

Levari, D. and Sheshinski, E. (1973). 'Experience and productivity in the Israel diamond industry', *Econometrica*, **41**(2), 239–253.

Lieberman, M. B. (1984). 'The learning curve and pricing in the chemical processing industries', *Rand Journal Economics*, **15**, 213–228.

Lucas, R. E., Jr. (1988). 'On the mechanics of economic development', *Journal of Monetary Economics*, **22**, 3–42.

Lucas, R. E., Jr. (1993). 'Making a miracle', *Econometrica*, **61**(2), 251–272.

Nelson, C. R. and Startz, R. (1990a). 'The distribution of the instrumental variables estimator and its *t*-ratio when the instrument is a poor one', *Journal of Business*, **LXIII**, S125–S410.

Nelson, C. R. and Startz, R. (1990b). 'Some further results on the exact small properties of the instrumental variables estimator', **LVII**, 967–976.

Preston, L. E. and Keachie, E. C. (1964). 'Cost functions and progress functions: an integration', *American Economic Review*, **54**, 100–107.

Rapping, L. A. (1965). 'Learning and World War II production functions', *Review of Economics and Statistics*, **47**, 81–86.

Romer, P. (1986). 'Increasing returns and long-run Growth', *Journal of Political Economy*, **94**, 1002–1037.

Romer, P. (1993). 'Two strategies for economic development: using ideas and producing ideas', *Proceedings of the World Bank Annual Conference on Development Economics 1992*, 63–115.

Sheshinski, E. (1967). 'Tests of the "Learning By Doing" Hypothesis', *Review of Economics and Statistics*, **49**, 568–78.

Stiglitz, J. E. (1989). 'Markets, Market failures, and economic development,' *AEA Papers and Proceedings*, **79**(2), 197–203.

Stiglitz, J. E. (1987). 'Learning to learn, localized learning and technological progress'. In Dasgupta, P. and Stoneman, P. (eds) *Economic Policy and Technological Performance*. Center for Economic Policy Research: Cambridge University Press.

Stokey, N. L. (1988). 'Learning by doing and the introduction of new goods', *Journal of Political Economy*, **96**, 701–716.

Stokey, N. L. (1991). 'Human capital, product quality, and growth', *Quarterly Journal of Economics*, **106**, 516–616.

Succar, P. (1987). The need for industrial policy in LDCs—A re-statement of the infant industry argument', *International Economic Review*, **28**, 521–534.

White, H. (1978). 'A heteroskedasticity consistent covariance matrix and a direct test for heteroskedasticity', *Econometrica.*

Young, A. (1991). 'Learning by doing and the dynamics of international trade', *Quarterly Journal of Economics*, **106**, 369–406.

Young, A. (1993). 'Invention and bounded learning by doing', *Journal of Political Economy*, **101**(3), 443–472.

APPENDIX

Appendix 1: definitions and means of variables used in productivity equations

Name	Definition	Mean
ln(VADDL)	The log of value-added divided by total labour input. Value-added is calculated as the sum of profits (i.e., the value of sales minus all variable costs) plus labour costs.	12.8209
ln(K/L)	The log of the capital-labour ration. Capital is defined as the replacement value of the total capital stock. This endogenous variable is lagged by one year.	12.9686
ln(FS)	The log of total labour input. The labour input is defined as the sum of all full-time (i.e. working 40+ hours per week) employees and apprentices. This endogenous variable is lagged by one year.	2.9181
Basic	The proportion of employees who have completed either primary school or junior secondary school (i.e. employees who have 6–9 years of schooling. This endogenous variable is lagged by one year.	0.5352
Secondary	The proportion of employees who have completed vocational school, senior secondary school, or polytechnic training (i.e., employees who have 11–15 years of schooling). This endogenous variable is lagged by one year.	0.2687
Tertiary	The proportion of employees who have completed either a university degree or professional training (i.e. employees who have 15–20 years of schooling). This endogenous variable is lagged by one year.	0.1152
ln(V_1)	The logarithm of cumulative value of a firm's output. Output was calculated from the year the firm was founded until 1993.	19.2856

Appendix 1: definitions and means of variables used in productivity equations

Name	Definition	Mean
$\ln(V_2)$	The logarithm of cumulative value of a firm's output divided by its cumulative labour input. Output was calculated from the year the firm was founded until 1993. Cumulative labour input was calculated by taking the average number of employees in the firm between 1991 and 1993 and then multiplying this number by the age of the firm.	18.8552
ln(EXP)	The logarithm of the average number of years of experience of the firm's work force. This number was calculated in two steps: (1) the average number of years of experience of workers in eight occupations was multiplied by the number of workers in each occupation, and (2) this sum was divided by the total number of workers in the firm.	1.3972
$\ln(\text{IndV}_1)$	The logarithm of aggregate industry output.	24.4949
ln(Indexp)	The logarithm of industry average experience.	5.9805
Indage	Industry average firm age.	14.2859
Firmage	The age of the firm in years.	16.6201
Food	Equals one if the firm has a SIC classification of 311–312; zero otherwise.	0.2011
Garm	Equals one if the firm has a SIC classification of 322; zero otherwise.	0.2011
Wood	Equals one if the firm has a SIC classification of 331; zero otherwise.	0.0726
Furn	Equals one if the firm has a SIC classification of 332; zero otherwise	0.2346
Metal	Equals one if the firm has a SIC classification of 381; zero otherwise.	0.2179
Mach	Equals one if the firm has a SIC classification of 382; zero otherwise.	0.0447
ln(Hrs)	The log of the firm's average number of hours worked per week.	6.6325
Union	Equals one if the firm is unionised; zero otherwise.	0.3073
Papp	The proportion of employees who are apprentices.	0.3522
Hand	Equals one if the manager reported that the firm uses only hand tools in production; zero otherwise.	0.2458
Comp	Equals one if the manager reported that the firm uses computers in production; zero otherwise.	0.0335

Appendix 1: definitions and means of variables used in productivity equations

Name	Definition	Mean
Ntasks	The level of specialisation (i.e., number of tasks) of workers in the production process. Proxied by the number of occupations (e.g., 1–18) that workers fill in the firm's organisational structure. For example, a firm with Ntasks=4 might have managers, administrative workers, supervisors, and production workers.	4.4508
Y1988	Deflated firm output in 1988 divided by 1,000,000 Cedis.	65.8841
Year93	Equals one if the data is from Wave III (i.e., 1993) of the RPED data.	0.4693