Why Do South Korean Firms Produce So Much More Output per Worker than Ghanaian Ones?

Simon Baptist
Francis Teal

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Why Do South Korean Firms Produce So Much More Output per Worker than Ghanaian Ones?

Simon Baptist
Economist Intelligence Unit, Singapore

Francis Teal
*CSAE, University of Oxford
and IZA*

ABSTRACT

Why Do South Korean Firms Produce So Much More Output per Worker than Ghanaian Ones?[*]

Macro analysis of the sources of income differences has produced very different results as to the importance of education. In this paper we investigate the roles of education and technology in explaining differences in firm level productivity across Ghana and South Korea. The labour productivity differentials across these firms exceed those implied by macro analysis. Median value-added per employee is over thirty times higher in South Korean than in Ghanaian manufacturing firms. We show that if we allow for a non-linear effect of education on output the whole of the average productivity differences across the countries can be explained. We discuss the policy implications that flow from this finding.

Corresponding author:

Francis Teal
Centre for the Study of African Economies
Department of Economics
University of Oxford
Manor Road, Oxford OX1 3UL
United Kingdom
E-mail: francis.teal@economics.ox.ac.uk

1   Introduction

In a widely cited paper Hall and Jones (1999) asked why some countries produce so much more output per worker than others. Their answer was that total factor productivity (TFP) could differ by a factor of nearly eight times across poor and rich countries and they attribute this to differences in institutions and government policy. Human capital played a relatively minor role. Using the same data the question was also posed by Caselli and Coleman (2006) and their answer was that differences could be explained by efficiency in the use of skilled labour. Rich countries got far more from their skills than poor ones. As both were exercises in calibration of a macro model we are left to argue about the plausibility of the assumptions that underlie the two exercises.

The importance of TFP as a determinant of output has also been influential in firm based studies. For example, the extensive work carried out by the World Bank on firm surveys has been used to argue that the investment climate, by which is meant a range of factors affecting the TFP of firms, "matters enormously and the relative impact of the various investment climate variables indicates where reform efforts should be directed", abstract from Escribano and Guasch (2005). The literature on firm formation and growth has also focused on the role of TFP with the probability of survival being a function of (possibly initially unknown) TFP. Aggregate productivity within the sector can grow by a process of selection if the firms with relatively high TFP are those which survive, Söderbom, Teal and Harding (2006) provide an analysis of survival of manufacturing firms in Africa. Both macro and micro data appear to have in common that much of the differences in outputs we observe cannot be explained by observable inputs, including human capital.

To link the macro with the micro analysis we ideally need to be comparing firms of a similar type across countries. Studies doing this have mainly been confined to developed countries, Bailey and Solow (2001) provide a review. Studies using comparative data across developing countries have focused more on factors affecting productivity as, for example, in the role of exports (Bigsten et at 2004, Aw and Hwang 1995, Clerides at al 1998). There has been much less work directly comparing the productivity of firms across poor countries. Such differences are indeed likely to be small relative to those that result from comparing poor with rich countries.

In this paper we propose to compare the productivity of firms in a relatively poor country, Ghana, with a relatively rich one, South Korea. More specifically we wish to answer the question set out above with the macro data: are the underlying differences in productivity the result of TFP (the finding of Hall and Jones (1999)) or can the differences be explained by the efficiency with which skilled labour is used (the finding of Caselli and Coleman (2006))?

Such a comparison is clearly difficult. How outputs and inputs are to be valued and to be made comparable across firms within a country is an important empirical issue in assessing whether returns to scale or market power is being identified. The problem of comparing firms across countries offers even greater empirical challenges. However it needs to be noted that the challenges of using macro data are greater. In this paper we begin in the next section by presenting a direct comparison of firm-level productivity across Ghanaian and South Korean firms. The productivity

differences at the firm level substantially exceed those derived from macro data for the whole economy. In the sections that follow we investigate the sources of this difference.

In section 3 we present the value-added production function from Hall and Jones (1999) that we will initially use to assess the roles of TFP and education in explaining productivity differences across the firms in the two countries. Issues of functional form, dynamics and the endogeneity of input choice, including education, are addressed in section 4. As the effect of education on productivity is central to our analysis in section 5 we carry out a range of robustness checks now focused on the effects of education on the total factor productivity differences derived from the production functions estimated in section 4. The implied earnings function from the specification is set out in section 6 so that the Mincerian return to education from using firm level data can be compared with that derived from labour market data. A final section concludes.

2   Productivity in Ghanaian and South Korean Firms

In Table 1 we present the comparison of our firm level data for Ghana and South Korea with the macro data for those two countries from Hall and Jones (1999). At the macro level labour productivity differs by a factor of seven and TFP by a factor of three. At the micro level the differences in labour productivity are very much larger. If we use the value-added measure, which is the most comparable with the macro data, median productivity differs by a factor of over thirty times. For output per employee median productivity differs by a factor of twenty. South Korea has nearly 50 times the median level of capital per employee as Ghana. Median firm size, measured by number of employees, is 22 in Ghana and 84 in South Korea.

The picture painted by the data is of very low productivity, small firms at the mid-point of the distribution for Ghana compared with very much higher productivity, larger plants in South Korea. The measures of human capital are not directly comparable between the macro and micro data as in the macro data the measure is imputed using an assumed Mincerian return to education while in the micro data it is simply the average number of years of education of the workers in the firm. In the macro data South Korea has 64 per cent more human capital while in the micro data the differential is less than half of that at 27 per cent, which is partly due to workers in Ghanaian manufacturing firms having above average levels of education. Thus our micro data suggest an even bigger challenge for any view that human capital matters for productivity differences.

In making outputs and inputs comparable both across firms within a country and across countries we proceed so that, as with the macro data, we can measure all output and input in constant price purchasing power parity prices. We do this in three stages. Firstly the variables are constructed so that the definitions are consistent across countries. Secondly, we deflate the values of outputs and inputs by firm specific domestic prices so as to render them comparable within the country. Finally, they are converted to international prices by means of the purchasing power parity price indices available from the PENN World Table (Heston et al 2001). The measure of labour input used in the regressions below is weekly hours worked. For Ghana we use the firm-level data on employment and average hours worked. In South Korea we do not have firm-level data on hours worked, and so we use the subsector (ISIC 2-digit) averages from the ILO. The result is a series of

constant price values of value-added and its components and a measure of labour inputs which provide data on comparative productivity across the manufacturing sectors of South Korea and Ghana.[1] Table 1 has already presented the key results of these calculations.

In Figure 1 we show labour productivity and capital per labour hours, where the top part of the figure uses value-added and the bottom part output in the measure of labour productivity. Both measures of productivity show the same pattern with a very clear shift up in the underlying productivity of Korean relative to Ghanaian firms, once we condition on capital per labour hour. The fitted line shown in the figure is the OLS regression line which confirms that the difference in capital per labour hour is associated with a substantial dispersion of labour productivity. As the measures of labour productivity are in logs the figure implies this shift up in the underlying productivity of Korean firms is very large, once differences in capital intensity are allowed for, of the order of 10 to 20 times. We will be more precise in the next section.

Critical for our question is whether this difference in productivity can be explained by differences in the use of human capital across the firms. For both Ghana and South Korea we have a common measure which is the average number of years of education of workers in the firm. For Ghana we have some time variation in this measure. The firm level human capital measure for Ghana is obtained from labour market surveys carried out at the same time as the firm surveys. These surveys contain information on the years of education by occupation of the worker and we have aggregated up the individual level data to a firm basis using the occupational structure of the firm. The proportion of the workforce that has completed various levels of education is used for South Korea and, from this, we impute an implied average by assuming that workers finish the level of reported education. This measure, however, has no time variation. To ensure comparability across the two countries we do not use the very little time variation which is available for the Ghana data so the measure used in the regressions is the average across the years. In the next section we turn to the specification of the production function we propose to use in the analysis.

3   The sources of productivity differences in value-added

We begin our investigation of how these very large productivity differences can be explained with the same production function as that used by Hall and Jones (1999).
(1)

$$V_{it} = K_{it}^{\alpha}(A_{it}H_{ih})^{(1-\alpha)}e^{u_{it}}$$

where V_{it} is a value-added measure of output, K_{it} is a measure of physical capital, A_{it} is a labour augmenting measure of total factor productivity, H_{it} is the amount of human capital augmented labour used in production, u_{it} is the error term. The equation imposes constant returns to scale and, given the Cobb-Douglas form of the production function, the share of capital in value-added is α. In the empirical analysis of the next section the assumption of constant returns to scale will be tested.

---

[1] More details of the calculations are given in the Appendix.

Hall and Jones (1999) link human capital to years of education by the following functional form:

(2)

$$H_{it} = e^{\phi(E_i)} L_{it}$$

where E_i is the number of years of education of workers in the labour force and L_{it} is a measure of labour hours. In empirical work the function ϕ is usually written in a form non-linear in the variables as:

(3)

$$\phi(E_i) = \delta_0 + \delta_1 E_i + \delta_2 E_i^2 + v_{it}$$

The implied value-added production function is:

(4)

$$LnV_{it} = \alpha LnK_{it} + (1 - \alpha)LnA_{it} + (1 - \alpha)\phi(E_i) + (1 - \alpha)LnL_{it} + u_{it}$$

which can be re-arranged in per capita terms as:

(5)

$$Ln\frac{V_{it}}{L_{it}} = \alpha Ln\frac{K_{it}}{L_{it}} + (1 - \alpha)LnA_{it} + (1 - \alpha)\phi(E_i) + u_{it}$$

This specification ensures that human capital acts as a shifter of the production function in the same way as TFP captured in A_{it}. Two assumptions have been made in specifying the function $\phi(E_i)$ in the macro analysis. The first is that the returns to human capital are concave, so that with higher levels of human capital the return falls, and the second is the assumption that the Mincerian returns to education are the same across countries. These assumptions ensure that as human capital expands the return falls so the net effect on underlying productivity of any increase in supply is mitigated.

From the definition in equation (2) we have

$$w_H H_{it} = w_H e^{\phi(E_i)} L_{it} = w_{L(it)} L_{it}$$

where in the second equality we move from an invariant price for human capital (w_H) to the observable price of labour (w_L) and we can write

$$log\ w_{L(it)} = log\ w_H + \phi(E_i)$$

In our empirical work above the function ϕ has been written in a form non-linear in the variables as:

(6)

$$\phi(E_{it}) = \delta_0 + \delta_1 E_i + \delta_2 E_i^2$$

So our implied Mincerian earnings function specified allowing for non-linearity in the returns to education is:

(7)

$$logw_{L(it)} = \delta_0 + \delta_1 E_i + \delta_2 E_i^2$$

and the most general form of the value-added production function is:

(8)

$$Ln\frac{V_{it}}{L_{it}} = \alpha Ln\frac{K_{it}}{L_{it}} + (1-\alpha)LnA_{it} + (1-\alpha)(\delta_0 + \delta_1 E_i + \delta_2 E_i^2) + u_{it}$$

In Table 2 we present the results of this value-added production function using OLS to establish, in terms simply of descriptive statistics, how allowing a free specification of our human capital measure, which relaxes the concavity assumption, affects the measure of the productivity differential between Ghanaian and South Korean firms. In Table 2 Column (1) we reproduce what we already know from Table 1 that the productivity differences in labour productivity defined by value-added are extremely large. In the sample used in the regression a factor of over 40 times ($e^{3.75}$). If we look across the columns of Table 2 we see that a control for physical capital reduces the differential by more than half to 13 times ($e^{2.58}$) while a control for non-linear returns on human capital has a modest effect reducing the differential to roughly 11 times ($e^{2.38}$). However in Column (4) we show that, if we allow for differential returns on human capital across the two countries, the dummy for Korea's differential productivity, which in the context of this equation is the differential for those firms employing labour with no education, has a negative point estimate and is wholly insignificant. While for Ghana the returns to education are convex those for South Korea are concave.

In the sections that follow we will investigate how far this result survives once we allow for issues with functional form, dynamics and the endogeneity of the inputs.

4   Functional Form and Endogeneity in the Production Function

The value-added production function presented in Table 2 abstracts from a large number of identification and estimation issues that have been prominent in the analysis of production functions with micro data. In order to show how we will attempt to address those issues we begin with a more general model than the one presented above. We use small letters to denote logs and write our production function as:

$$v_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \beta_e E_i + \mu_{it} + \epsilon_{it}$$

We are now explicit as to the form of the error term in the equation and constant returns to scale is not imposed. In this specification it is assumed that μ_{it} is observed by the firm but not by the econometrician while ϵ_{it} is an error term unobserved by both firm and econometrician. In their review of methods which have been developed to estimate this equation Ackerberg el al (2006) contrast the more structural approach of Olley and Pakes (1996) and Levinsohn and Petrin (2003) with that which uses the methods of dynamic panel which have evolved into the difference and system GMM estimators of Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998, 2000). The dynamic panel methods set out a general model of the form:

(9)

$$v_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \beta_e E_i + c_i + w_{it} + \epsilon_{it}$$
$$w_{it} = \rho w_{i,t-1} + \xi_{it}$$

The assumption that underlies the estimation is that both ϵ_{it} and ξ_{it} can be assumed uncorrelated with the inputs. As Ackerberg et al (2006) point out this is an economic assumption as well as an econometric one. It assumes, as is also done in the more structural approach, that the time varying, but unobserved by the econometrician, determinants of productivity can be represented by this first order autoregressive process. The economic meaning of this is that once we condition on $w_{i,t-1}$ the remaining 'innovation', ξ_{it}, is news to the firm to which it cannot react. We can now transform the equation into a form which can be estimated by either difference or system GMM.

(10)

$$v_{it} - \rho v_{i,t-1} = \beta_1 k_{it} - \beta_1 \rho k_{i,t-1} + \beta_2 l_{it} - \beta_2 \rho l_{i,t-1} + (1-\rho)E_i + (1-\rho)c_i + \xi_{it} + \epsilon_{it}$$

Where we now have an error term, $\xi_{it} + \epsilon_{it}$, which is uncorrelated with the inputs. We can allow for any possible correlation between the time invariant unobservable, $(1-\rho)c_i$, by differencing. The differenced GMM estimator then proceeds to use lagged levels as instruments for the differences. The system GMM estimator then combines this instrumented difference equation with an estimate of the levels equation where lagged differences are used as instruments.

The potential problems with these two estimators have been extensively discussed, Roodman (2005, 2009). Lagged levels may be poor instruments of differences and the validity of the differences as instruments for levels depends on the deviations from the steady state level not being correlated with the fixed effects. When we have no time variation in our education measure the effect can only be identified in the system GMM estimator so in this context the conditions for this estimator to provide consistent estimates is particularly important. Tests for instrument validity can be weak if there are 'too many' instruments, Roodman (2009), with the problem that how many is 'too many' is not easy to determine.

In the more structural approach of Olley and Pakes (1996) and Levinsohn and Petrin (2003) the procedure is to be more explicit as to the behavioural assumptions that underlie the decision making process in the firms and model μ_{it}. Olley and Pakes (1996) derive a model of μ_{it} as a function of investment flows and Levinsohn and Petrin (2003) as a function of intermediate inputs. Ackerberg et al (2006) propose an extension to these methods and provide a discussion of the assumptions underlying these estimators.

The endogeneity issue that is seen across all these methods as crucial is that there are unobserved determinants of firm output observed by the firm, but unobserved by the econometrician, and that these factors are correlated with what we do observe. Identification using these methods hinges on having panel data. Bond and Söderbom (2005) consider whether identification is possible with only cross-section data. They observe that identification without panel data hinges on why inputs vary across firms and their own investigation focuses on the possible role of adjustment costs within the context of cross-section data. As we have panel data we will use the system GMM estimator but present general dynamic models consistent with the specification suggested by the Olley and Pakes (1996) approach.

A functional form issue which has featured in the literature has been the use of gross-output or value-added specifications, Basu and Fernald (1995). We investigate a gross output specification as it is both more general than the value-added specification used above and can be used to derive an implied value-added function. Further, as Levinsohn and Petrin (2003) use intermediate inputs as a determinant of the unobservable component of productivity it seems desirable to investigate a gross output specification. We therefore compare a gross output specification with that using value-added and test whether important information is lost when a value-added specification is used.

(11)

$$o_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 i_{it} + \beta_4 m_{it} + \beta_e E_i + c_i + w_{it} + \epsilon_{it}$$
$$w_{it} = \rho w_{i,t-1} + \xi_{it}$$

We now have output, o_{it}, intermediate inputs, i_{it} and raw material inputs, m_{it}. A similar transformation to that considered above provides us with the specification we propose to estimate.

(12)
$$o_{it} - \rho o_{i,t-1} = \beta_1 k_{it} - \beta_1 \rho k_{i,t-1} + \beta_2 l_{it} - \beta_2 \rho l_{i,t-1} + \beta_3 i_{it} - \beta_4 \rho i_{i,t-1} + \beta_5 m_{it} + \beta_6 \rho m_{i,t-1}$$
$$+ (1 - \rho) E_i + (1 - \rho) c_i + \xi_{it} + \epsilon_{it}$$

This can be converted to a value-added specification under certain assumptions.
(13)
$$v_{it} = o_{it} - im_{it}$$

where we have combined intermediate inputs and raw materials for exposition purposes. If we are willing to assume that shares are constant then we can write:
(14)
$$v_{it} = o_{it} - s_{im} o_{it} = (1 - s_{im}) o_{it}$$

Then value added and output will grow at the same rate and we can move from the gross output to the value-added specification by writing for the static specification:
(15)
$$o_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 o_{it} + \beta_4 o_{it} + \beta_e E_i + c_i + \epsilon_{it}$$

We can then obtain the implicit value-added parameters by using:
(16)
$$v_{it} = (\beta_1 k_{it} + \beta_2 l_{it} + \beta_e E_i + c_i + \epsilon_{it}) / (1 - \beta_3 - \beta_4)$$

Such an approach can be compared with directly estimating the value added specification:
(17)
$$v_{it} = \beta_1' k_{it} + \beta_2' l_{it} + \beta_e' E_i$$

Finally we note that the parameters from the implicit Mincerian earnings function implied by this production function can be obtained from:
(18)
$$\beta_e' = (1 - \beta_1')(\delta_0 + \delta_1 E_i)$$

using the algebra set out in (8) above.

We now proceed to estimate both the static and dynamic form of the gross output production function for Ghana and South Korea. Tables 3 and 4 present the general dynamic and static specifications for Ghana and Tables 5 and 6 those for South Korea. Column (1) shows the OLS estimates and Column (2) the fixed effect. In the remaining columns of the tables we show the results of using a system GMM estimator. In Column (3) education is treated as exogenous endogenous. In Columns (4) and (5) attention is confined to identifying a linear impact of education and in Column (4) is treated as endogenous and in Column (5) as exogenous.

As well as tests for constant returns to scale and tests for the validity of the instruments Tables 3 and 5 show the implied long run estimates from the dynamic specification which can be compared with the results in the static specification of Table 4 and 6. Considering first the results for the dynamic specification, Tables 3 and 5. In a short panel separately identifying both dynamics and fixed effects can be problematic and that appears particularly to be the case for our Korean data where the implied long run coefficient from the dynamic specification are very imprecisely estimated compared with the static specification of Table 6. The panel dimension of the Ghana data is longer and here it has proved possible to produce significant long run coefficients from the dynamic specification in Table 3 which are broadly consistent with those for the static specification. This is important as the autocorrelation tests imply the possibility of first order autocorrelation in the levels equation as in equation (8) above. In the case of South Korea we cannot identify a dynamic production function but the tests indicate no first order autocorrelation in the differences implying that a moving average process may be present in the levels, making the instruments in the GMM estimation valid.

As Tables 4 and 6 Columns (3) - (5) show it is possible to identify the capital and other inputs for the gross output static production function for both Ghana and Korea using the system GMM estimator. Constant returns to scale are not rejected for both countries and the technology clearly differs across the two with raw material inputs far more important in the Ghana technology than is the case for Korea. In comparing columns (1) and (2) for both countries the fixed effects estimates are virtually identical, implying the time invariant unobservables are not correlated with the observables. It needs to be noted that this result imposes constant returns to scale which is rejected for the fixed effects estimators in Column (2). The Hansen test for the validity of the instruments is not rejected for any of the system GMM estimators for both countries.

Of central interest is the possibility of bias in the education parameters and the results here differ across the two countries. In Table 4 there is clear evidence for Ghana of a convex relationship for education when education is treated as exogenous, Column (1) and (3). In Column (4) education is treated as endogenous which produces a similar pattern but the parameters are no longer significant. This is open to two very different interpretations. One is that endogeneity is not an issue the other is that the instruments are not valid, we turn to the second possibility in the next section. It is also possible that the true underlying return is linear but if linearity is imposed, and we allow education to be endogenous, the result is a point estimate that is negative and wholly insignificant (not

reported). In view of this inability to identify a linear effect allowing for the possibly endogeneity of education we report in column (5) the results assuming education is exogenous where the point estimate is 0.2 and significant at the 10 per cent level. We turn below to further tests for the endogeneity of education..

In Table 6 we show results for South Korea. When education is treated as exogenous in Table 6, columns (1) and (3), there is a concave pattern to the returns to education but the quadratic term is not significant. In Columns (4) and (5) we show a linear specification where in Column (4) education is treated as endogenous and in Column (5) as exogenous. As with the non-linear specification for Ghana the result of allowing education to be endogenous is a point estimate virtually identical to that obtained when education is treated as exogenous although less precisely estimated (see Columns (4) and (5)).

5.   Robustness tests for the endogeneity of education

Central to our analysis is the wish to identify the sources of productivity differences across the two countries and assess if the differing impact of education across the two countries identified above can be given a causal interpretation. It is reasonable to doubt if the system GMM estimator used in the previous section has succeeded in showing that education can be treated as exogenous. As recent work on instruments has stressed the validity of the instrument is critical for being able to draw conclusions as to how robust are the results. One problem with the system GMM estimator is that it is hard to assess how much the instruments do shift the education variable, a concern accentuated by the fact that the parameter is identified through the levels equation as there is no time variation in the data we use. In this section we provide some robustness checks on the results of the last section.

We exploit the finding above that the parameter estimates on other inputs are not affected significantly by the inclusion of education. We thus obtain a measure of total factor productivity (TFP) as the residual from a regression excluding education so enabling us to focus on the potential endogeneity of education. So we model that measure of TFP as a function of education
(19)
$$tfp_{it} = \alpha_0 + \alpha_1 E_i + \alpha_2 E_i^2 + v_{it}$$

We will use two approaches to assess the possible importance of endogeneity bias in our results. One is a standard instrumental variable approach where we instrument both the linear and quadratic terms along lines suggested by Wooldridge (2010) and the second is a control function as outlined in Wooldridge (2007). For instruments we exploit the fact that the firm's current level of demand for educated labour is determined by past investment in capital and material inputs. As is argued by Angrist and Pischke (2009, pp 215-216) it is important to establish that separate instruments drive the two terms we wish to identify so we report the Angrist and Pischke F statistic. The control function approach avoids this problem but comes at the cost of requiring stronger assumptions.

If we denote our instrument set as $\mathbf{z}_{it}$ then the exogeneity assumption we need for our standard IV estimator is $E(\mathbf{z}'_{it}v_{it}) = 0$. This assumption that the instruments are uncorrelated with the errors in

our equation of interest is necessary for the IV estimator to be consistent. However it is not unbiased and the size of the bias depends on how good are our instruments. Recent research has emphasised the importance of tests for the strengths of the instruments and for assessing the size of the potential bias from using IV relative to that from the OLS, in summary we need to be sure the cure is not worse than the disease. How good are our instruments depends on the first stage regression which for education is:

(20)
$$E_i = \pi_0 + \mathbf{z}_{it}\boldsymbol{\pi}_1 + \varepsilon_{it}$$

The problem for our model is twofold. First to identify the quadratic term in education we need instruments that shift E_i^2 separately from the linear term and secondly that for both the instruments are 'sufficiently' powerful. In principle the first of these requirements can be met by using $\mathbf{z}_{it}^2$ and interaction terms as suggested by Wooldridge (2007). As we will see for these instruments there is a problem in identifying both the linear and non-linear terms.

An alternative is to use a control function approach which requires both an extension of the above model and an assumption as to how v_{it} and ϵ_{it} are related. The extension required is that we write the model as:

(21)
$$E(v_{it}|\mathbf{z}_{it}) = 0$$

Under this mean independent assumption, which is stronger than the zero correlation assumption above, we can use $\mathbf{z}_{it}^2$ as an instrument for E_i^2. So the IVs would be $(\mathbf{z}_{it}, \mathbf{z}_{it}^2)$ for (E_i, E_i^2).

Assumptions which allow the residual from the reduced form for education to act as a control are:

(22)
$$E(v_{it}|z_{it}, E_i) = E(v_{it}|\epsilon_{it}) = \rho_1\epsilon_{it}$$

The first of these equalities would hold if (v_{it}, u_{it}) is independent of $\mathbf{z}_{it}$ which is clearly a strong assumption. The second equality imposes linearity which is also a restrictive assumption. With these assumptions we have that

(23)
$$E(tfp_{it}|\mathbf{z}_{it,}, E_i) = \alpha_0 + \alpha_1 E_i + \alpha_2 E_i^2 + \rho_1\epsilon_{it}$$

A control function approach is immediate by replacing ϵ_{it} by $\hat{\epsilon}_{it}$. The advantage of this approach is that we do not need to instrument both the linear and the quadratic term in education. The disadvantage, clearly, is the strong assumptions entailed in (22). The results of both approaches are reported for Ghana in Table 7 and for South Korea in Table 8. For Ghana preliminary tests showed that modelling dynamics was important but this was not the case for Korea, possibly reflecting the different lengths of the panels.

In Table 7 columns (1) and (2) we confirm that we can reproduce the results from the general specification shown in Table 4 for Ghana. In columns (3) – (4) we report the results for the results of estimating the equation using a two-step GMM estimator for both the quadratic specification

(Column (3)) and the linear (Column (4)). In Column (5) we show the results of the control function approach.

On the basis of the control function specification reported in Column (5) there is no evidence for the endogeneity of education. The two-step GMM estimator reported in Table 7 Column (3) generates a significantly higher degree of convexity than the OLS results reported in Column (2), although it will be noted that the Angrist and Pischke statistics — reported as AP — give grounds for doubts as to whether the instruments have separately identified both the linear and quadratic effect. The results for a specification linear in education reported in Column (4) produce virtually identical results to those in Table 3 Column (5). In summary, what evidence we have from the instrumenting procedures used in Table 7 is that OLS understates both the average return to education and its degree of convexity.

In Table 8 we carry out a similar exercise for South Korea. In Columns (1) and (2) we again confirm that using TFP gives the same results as the general specification shown in Table 6. In Column (3) we show the two-step GMM estimates and in Column (4) we show the control function for the non-linear specification. As with Ghana we find no evidence from the control function that education is endogenous. The point estimate from the two-step GMM estimator in Column (3) at 0.16 is much higher than the OLS result in Column (2) of 0.09 but much less precisely identified which is exactly what we would anticipate if indeed education is exogenous.

We now return to the importance of allowing for non-linearity, particularly in the case of Ghana. In Figure 2 we take the predicted value of labour productivity for Ghana (from Table 4, Column (3)) and for South Korea (Table 6, Column (3)) and graph these predicted values against education. We report a regression which shows how the impact of education affects these predicted values of labour productivity in Table 9. That regression confirms what is apparent from Figure 2; namely, that the whole of the differences in underlying productivity can be explained by the different pattern of education on productivity. Indeed, for Ghana over the whole range from zero to fifteen years of education there is no evidence that education raises output despite a clear positive average effect.

6. The implied earnings function

Measures of the firm level effect of education on productivity are rare relative to the returns to education using labour market data. In this section we investigate the implied earnings function from our productivity measures and compare the results with those obtained, quite independently, from labour market data. To do this we need to move from the production function to the wages implicit in the interpretation of the underlying model given above. In Table 10 we report the implied parameters of the earnings function derived above in equation (7).

$$logw_{L(it)} = \delta_0 + \delta_1 E_i + \delta_2 E_i^2$$

The linear specification gives an average Mincerian return to education for South Korea of 0.29 per cent, some 70 per cent higher than that for Ghana of 0.17 per cent. As we have already seen, focusing on the average is very misleading - particularly for Ghana. That however is what much of

the research on the return to human capital has done when labour market data has been used. We now consider how these implied return of human capital wages in Table 10 compare with evidence from labour market surveys.

In surveys which date back to the 1970s George Psacharopoulos has presented comparative estimates of the returns to education across countries, Psacharopoulos, (1981, 1985, 1994) and Psacharopoulos and Patrinos (2002, 2004). The argument throughout these surveys has been the same and is well summarised by the abstract to Psacharopoulos (1994) 'the rate of return patterns established in earlier reviews are upheld: namely, that primary education continues to be the number one investment priority in developing countries; the returns decline by the level of schooling and the country's per capita income'. While Psacharopoulos' interpretation of the evidence has been contested, Bennell (1996, 1998), it has been very influential and his data underlies the estimates of human capital in both Hall and Jones (1999) and Caselli and Coleman (2006) and version 8 of the PENN World Tables, Feenstra (et al) (2015). The pattern of returns in our data is clearly the opposite to that claimed by Psacharopoulos. One of Bennell's objections to the argument advanced by Psacharopoulos is that the data on which his overviews are based is very unreliable. Indeed many of the estimates date from the time before there was either firm or household level surveys to use. However in the case of both Ghana and South Korea the estimates cited in the most recent survey, Psacharopoulos and Patrinos (2004, Table A2), are both based on labour market data, Jones (2001) and Ryoo et al (1993).

In the case of Ghana we have three sources of labour market data. The first is based on household surveys, (see GSS (2000, 2007)), the second is based on workers in manufacturing firms (see (http://www.csae.ox.ac.uk/datasets/ghana-rped/Ghmain.html) and the third is based on a panel labour force survey (http://www.csae.ox.ac.uk/datasets/Ghana-Tanz-UHPS/default.html). The first of these sources is used by Glewwe (1996) who draws from the second year of the Ghana Living Standards Survey (GLSS), which covered 3200 households from all regions of Ghana from October, 1988 to August, 1989. The focus of the Glewwe paper is the possibility of bias in the estimates for education due to the failure to account for ability and differences in school quality. However he does report an OLS estimate of the Mincerian return to education of 0.085 (Table 2 page 275) which he argues overstates the causal return to education. Jones (2001) uses the early rounds of the manufacturing labour force survey that was conducted in Ghana from 1992 until 2003. While the Glewwe data is based on a household sample, that used by Jones (2001) is based on workers in manufacturing firms. In spite of the different population from which the sample is drawn the point estimate is virtually identical at 0.071 (Table 2 page 71). Both Glewwe and Jones chose a linear specification. The third source is used in Falco et al (2011) where the specification chosen allows for non-linearity and the results imply a clearly convex patterns of Mincerian returns to education.

This convexity for sub-Sahara Africa (SSA) has been noted in previous work. In a study for SSA, Appleton et al (1996) show that for all the countries they survey returns increase with the level of education. Convexity also characterises much of the more recent data surveyed by Schultz (2004). Bigsten et al (2000), using the labour force data from firms across a range of African countries, which include Ghana, show that the returns are highly non-linear and convex. Söderbom and Teal

(2004) provide more detailed analysis of the Ghana data. Söderbom et al (2006) show convexity in the returns for both Kenya and Tanzania again using labour force data from manufacturing firms. A similar pattern has been found for India, Kingdon and Unni (2001) and Duraisamy (2002).

For our purposes we need to compare the return in Ghana with those for South Korea. We have access to less data for South Korea but the study by Ryoo et al (1993), used by Psacharopoulos and Patrinos (2004), provides a time series for the returns to education over the period 1974 to 1988 and implicitly allows an assessment of the shape of the earnings function. Ryoo et al (1993, Table 1) shows a steady rise in the Mincerian return to education as the level rises, the same pattern as that observed for Ghana, but with very much higher returns. Returns to both men and women, with middle school or higher, have returns between 6 and 20 per cent, although there is some indication they were falling over the period. We have not been able to find how this pattern for the 1980s compares with that for the 1990s, the period to which our data refers. There is a study covering this period, Kwack (et al) (2007), but it adopts a linear specification so cannot throw light on whether the earlier pattern of returns rising with the level has changed.

This survey of the returns to education available from the labour market literature confirms the importance of allowing for non-linearities in the return to education. In fact all the empirical literature we have surveyed for Ghana and South Korea using labour force data show a convex earnings function. Given the extent of the convexity apparent in the Ghana data it is clear that the average tells one very little of the marginal value at any point. The average Mincerian returns reported by Glewwe (1996) and Jones (2001) of 7 per cent are lower than our point estimates in Table 10 but, given the imprecision of our estimate, not significantly different. Figure 2 shows that over the whole range we have plotted the returns to South Korea are concave. However in the data we only observe average years of education of the workforce from 8 to 16 years and, over this range not only is linearity not seriously misleading but the average is lower than over the whole range where we have extrapolated beyond the observed data. In summary over the range from 8 to 16 the returns are not significantly different from the upper part of the range reported in Ryoo et al (1993).

In summary this survey of rates of return from the labour force data is wholly consistent with our findings from the firm level data of very much higher returns in South Korea than in Ghana, a finding which is also inconsistent with the argument of Psacharopoulos that returns are higher in poorer countries.

7 Conclusions

This paper set out to investigate why output per worker is so much higher for firms in South Korea than for firms in Ghana. Two firm level datasets were constructed where care was taken to ensure consistent variable definitions and appropriate deflators. Production was modeled using both a value-added and a gross-output production function where we allowed for the input coefficients and the Mincerian returns to education to differ across countries. Endogeneity was controlled using a variety of methods, including through the use of the system GMM estimator for all the inputs including education.

Our results for education can provide an insight into the differences raised in the introduction between the results of Hall and Jones (1999) and those of Caselli and Coleman (2006) which rely on calibration rather than estimation. The results are consistent with the findings of Caselli and Coleman (2006) that differences in output can be explained by differences in the efficiency with which human capital is used across poor and rich countries. In particular our findings suggest that not only do the returns on human capital differ dramatically across the two countries but so does the shape of the earnings function.

Before considering the implications of these results for policy we need to note their limitations. The advantage of the use of macro data is that it provides an overview of outcomes. Thus the generality of the (opposing) conclusions drawn by Hall and Jones (1999) and Caselli and Coleman (2006). Our results are specific to two countries and we have no reason to think they generalise. Thus our finding is that, in our two countries over this period, there are very substantial differences in the efficiency with which human capital is used. That accords with the Caselli and Coleman (2006) result. However it says nothing about the possible relative importance of TFP and the efficiency of education in other countries or indeed in other sectors.

Having noted the limitations we also note that our results fit well into wider patterns noted in micro data. The result that returns to education in Ghana are highly convex has been found in a wide range of studies using labour force data two sources of which, the GLSS surveys and the panel labour force data, have been collected quite independently of the data that underlies our results. Further the implied convex earning function which was derived in section 6 above is a result that has been observed in nearly all studies of earnings functions in poor countries. We do observe concavity in the implied earning function for South Korea. However all the Korean firms have average educational levels between 8 and 16 years of education where the return are close to being linear so little weight should be put on our ability with firm data to identify the returns at lower levels of education.

Two important policy questions flow from this general finding of convexity in SSA and our particular finding that, in the case of manufacturing firms in Ghana and South Korea, the efficiency with which human capital as measured by education is used differs dramatically across the two countries. The first is the implication noted by Heckman et al (2009) that low levels of education will have an option value as to progress to the higher level requires the earlier stage. Thus students will wish to acquire primary education not for its immediate value but for the value of the option of being able to progress through the educational system. As progress through the system is constrained the implications are stark. Most of those educated will have levels of education with no returns. The second is how misleading it is to cite averages when non-linearities are an important feature of the data. This is most readily seen from Figure 2 over the range of education from 5 to 15 years there is no increase in output for Ghana while output increases by more than four times in South Korea.

None of which addresses what the data suggests is the most important question. Why is education so much more effective at producing output in South Korea than it is in Ghana? That indeed is the

implicit question that arises from the findings of Caselli and Coleman (2006) that the efficiency of the use of education is much higher in richer than poorer countries.

Several possible explanations suggest themselves for our result. The first and most obvious is that the quality of education differs across the countries. In results not reported we confined attention simply to textile and garments and reproduced similar results to those reported for the whole of the manufacturing sector. That suggests some caution with the quality of education explanation. This is a sector in which skills, at least in garments, is low and the fact that education is so much more productive in South Korea makes it unlikely that quality can be so different. A second possibility is that human and physical capital are being combined in fundamentally different ways in the two countries which our simple Cobb-Douglas specification does not capture. It seems clear that the nearly fifty fold difference in capital per labour hour in the median South Korean firm relative to the Ghanaian reflects a radical change in the type of capital being installed. If that capital is such as to make education much more productive then we need a much more sophisticated modelling of capital than is possible with our data. To take a simple example. A machine operator in South Korea operating a robot to produce garments may have much the same education in years as a machine operate in Ghana operating an electrical sewing machine. Even allowing for the substantial differences in capital it seems more than possible that one can be five times as productive as the other conditioning on that capital. A third possibility is that the quality of the management is the key difference. Recent micro work looking at issues in firm management, surveyed in Bloom et al (2014), certainly provides some evidence that management may explain low productivity within firms. Although it would need to explain how management was linked to the uses made of educated labour.

Clearly these explanations are speculative and are not mutually exclusive. What they all have in common is the implication that assessing the return to education without a consideration of its quality, the factors with which it is being combined and the quality of the management in the firm may mislead as to the potential value of that education. In one sense we have made progress in showing that it is not, in this instance, an unknown factor termed TFP which explains differences in outcomes. However in showing that it is the efficiency with which education impacts on productivity pushes the analysis only one step forward. We need to know why and we do not know the answer to that question.

References

Ackeberg, D.A., Caves, K., and Frazer, G. (2006) 'Structural Identification of Production Functions'. UCLA Working Paper.

Angrist, J.D. and J-S Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's companion*, Princeton University Press.

Appleton, S., Hoddinott, J. and J. Mackinnon (1996) 'Education and health in sub-Saharan Africa' *Journal of International Development* Vol. 8, No. 3,307-339.

Arellano, M. and Bond, S. (1991) 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations'. *Review of Economic Studies*, 58:277–297.

Arellano, M. and O. Bover (1995) 'Another look at the instrumental variable estimation of error-component models', *Journal of Econometrics*, vol. 68, issue 1, pages 29-51

Aw, B. and Hwang, A. (1995) 'Productivity and the Export Market: A Firm-level Analysis'. *Journal of Development Economics*, 47:313-332

Baily, M. N. and R. M. Solow (2001) 'International Productivity Comparisons Built from the Firm Level', *Journal of Economic Perspectives*, Vol. 15, No. 3, Summer: 151–172.

Basu, S. and Fernald, J. G. (1995) 'Are apparent productive spillovers a figment of specification error?', *Journal of Monetary Economics*, 36:165–188.

Bennell, P. (1996) 'Rates of Return to Education: Does the Conventional Pattern Prevail in sub-Saharan Africa?' *World Development*, Vol. 24, No. 1, pp. 183-199.

Bennell, P. (1998) 'Rates of Return to Education in Asia: A Review of the Evidence' *Education Economics,* Vol. 6, No. 2.

Bigsten, A., Collier, P., Dercon, S., Fafchamps, M., Gauthier, B. Willem Gunning, J.W., Isaksson, A., Oduro, A., Oostendorp, R., Pattillo, C., Söderbom,M., Teal, F., Zeufack, A. and S. Appleton (2000) 'Rates of return on physical and human capital in Africa's manufacturing sector.' *Economic Development and Cultural Change*, Vol. 48 (4) pp. 801-827.

Bigsten, A., Collier, P., Dercon, S., Fafchamps, M., Gauthier, B., Gunning, J. W., Oduro, A., Oostendorp, R., Pattillo, C., Söderbom, M., Teal, F., and Zeufack, A. (2004) 'Do African Manufacturing Firms Learn from Exporting?' *The Journal of Development Studies*, 40(3):115–141.

Bloom, N., Lemos, R., Sadun, R., Scur, D. and J. Van Reenen (2014) 'The new empirical economics of management', *Journal of the European Economic Association*, Volume 12, Issue 4, pages 835–876, August.

Blundell, R. and S. Bond (1998) 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics* 87, 115-143.

Blundell, R. and S. Bond (2000) 'GMM estimation with persistent panel data: an application to production functions', *Econometric Reviews* 19, 321-340.

Bond, S. and M. Söderbom (2005) 'Adjustment costs and identification of Cobb Douglas production functions', The Institute of Fiscal Studies, WP05/04.

Caselli, F. and W.J. Coleman (2006) 'The World Technology Frontier'. *American Economic Review*, 96(3):499-522.

Clerides, S., Lach, S., and J. Tybout (1998) 'Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico and Morocco'. *Quarterly Journal of Economics*, 113:903-947.

Duraisamy, P. (2002) 'Changes in returns to education in India, 1983-94: by gender, age-cohort and location'. *Economics of Education Review*, 21: 609-622.

Escribano, A. and J. Guasch (2005) 'Assessing the Impact of the Investment Climate on Productivity using Firm-Level Data: Methodology and the Cases of Guatemala, Honduras, and Nicaragua', World Bank Policy Research Working Paper 3621.

Falco, P., Kerr, A. Rankin, N., Sandefur, J. and F. Teal (2011) 'The returns to formality and informality in urban Africa' *Labour Economics,* Volume 18, Supplement 1, December, Pages S23–S31.

Feenstra, R. C., Inklaar, R. and M. P. Timmer (2015) 'The Next Generation of the Penn World Table', forthcoming *American Economic Review,* available for download at www.ggdc.net/pwt.

GSS (Ghana Statistical Service) (2000) *Poverty Trends in Ghana in the 1990s*, Accra, Ghana.

GSS (Ghana Statistical Service) (2007) *Pattern and Trends of Poverty in Ghana: 1991-2006*, Accra, Ghana.

Glewwe, P. (1996) 'The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment' *Journal of Development Economics* Vol. 51 (1996) 267-290

Hall, R. E. and Jones, C. I. (1999) 'Why Do Some Countries Produce So Much More Output Per Worker Than Others?' *Quarterly Journal of Economics*, 114(1):83–116.

Hallward-Driemeier, M. (2001) 'Firm-Level Survey Provides Data on Asia's Corporate Crisis and Recovery'. World Bank Working Paper Series, 2515.

Hallward-Driemeier, M., Iarossi, G., and K. L. Sokoloff (2002) 'Exports and Manufacturing Productivity in East Asia: A Comparative Analysis with Firm-Level Data'. NBER Working Paper Series, 8894.

Heckman, J.J., L.J. Lochner and P.E. Todd (2009) 'Earnings functions, rates of return, and treatment effects: the Mincer Equation and Beyond', Chapter 7 in *Handbook of the Economics of Education*, Vol. 1 (eds.) E.A. Hanushek and F. Welch, North-Holland, Amsterdam.

Heston, A., Summers, R., and Aten, B. (2001). Penn World Table Version 6.1.

Jones, P. (2001) 'Are educated workers really more productive', *Journal of Development Economics*, Vol.64, pp. 57-79.

Kingdon, G. G. and Unni, J. (2001) 'Education and women's labour market outcomes in India'. *Education Economics*, 9: 173-194.

Kwack, S. Y., Lee, Y. S. and H. Choi (2007) 'Gender Earnings gaps and return to schooling in Korea', *Journal of Economic Research*, 12, pp.79-101.

Levinsohn, J., and A. Petrin (2003) 'Estimating production functions using inputs to control for unobservables'. *Review of Economic Studies* 70: 317–342.

Olley, G. S., and A. Pakes (1996) 'The dynamics of productivity in the telecommunications equipment industry'. *Econometrica* 64: 1263–1297.

Patrinos, H. A. and G. Psacharopoulos (2011) 'Education: past, present and future global challenges' Policy Research Working Paper Series 5616, The World Bank.

Psacharopoulos, G., (1981) 'Returns to Education: an updated international comparison', *Comparative Education* Volume 17 No. 3 1981, 321-341.

Psacharopoulos, G., (1985) 'Returns to education: a further international update and implication,' *The Journal of Human Resources*, Vol. 20, No. 4 (1985), pp. 583-597.

Psacharopoulos, G., (1994) 'Returns to Investment in Education: A Global Update', *World Development*, Vol. 22, No 9, pp. 132.5-l 343, 1994.

Psacharopoulos, G., and H. A. Patrinos (2002) 'Returns to investment in education: a further update,' Policy Research Working Paper Series 2881, The World Bank. Published as:
Psacharopoulos, G. and H. A. Patrinos (2004) 'Returns to investment in education: a further update,' *Education Economics*, vol. 12(2), pages 111-134.

Roodman, D. (2005) 'xtabond2: Stata module to extend xtabond dynamic panel data estimator'. Center for Global Development, Washington.

Roodman, D. (2009) 'Practitioners' Corner: A Note on the Theme of Too Many Instruments' *Oxford Bulletin of Economics and Statistics*, 71, 1, pp. 135-158.

Ryoo, J.K., Y.S. Nam and M. Carnoy (1993) 'Changing rates of return to education over time: A Korean case study', *Economics of Education Review*, Vol. 12, No.1, pp. 71-80.

Schultz, T. P. (2004) 'Evidence of returns to schooling in Africa from household surveys: Monitoring and restructuring the market for education', *Journal of African Economies*, Vol. 13, AERC Supplement 2, pp. ii95-ii148.

Söderbom, M. and F. Teal (2004) 'Size and Efficiency in African Manufacturing Firms: Evidence from Firm-level Panel Data'. *Journal of Development Economics*, 73:369–394.

Söderbom, M., Teal, F. and A. Harding (2006) 'The determinants of survival among African manufacturing firms', *Economic Development and Cultural Change*, Vol. 54, No. 3, April 2006, pp. 533-555.

Söderbom, M., Teal, F., Wambugu, A. and G. Kahyarar (2006) "The dynamics of returns to education in Kenyan and Tanzanian manufacturing", *Oxford Bulletin of Economics and Stat*istics, 2006, Vol. 68, No.3, pp. 261-288.

Wooldridge, J. M. (2007) 'What's new in econometrics Lecture 6 Control functions and related methods' at http://www.nber.org/WNE/lect_6_controlfuncs.pdf

Wooldridge, J. M. (2010) *Econometric Analysis of Cross-Section and Panel Data*, The MIT Press, Second Edition.

Table 1   **Summary Statistics for Productivity in Firms in Ghana and South Korea**

	Ghana		South Korea	
	Mean	Median	Mean	Median
Firm level data				
Firm size (Number of Employees)	73 (156.4)	22	122 (125.6)	84
Hours Worked per Week	46.1 (8.8)	45	46.4 (1.6)	46.2
Value-added per Employee	7.7 (16.9)	3.4	190.5 (294.5)	116.6
Output per Employee	20.2 (34.8)	9.8	288.5 (350.2)	192.8
Capital per Employee	7.4 (28.5)	1.3	114.3 (207.1)	61.2
Capital to Output	0.512 (1.4)	0.105	0.508 (0.9)	0.316
Years of Education	10.0 (1.9)	10.0	12.7 (1.2)	12.7
Number of observations	1774		907	
Macro data				
Labour Productivity	0.052		0.380	
Capital to Output	0.567		0.880	
Human Capital per Worker	0.464		0.761	
Total Factor Productivity	0.198		0.568	

Note: The monetary values for the firm data are all expressed in '000 1996 PPP$. Standard deviations are in parentheses. The macro data are from Hall and Jones (1999) and are benchmarked with reference to the US.

Table 2   Pooled OLS Regressions
Dependent Variable Ln (Value-added per labour hour)

	(1)	(2)	(3)	(4)
Ln (Capital per labour hour)		0.31***	0.28***	0.27***
		(0.025)	(0.028)	(0.029)
Education (in years)			-0.45***	-0.60***
			(0.087)	(0.133)
Education (in years)^2 (a)			2.57***	3.5***
			(0.425)	(0.749)
Korea dummy	3.75***	2.58***	2.38***	-3.74
	(0.109)	(0.131)	(0.128)	(2.405)
Education*Korea Dummy				1.09***
				(0.403)
Education^2*Korea Dummy				-4.82***
				(1.72)
Constant	-2.80***	-3.86***	-1.97***	-1.32**
	(0.112)	(0.126)	(0.447)	(0.602)
Observations	2,681	2,681	2,681	2,681
R-squared	0.67	0.75	0.75	0.76

Robust standard errors in parentheses   *** p<0.01, ** p<0.05, * p<0.1

(a) The Education squared variable is divided by 100 in this and the following tables to enhance the readability of the parameter values.

Note: Time dummies are included in all regressions in this and the following tables but not reported.

Table 3 The Dynamic Output Production Function for Ghana: OLS, FE and System GMM
Dependent variable: Ln(Output per labour hour)

	(1)	(2)	(3)	(4)	(5)
			System-GMM	System-GMM	System-GMM
	OLS	FE	Exogenous education	Endogenous education	Exogenous education
Ln (Output per labour hour)$_{i,t-1}$	0.61*** (0.031)	0.33*** (0.031)	0.57*** (0.070)	0.55*** (0.066)	0.57*** (0.070)
Ln (Capital per labour hour)$_{it}$	0.06*** (0.015)	0.05*** (0.015)	0.03 (0.034)	0.04 (0.034)	0.04 (0.034)
Ln (Capital per labour hour)$_{I,t-1}$	-0.05*** (0.014)	-0.04*** (0.013)	-0.02 (0.029)	-0.03 (0.029)	-0.02 (0.030)
Ln (Indirect costs per labour hour)$_{it}$	0.16*** (0.028)	0.12*** (0.019)	0.11** (0.054)	0.09** (0.040)	0.11** (0.055)
Ln (Indirect costs per labour hour)$_{I,t-1}$	-0.09*** (0.021)	-0.04*** (0.012)	-0.06 (0.044)	-0.03 (0.032)	-0.06 (0.045)
Ln (Material costs per labour hour)$_{it}$	0.61*** (0.040)	0.66*** (0.025)	0.66*** (0.040)	0.69*** (0.038)	0.66*** (0.040)
Ln (Material costs per labour hour)$_{i,t-1}$	-0.34*** (0.039)	-0.18*** (0.025)	-0.34*** (0.060)	-0.34*** (0.055)	-0.35*** (0.056)
Education (in years)$_i$	-0.04* (0.021)		-0.03 (0.030)	-0.14 (0.096)	0.02** (0.009)
Education (in years)^2_i	0.25** (0.118)		0.28 (0.179)	0.83 (0.516)	
Constant	0.82*** (0.144)	1.25*** (0.100)	0.71*** (0.237)	1.23*** (0.469)	0.48*** (0.163)
Observations	1,598	1,598	1,598	1,598	1,598
R-squared	0.94	0.84			
Number of firms		239	239	239	239
AR1			0.00	0.00	0.00
AR2			0.0654	0.0457	0.0655
Sargan P			0.557	0.369	0.466
Hansen P			0.315	0.416	0.313
CRS (P value of F test)	0.04	0.13	0.85	0.57	0.91
Long run parameters					
Capital	0.03* (0.02)	0.004 (0.017)	0.027 (0.038)	0.023 (0.040)	0.033 (0.038)
Indirect costs	0.18*** (0.03)	0.11*** (0.023)	0.121 (0.050)**	0.121 (0.040)***	0.128 (0.048)**
Material costs	0.68*** (0.04)	0.72*** (0.026)	0.753 (0.059)***	0.786 (0.052)***	0.753 (0.061)***

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4 The Static Output Production Function for Ghana: OLS, FE and System GMM
Dependent variable: Ln(Output per labour hour)

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	System-GMM Exogenous education	System-GMM Endogenous education	System-GMM Exogenous education
Ln (Capital per labour hour)$_{it}$	0.03*** (0.012)	0.03* (0.014)	0.04 (0.024)	0.06** (0.028)	0.05** (0.023)
Ln (Indirect costs per labour hour)$_{it}$	0.18*** (0.018)	0.15*** (0.019)	0.15*** (0.032)	0.13*** (0.029)	0.15*** (0.032)
Ln (Material costs per labour hour)$_{it}$	0.65*** (0.024)	0.64*** (0.026)	0.67*** (0.048)	0.67*** (0.045)	0.65*** (0.051)
Education (in years)$_i$	-0.09** (0.039)		-0.12** (0.056)	-0.07 (0.241)	0.02* (0.023)
Education (in years)^2_i	0.62*** (0.222)		0.81** (0.330)	0.37 (1.236)	
Constant	2.12*** (0.217)	1.94*** (0.097)	2.17*** (0.389)	2.10* (1.165)	1.58*** (0.217)
Observations	1,880	1,880	1,880	1,880	1,880
R-squared	0.91	0.80			
Number of firms		254	254	254	254
AR1			0.00	0.00	0.00
AR2			0.00	0.00	0.00
Sargan P			0.00	0.00	0.00
Hansen P			0.651	0.561	0.702
CRS (P value of F test)	0.0793	0.0091	0.7835	0.6273	0.6202

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5 The Dynamic Output Production Function for South Korea: OLS, FE and System GMM
Dependent variable: Ln(Output per labour hour)

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	System-GMM Exogenous education	System-GMM Endogenous education	System-GMM Exogenous education
Ln (Output per labour hour)$_{i,t-1}$	0.93***	-0.11	0.88***	0.87***	0.95***
	(0.030)	(0.152)	(0.191)	(0.186)	(0.212)
Ln (Capital per labour hour)$_{it}$	0.10**	0.10	0.07	0.07	0.08
	(0.042)	(0.069)	(0.069)	(0.076)	(0.068)
Ln (Capital per labour hour)$_{I,t-1}$	-0.09**	-0.03	-0.14***	-0.14***	-0.13***
	(0.041)	(0.055)	(0.042)	(0.042)	(0.040)
Ln (Indirect costs labour hour)$_{it}$	0.20***	0.25***	0.22***	0.22***	0.22***
	(0.033)	(0.049)	(0.045)	(0.045)	(0.045)
Ln (Indirect costs labour hour)$_{I,t-1}$	-0.18***	0.04	-0.15*	-0.15*	-0.17**
	(0.035)	(0.059)	(0.076)	(0.075)	(0.077)
Ln (Material costs labour hour)$_{it}$	0.25***	0.23***	0.27***	0.27***	0.25***
	(0.044)	(0.058)	(0.073)	(0.077)	(0.073)
Ln (Material costs labour hour)$_{I,t-1}$	-0.22***	0.10	-0.21**	-0.20**	-0.23**
	(0.044)	(0.108)	(0.097)	(0.098)	(0.103)
Education (in years)$_i$	0.16		0.25	0.03	0.01
	(0.110)		(0.164)	(0.124)	(0.023)
Education (in years)^2_i	-0.61		-0.94		
	(0.431)		(0.611)		
Constant	-0.81	5.48***	-0.80	0.63	0.49
	(0.698)	(1.180)	(0.941)	(1.37)	(0.747)
Observations	560	560	560	560	560
R-squared	0.90	0.58			
Number of firm		320	320	320	320
Sargan P			0.00866	0.005	0.0319
Hansen P			0.142	0.104	0.532
CRS (P value of F test)	0.8930	0.0008	0.7350	0.6021	0.6160
Long run parameters					
Capital	0.250	0.066	-0.558	-0.461	-0.465
	(0.217)	(0.086)	(1.256)	(1.004)	(1.058)
Indirect costs	0.214	0.260***	0.582	0.521	0.532
	(0.177)	(0.087)	(1.117)	(0.917)	(0.973)
Material costs	0.459***	0.299	0.530	0.493	0.508
	(0.167)	(0.101)	(0.911)	(0.820)	(0.837)

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 6 The Static Output Production Function for South Korea: OLS, FE and System GMM
Dependent variable: Ln(Output per labour hour)

VARIABLES	(1) OLS	(2) FE	(3) System-GMM Exogenous education	(4) System-GMM Endogenous education	(5) System-GMM Exogenous education
Ln (Capital per labour hour)$_{it}$	0.11*** (0.028)	0.08* (0.044)	0.13 (0.080)	0.12 (0.094)	0.13 (0.084)
Ln (Indirect costs per labour hour)$_{it}$	0.22*** (0.030)	0.20*** (0.035)	0.14** (0.068)	0.15** (0.065)	0.14** (0.069)
Ln (Material costs per labour hour)$_{it}$	0.25*** (0.037)	0.25*** (0.044)	0.38*** (0.098)	0.38*** (0.099)	0.39*** (0.099)
Education (in years)$_i$	0.47* (0.253)		0.48** (0.239)	0.12 (0.227)	0.10*** (0.026)
Education (in years)^2_i	-1.41 (1.035)		-1.49 (0.974)		
Constant	1.14 (1.571)	5.39*** (0.413)	0.76 (1.600)	2.86 (2.427)	3.07*** (0.636)
Observations	918	918	918	918	918
R-squared	0.58	0.56			
Number of firm		357	357	357	357
AR1			0.410	0.404	0.413
AR2					
Sargan P			0.00319	0.000411	0.00112
Hansen p			0.740	0.447	0.658
CRS (P value of F test)	0.3159	0.0000	0.1707	0.2144	0.1511

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 7 TFP and Education in Ghana: Dependent Variable Total Factor Productivity (TFP)

VARIABLES	(1) OLS Static	(2) OLS Dynamic	(3) GMM 2 step	(4) GMM 2 step	(5) Control Function
$TFP_{i,t-1}$		0.60***	0.54***	0.53***	
		(0.032)	(0.058)	(0.058)	
Education (in years)$_i$	-0.11***	-0.05***	-0.34**	0.02**	-0.12***
	(0.034)	(0.018)	(0.142)	(0.010)	(0.042)
Education (in years)^2_i	0.69***	0.36***	1.82***		0.75***
	(0.198)	(0.110)	(0.707)		(0.206)
Control for education					-0.00
					(0.024)
Constant	0.36**	0.16**	1.54**	-0.23**	0.37
	(0.141)	(0.076)	(0.693)	(0.099)	(0.260)
Observations	1,880	1,598	1,342	1,342	1,360
R-squared	0.02	0.35	0.33	0.39	0.03
Long run parameters					
Education (in years)$_i$		-0.14**	-0.74*	0.05**	
		(0.044)	(0.303)	(0.021)	
Education (in years)^2_i		0.89***	3.93***		
		(0.261)	(1.50)		
First stage F (P value) for Education			87.40	76.38	
			(0.00)	(0.00)	
AP Adjusted F (P value) for Education (b)			0.69	200.56	
			(0.56)	(0.00)	
First stage F (P value) for Education^2			104.68		
			(0.00)		
AP Adjusted F (P value) for Education^2			0.72		
			(0.54)		
First stage F for $TFP_{i,t-1}$			48.57	76.38	
			(0.00)	(0.00)	
AP Adjusted F (P value) for $TFP_{i,t-1}$ (b)			75.69	109.09	
			(0.00)	(0.00)	
Hansen J statistic P			0.8920	0.9860	
Instruments used			Lagged 2 Ln (capital per labour hour) and its quadratic First difference of lagged tfp Lagged 2 Ln (capital interacted with material per labour hour) and its quadratic)	Lagged 2 Ln (capital per labour hour) and its quadratic First difference of lagged tfp	Lagged 2 Ln (capital per labour hour) and its quadratic

Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1

Table 8 TFP and Education in South Korea
Dependent Variable: Total Factor Productivity (TFP)

VARIABLES	(1) OLS Static Non-Linear	(2) OLS Static Linear	(3) GMM 2 step (a)	(4) Control Function
Education (in years)$_i$	0.46*	0.09***	0.16*	0.57*
	(0.237)	(0.022)	(0.096)	(0.342)
Education (in years)^2_i	-1.45			-1.90
	(0.961)			(1.353)
Control for education				0.01
				(0.078)
Constant	-3.47**	-1.19***	-2.08*	-4.12*
	(1.454)	(0.272)	(1.216)	(2.246)
Observations	918	918	241	241
R-squared	0.05	0.04	0.004	0.04
First stage F (P value) or Education			9.30 (0.00)	
Hansen J statistic P value			0.1462	
Instruments			Lagged 2 Ln (capital per labour hour) and its quadratic Lagged 2 Ln (materials per labour hour) and its quadratic	Lagged 2 Ln (capital per labour hour) and its quadratic
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Table 9   Predicted Ln (Output per Labour Hour)

	(1)	(2)	(3)
Korea Dummy	3.07***	2.40***	-5.17**
	(0.081)	(0.107)	(2.263)
Education		-0.68***	-1.14***
		(0.127)	(0.216)
Education^2		4.23***	6.91***
		(0.609)	(1.177)
Education* Korea Dummy			1.57***
			(0.395)
Education^2* Korea Dummy			-7.85***
			(1.782)
Constant	5.33***	7.73***	9.54***
	(0.073)	(0.683)	(0.992)
Observations	2,798	2,798	2,798
R-squared	0.64	0.68	0.70

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Note: See Figure 2 for the basis of this regression.

Table 10   The Implied Earnings function with education treated as an exogenous variable

$$logw_{L(it)} = \delta_0 + \delta_1 E_i + \delta_2 E_i^2$$

VARIABLES	(1) Ghana Non-Linear	(2) Ghana Linear	(3) Korea Non-Linear	(4) Korea Linear
Education (in years)$_i$	-.801**	0.17*	1.354*	0.29***
	(0.25)	(0.085)	(0.788)	(0.078)
Education (in years)^2_i	5.41***		-4.25	
	(1.38)		(3.10)	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Column (1) is taken from Table 4 Column (3)
Column (2) is taken from Table 4 Column (5)
Column (3) is taken from Table 6 Column (3)
Column (4) is taken from Table 6 Column (5)

Data Appendix

The data used are a balanced panel of 863 South Korean manufacturing firms observed for 3 years and an unbalanced panel of 312 Ghanaian manufacturing firms observed for 12 years. The Ghanaian data was collected in a series of interviews with firm management and cover the period 1991-2002. Along with the survey questionnaire, this data is publicly available from the Centre for the Study of African Economics at the University of Oxford. The data set (including definitions and variable construction) is that used in Soderbom and Teal (2004) with the addition of the more recent survey rounds. The data used to estimate the production function are the value of physical capital stocks, number of employee-hours (number of employees multiplied by average weekly hours), expenditure on materials and other inputs (mostly electricity, other energy and rented land, buildings and equipment) and average firm worker years of education. There are also price indexes for output and material inputs used to convert the variables into real 1991 domestic currency prices. Gross output is measured as total sales, adjusted by changes in inventories.

The South Korean data were collected in face-to-face interviews covering 1996-1998, and are described in Hallward-Driemeier (2001). The data and the survey questionnaire are publicly available from the World Bank. Output is calculated as sales plus the change in inventories, with other production variables being value of physical capital stock, number of employees and expenditure on materials, electricity, other energy and rented land, equipment and buildings. These final three inputs are aggregated into a single factor so as to be comparable with the Ghanaian data. Again, there are firm-level price indexes for output and material inputs used to convert the data into real 1996 domestic currency. Where these price indexes were not available the sectoral averages were used. The variables in the South Korean dataset were transformed into ratios and truncated before they were made publicly available, and the recovery of the levels of the variables from these transformations will have introduced some additional measurement error into the data. Firms with more than 500 employees or within the top 5% for asset value were truncated and so these have been dropped from the sample. This means that the South Korean sample contains more small firms than the true firm population which, if anything, should reduce any apparent differences with the smaller Ghanaian firms. Average hours worked by sector were obtained from the ILO and used in the construction of the employee-hour variable.

For both Ghana and South Korea we have firm level price deflators for output and some of the inputs. We have used the primary data from the surveys to obtain measures of outputs and inputs classified as raw material inputs and "other inputs" which are primarily indirect costs (electricity, other energy and rent). We have firm level deflators for both output and raw material inputs but not for indirect costs. We proceeded by deflating output and raw material prices by these firm-level deflators and then converting these values to PPP numbers by using the PPP deflator for the base year.[2] For indirect costs, where no appropriate firm-level deflator was available, we have used the PPP consumption deflator for each year. These deflators render the data comparable over time within the country and, as they are firm based, ensure that the variables measure changes in real

[2] In cases where these firm level price indices were not reported they were assumed to be equal to the average of firms in the same manufacturing sub-sector.

quantities (at least for output and raw materials) and not higher revenues or lower costs associated with market power.

In order to establish how much of this labour productivity differential can be explained by factor intensity we also need to impute a value to the capital stock for each firm in each country. For the South Korean firms, we take the firm-level reported value of the nominal capital stock for each year and deflate it using the annual PPP investment deflator. For the Ghanaian firms, as we have a longer panel available and due to concerns about capital valuation in this less developed market, we aggregate the deflated value of the investment series and assume a depreciation rate rather than using the reported values of the capital stock.

PPP deflators from PWT 6.1 were used to convert the real domestic quantities in each country into international quantities. The PPP investment deflator was used for the capital stock, the consumption deflator for indirect and material inputs, and the GDP deflator for output.

**Figure 1 Productivity Differences Between South Korean and Ghanaian Firms
(Value-added , output and capital per worker are expressed in 1996$PPP)**
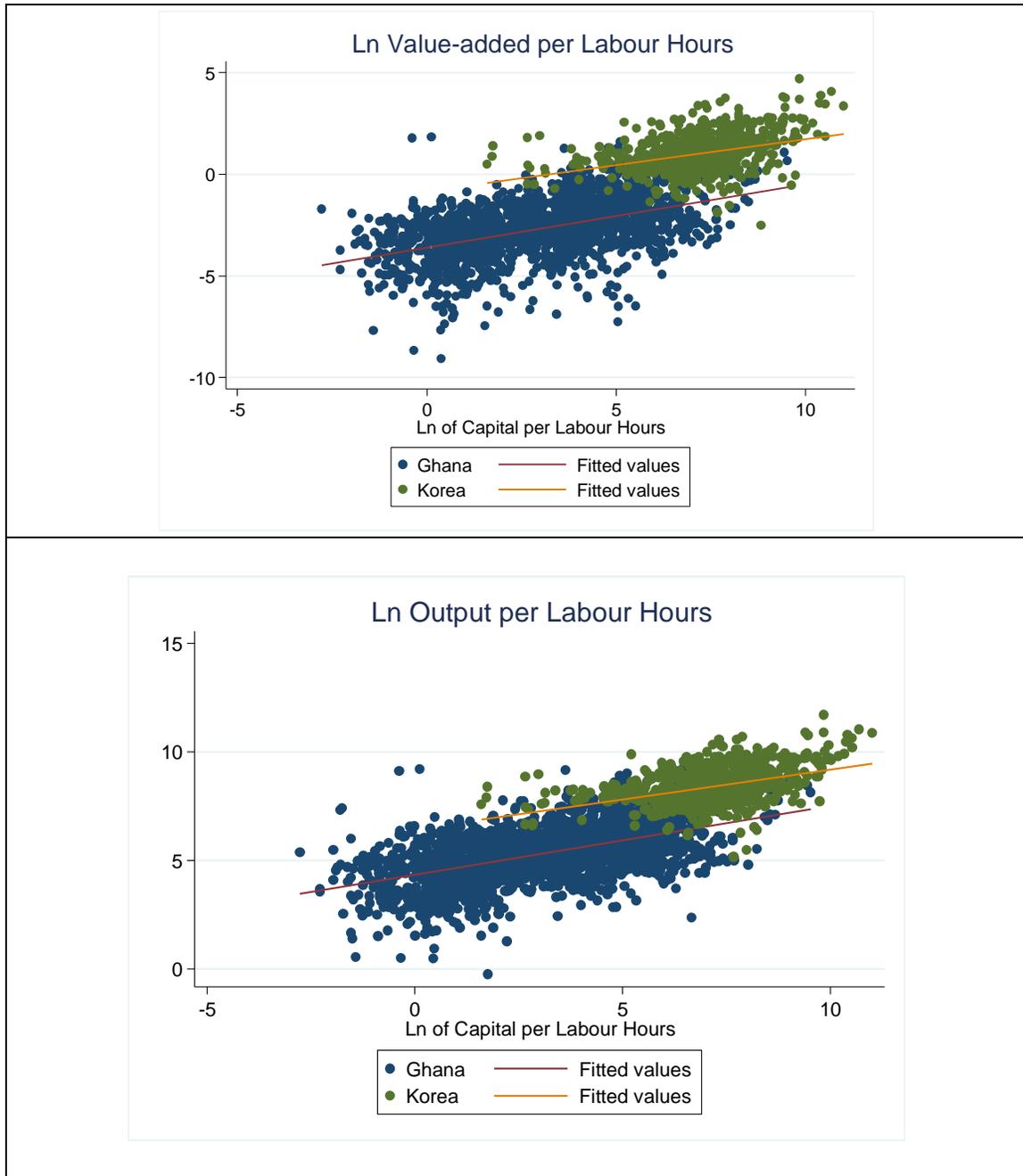
Figure 2   The Impact of Education on Productivity

Predicted Ln Output per Labour Hour