Education Externalities in Rural Ethiopia: Evidence from Average and Stochastic Frontier Production Functions

Sharada Weir and John Knight

Centre for the Study of African Economies
Department of Economics, University of Oxford

Manor Road Building, Oxford OX1 3UQ, United Kingdom

Office: +44 1865 271 084
Fax: +44 1865 281 447

sharada.weir@Economics.ox.ac.uk

CSAE WPS/2000.4

March 2000

Abstract Education will have externality effects in agriculture if, in the course of conducting their own private economic activities, educated farmers raise the productivity of uneducated farmers with whom they come into contact. This paper seeks to determine the potential size and source of such benefits for rural areas of Ethiopia. Average and stochastic frontier production function methodologies are employed to measure productivity and efficiency of farmers. In each case, internal and external returns to schooling are compared. We find that there are substantial and significant externality benefits of education in terms of higher average farm output and a shifting outwards of the production frontier. External benefits of schooling may be several times as high as internal benefits in this regard. However, we are unable to find any evidence of externality benefits to schooling in terms of improvements in technological efficiency in the use of a given technology. This suggests that the source of externalities to schooling is in the adoption and spread of innovations, which shift out the production frontier.

Acknowledgements The UK Department for International Development (DFID), which supports policies, programmes and projects to promote international development, provided funds for this study as part of that objective. The views and opinions expressed are those of the authors alone. We thank Marcel Fafchamps and participants at a CSAE seminar for useful comments; the usual disclaimer applies.

JEL Classification: D62, I2, O12, Q12

1. Introduction

The potential benefits of schooling are manifold. They may accrue to the individuals receiving education, to their households, and to their community or society. They may be internalised by the decision-taker(s) or not taken into account in the decisions. To the extent that individuals or households fully perceive the benefits of schooling to themselves and there are no constraints on schooling, they will invest in the amount of education that maximises their private utility. However, there will be a socially sub-optimal level of investment in the presence of positive externalities.

Externalities are generally defined to include the effects of economic actions which are not taken into consideration by the economic agent. These external effects may be negative or positive. Economic agents may not be aware of the effects of their actions on others or they may not care about them. Even if external benefits or costs are recognised by the economic agents involved, there is often no market for the externality. That is, those who are affected by the activities of others are usually able neither to pay them to produce more of a positive externality nor to impose on them the full social costs of a negative externality.

In considering external benefits, two avenues suggest themselves. First, education may have externality effects if educated people raise the productivity of others, e.g., early adopters of innovations may be educated farmers who demonstrate the value of new inputs and techniques to less educated individuals. Schooling can have positive externalities if the friends or neighbours of an educated person are able to share in the benefits of schooling (for example, through the informed choices made by educated people which may be mimicked by those without education) without financing the schooling themselves. Second, education, particularly of girls, may be associated with improved health and nutrition, and with reduced fertility. While these avenues have each been explored in parts of the developing world, research is needed in order to assess the applicability of findings in an African context (Appleton and Balihuta 1996). Furthermore, the mechanisms by which externalities operate are not well-understood in any context.

This paper concentrates on the first avenue of research: the externality link between education and productivity in agriculture. We employ average and stochastic frontier production function methodology to show that there are externality effects of education upon farm productivity in Ethiopia. Our evidence will suggest that the external benefits of education operate by pushing out the production frontier, but that private education is required to enable farmers to become efficient in the use of a given technology.

It is difficult to devise conclusive tests of the externality benefits of schooling for household agricultural production. The correlation that we seek between education in the community (which we refer to as the site) and individual farm productivity may be non-causal, for instance being mediated by a third factor, or causation may be reversed. The natural resources and fertility of the site may have permitted and encouraged prior private investment in education. In that case, the current educational levels of individual farmers and the productivity of farms in the site may be positively correlated despite a lack of educational externalities. The paper has to examine not only the evidence consistent with our main hypothesis - that externalities of schooling in agriculture exist - but also alternative explanations of the results.

Research on the economic benefits of schooling in Ethiopia is timely. The country has one of the lowest school enrolment rates in Africa. It is massively under-researched in relation to the size of the aid effort and in relation to other African countries. The government faces extreme budgetary constraints, and difficult expenditure choices must be made at the margin. There is an urgent need to discover the full consequences of the country's extremely low rates of school attendance.

The literature on knowledge spillovers is considered in Section 2. Section 3 outlines the data and methodology that we use to document the existence of externalities to schooling in agricultural production. Our results and interpretation are presented in Section 4. The scale of externality benefits in each site is discussed in Section 5. Section 6 considers alternative interpretations of the results. Section 7 concludes.

2. Some Relevant Literature

Knowledge externalities represent probably the most exciting, and intuitively appealing, theoretical development from the recent human capital literature. New growth theory (initiated by Romer 1986, 1990; Lucas 1988) has focused on intergenerational transfers of knowledge under conditions of increasing returns whereby each generation benefits from the stock of knowledge left by the previous generation. The possibility of a positive externality exists: too little knowledge is created by previous generations, since there is no market in which knowledge can be sold to future generations. The theory suggests that economic growth is driven by endogenous technological progress, which arises as a result of knowledge spillovers. Knowledge capital differs from other types of capital in that it is a non-rival input (i.e., its use by one agent does not preclude its use by others) and it is at least partially non-excludable (i.e., it is not possible for one economic agent completely to appropriate the rents from knowledge). Since knowledge is both an input into innovation and a product of such innovation, it is self-perpetuating (Grossman and Helpman 1991).

Although new growth theory emphasises the key role of the creation of new products and manufacturing processes, deliberate research and development is not the only type of innovating behaviour in which firms engage. In the developing world in particular, the primary avenue for technical progress is the adaptation of pre-existing products and production methods to local economies (Grossman and Helpman 1991). Often, tiny changes in production processes already in existence provide important improvements in productivity (Lall 1987). In developing countries, the skills and knowledge required to adapt new technology to the local environment are typically scarce, and thus a learning process is necessary. Learning occurs through a combination of three factors: experience gleaned from the act of production; knowledge capital imported from the developed world; and conscious accumulation of know-how. Producers must develop the know-how to choose which technology to purchase, adapt it to the local economy, and then use it efficiently (Lall 1987). Formal schooling may reduce the cost of acquiring such know-how, thereby facilitating the adoption of innovations. The externality arises if those without schooling copy the adoption behaviour and productive practices of the educated.

The logic of the literature on externality effects of knowledge as applied to firms is applied to individuals in the literature on ethnicity and geographical spillovers of human capital within neighbourhoods. Borjas (1995) postulates and tests a model whereby human capital is passed from one generation to the next not only through the human capital of the parents but also through the average human capital of the ethnic group in the parents' generation. His paper, like many others from the more general literature on neighbourhood spillovers (see, for example: Glaeser et al. 1992; Glaeser and Mare 1994; Rauch 1992; and Moretti 1998), focuses on urban neighbourhood effects and the wage labour market.

Our work examines the possibility of human capital spread effects within rural neighbourhoods where subsistence agriculture is the primary household activity. Foster and Rosenzweig (1995) consider a similar question using Indian data, but focus on external returns to experience with high-yielding seed varieties, rather than to education *per se*. By including average village experience with high-yielding seed varieties, they are able to show that both the experiences of the household and the experience of the household's neighbours was important to profitability in the use of high-yielding seed varieties and to rates of adoption. The only directly comparable previous research that we know of for Africa is Appleton and Balihuta (1996), who study the externality effects of schooling upon agricultural productivity in Uganda. They note that, whereas Lucas (1988) discussed knowledge externalities in a time-series context, allowing for changing technology, there can also be observable externalities in a cross-section of data if there are differing rates of adoption of new technologies across space. To test for externality effects of education, they add average levels of primary and secondary schooling of other farmers in the site to their equation, and find quite large externalities. They caution that these effects may be overstated, because they could not control for unobserved community fixed effects in the equation. Nevertheless, they conclude that externality effects may be even more important than the internalised effects of schooling upon productivity in agriculture.

Although we are not aware of other direct evidence of externalities to schooling using Ethiopian data, researchers have reported some suggestive findings. For example, Croppenstedt, Demeke and Meschi (1998), using data from a 1994 USAID fertiliser marketing survey, find that literate farmers are more likely to adopt use of fertiliser than those who are illiterate. This indicates that the educated may be early adopters. If innovations diffuse from the educated to those without schooling, there are externalities of education at work. Using the same data, we examine the adoption and diffusion of innovations, as a particular pathway for the transmission of externalities, in a complementary paper (Weir and Knight 2000). The focus of the present paper is to examine cross-sectional evidence of the static effect of education externalities on production; that of the complementary piece is directly to analyse the dynamic process of innovation and diffusion.

3. Data, Hypotheses and Methods

Our data are drawn from the first round of the Ethiopia Rural Household Survey (ERHS)[1] (Dercon and Krishnan 1994a). The survey covered 1477 households in 18 Peasant Associations (villages or clusters of villages which make up a survey site) spanning 15 Woredas (districts) in six regions. Six of the sites, primarily located in drought-prone areas, had previously been surveyed by IFPRI in 1989. The remaining nine were chosen to reflect

[1] The initial round of this large panel survey was conducted by the Department of Economics, Addis Ababa University, in collaboration with the Centre for the Study of African Economies, Oxford, and the International Food Policy Research Institute (IFPRI), Washington, in 1994.

most of the important agro-economic variations found in rural Ethiopia. Together, the 15 sites provide a realistic mix of cultivation categories and standard of living strata.

The number of households surveyed in each site reflects the size of the Peasant Association (PA) in relation to the total size of all PA's surveyed. Households were selected randomly using the PA registers, with female-headed households proportionally represented. Each household was surveyed three times (three rounds) within approximately twelve months (early 1994, later in 1994 and early in 1995), providing a picture of both past and present. Questions were asked on a wide range of issues affecting rural households, including production (crop output, land, labour and other inputs), innovation, consumption, assets, credit, migration, anthropometric measures, education and health (Dercon and Krishnan 1994a-c).

We test the following basic hypothesis: education at the site- as well as the household-level has a positive and significant effect upon cereal crop production in rural Ethiopia (i.e., there are external as well as internal benefits of schooling in terms of farm productivity). We find evidence consistent with important external effects. This leads us to examine a second hypothesis: education at the household- and site-levels affects *placement* of the production frontier (i.e., the internal and external effects of schooling include increasing potential output) and *deviations* from that frontier (i.e., the internal and external effects of schooling include increasing production efficiency). The results throw light on the mechanisms governing externalities to education in farm production.

The internal benefits of schooling in terms of household farm output will be examined by estimating a production function and production frontier with education at the household-level as an independent variable. Externality benefits will be captured by adding to the production function and production frontier a variable representing education at the site-level. The coefficients on household- and site-level education in the production function, respectively, provide information on the internal and external effects of schooling upon farm productivity. The coefficients in the frontier estimation provide evidence on the internal and external benefits of education in terms of placement of the frontier. Deviations from the frontier indicate the degree of production efficiency of the farm. This information may be used to estimate the internal and external benefits of schooling in terms of reducing farmer inefficiency.

Ironically, the presence of externalities may prevent us from finding evidence that education affects productivity at the household level (Jamison and Lau 1982; Phillips and Marble 1986). The effects of education on farm productivity may not be apparent if less educated farmers copy the better agricultural practices of their more educated neighbours. As well as presenting an empirical difficulty, this point is highly relevant from a policy perspective, since the presence of externalities may reduce the private demand for schooling while at the same time raising its social value.

There are two alternative production function methods to investigate externalities to schooling. One approach is to estimate production functions at higher levels of aggregation (Chaudhri 1979). For most countries, it is necessary to use highly aggregated data to capture all educational externalities. However, for Ethiopia, one would expect externalities to be more confined geographically, since communities are further apart and road networks incomplete. Thus, it may be that data aggregated at the site-level will reveal most of the externality effects of education. A site-level production function can be estimated using information averaged across the households in each site. However, not only does this approach require the assumption of homogeneity and linearity of individual farm production functions within each site but also the sample size - only 15 sites are contained in the survey - may be too small to generate reliable estimates.

Appleton and Balihuta (1996) suggest an alternative approach using individual-level data to estimate an equation with site-level aggregate education levels included:

(1) $$\ln Q_i = \alpha_0 + \Sigma\, \alpha_j \ln X_{ji} + \beta S_i + \Sigma\, g_k Z_{ki} + ED_v + \varphi_i$$

where Q_i is farm output for household i; X_{ji} is a vector of other inputs j for household i; S_i is a variable(s) representing average education for household i; Z_{ki} is other household characteristics for household i; ED_v is average education for site v; and φ_i is a stochastic error term. To facilitate comparison of internal and external effects of schooling, the site schooling variable, *ED*, can be specified as average years of schooling of all adults in the site. Its coefficient may then be interpreted as the average increase in household output for each additional year of average schooling in the site. Alternatively, the site-level education variable can be the average years of schooling of the most educated *x*-percent of the population or some other indicator of externality-generating education.

To determine whether there are internal and externality benefits of schooling in terms of the placement of the production frontier and in terms of increased efficiency of farmers in producing output using current technology, we also estimate stochastic frontier production functions with site-level education as an explanatory variable.[2] A production frontier is estimated based on the most efficient observed use of inputs to produce each level of output. The extent to which farm production departs from the frontier provides a measure of technical inefficiency for the sample as a whole or for each farm individually. The causes of technical inefficiency can be investigated by regressing inefficiency on education and other explanatory variables (Ali and Byerlee 1991).

Stochastic frontier estimation involves specification of a two-part error term, $Y = f(X)e^{v-u}$, where v represents random shocks, such as measurement error or factors which are external to the firm (e.g., weather), and is symmetric and distributed normally. The second component, u, is a one-sided, strictly non-negative, error representing technical inefficiency. Jondrow et al. (1982) show how to decompose the v-u term to provide estimates of technical inefficiency by calculating the expected level of inefficiency for each farm, $E(u_i)$, conditional on the random component, v_i (Ali and Byerlee 1991). Battese and Coelli (1988) provide a formula to estimate farm-specific efficiency in the case of a logged dependent variable.

Two approaches have been described to study the effects of schooling on productive efficiency using frontier production function analyses. The first approach is to estimate the production frontier with all relevant inputs, including education, in one stage as follows:

(2) $$Y = f_1(X, Z, H, E_n)e^{v-u}$$

where: X are direct inputs under the control of the farm manager; Z are environmental variables; H are other exogenous variables specific to the household, including education; and

[2] See Weir (1999) for a detailed discussion of this technique.

E_n is a site-level measure of educational attainment. The coefficient on household education in the production frontier indicates the effect of education on the productivity of the most efficient farm operators, and does not provide information on the effect of education upon typical farms in the sample.

In an alternative two-stage approach, a stochastic frontier production function is estimated using only inputs under the direct control of the farmer and exogenous environmental factors in the first stage in order to obtain estimates of inefficiency. Measured inefficiency is explained in the second stage using exogenous characteristics of the household and productive environment, including education at the household- and site-levels. The coefficient on years of schooling in the second stage, expected to be negative, represents the reduction in inefficiency owing to an extra year of education. A general two-stage model is as follows:

$$Y = f_1(X, Z)e^{v-u} \quad (3)$$

$$u = f_2(H, E_n) \quad (4)$$

where: X are direct inputs under the control of the farm manager; Z are environmental variables; H are other exogenous variables specific to the household, including education; and E_n is site-level education. Since the role of education in reducing farmer inefficiency is of primary interest, rather than an exhaustive consideration of the sources of inefficiency, only the education policy variables will be included in the second stage, while other exogenous household characteristics will be controlled for in the first stage.

In each of the above methods, a potential disadvantage of introducing the site-level education variable is that it may be correlated with other unobserved site-level variables, and the coefficient on this variable may incorporate not only the effects of average levels of schooling in the site but also other community fixed effects on farm output not caused by schooling. We may reduce this possibility of bias by inclusion of several other community-level variables. However, it is difficult to be certain that it has been eliminated, particularly without data on a large number of sites. Another approach will be to measure education at the neighbourhood-level (i.e., groups of households within each site) and control for all other sources of community fixed effects by including site dummy variables in the production function. Our data set allows us to pursue this modification of the methodology suggested by Appleton and Balihuta (1996).

Strauss and Thomas (1995) point out that, although adult educational attainment is a pre-determined variable, endogeneity may exist if investments in education made many years ago were correlated with unobserved variables which affect productivity today, such as ability and motivation or soil fertility. This issue is also relevant to the relationship between education and productivity at the site-level, since historically favourable, and unobservable, agro-economic conditions in the site may have led to greater investment in education as well as in other productivity-enhancing agricultural inputs. This may be relevant if education is determined by output at the site-level (e.g., if schools were set up in prosperous areas rather than school provision being decided centrally and determined by exogenous criteria).

In considering a production function with education aggregated at the site-level, controlling for site-level sources of spurious correlation between education and agricultural productivity is a necessary, though not a sufficient, condition for establishing causality. Historical information on the availability of school places and other constraints on the responsiveness of household investments in schooling to perceptions of increasing benefits of schooling may be used to show that the direction of causality is from education to productivity and that this relationship is not mediated by a third factor.

Table 1 describes the data to be used in both frontier and non-frontier production function estimation. Means are presented for the sub-sample of observations used in the econometric analysis. Data to estimate the relationship between education and farm productivity are drawn from the first round of the ERHS. The full sample for the ERHS contains 1477 households. However, the final sample employed for the present analysis includes only 616 households. Observations with missing or inconsistent information were omitted. To ensure that the estimated production function represented a single farming system and production technology, only those sites where cereals are cultivated primarily using oxen-plough technology were chosen. In addition, various outliers were omitted from the sample[3]: the lower bounds ensure that only households which are primarily engaged in farming were included, and the upper bounds were set to exclude data which are likely to have been erroneously recorded.

The decision to focus upon production of cereal crops alone, and to exclude sites where cereals are not cultivated using oxen-plough technology, may lead to bias in coefficients. However, education does not seem to increase the likelihood of employing oxen-plough cereal production technology. In fact, households in sites which primarily produced *enset* were found to have significantly higher average investments in schooling than those in the oxen-plough cereal-growing areas. Thus, by omitting the *enset*-growing region from the sample, no upward bias is expected in the estimates of the effects of schooling upon farm production. If there is any bias in the coefficients on years of schooling in the production function, it is expected to be downwards. Although it would be desirable to provide evidence on the maximum impact of schooling upon farm output, the focus on cereal crop producing sites alone will be maintained, since consistency is needed to facilitate the estimation, later in the paper, of a production frontier.

The dependent variable is the natural log of the value of cereal production deflated by a Laspeyre's price index.[4] Several of the farm variables are presented in logarithmic form. Input values of zero were transformed by adding the constant one to facilitate taking logs, as is common practice (see for example: Jacoby 1992; and MaCurdy and Pencavel 1986).[5] Land is

[3] These include: households using less than 0.2 ha of land; households using less than six or more than 800 person days of labour in total; households using less than two person days of labour for ploughing, or less than one person day of labour on weeding or harvesting; households which recorded cereal productivity per person day of labour ploughing of greater than 750, production of more than 620 per person day of weeding or in excess of 780 per person day of harvesting; and households which reported that the value of their capital assets exceeded 990 Birr.

[4] Cereal value is given by the sum of *tef*, wheat, maize, millet, barley and sorghum produced each multiplied by its price. An average price index is constructed for each household by weighting the price of each output by its share in total cereal production for that household (see Croppenstedt and Muller 1998).

[5] Jacoby (1992) cautions that use of different arbitrary constants to adjust inputs can result in very different coefficient estimates for some variables. Johnson and Rausser (1971) find that bias in parameter estimates increases with the size of the constant added, but modifying only the sub-sample of zero values results in less bias than modifying the entire sample. To test whether our results are sensitive to the choice of one as an arbitrary constant, alternative sets of results were obtained using the scalars 0.5 and 2 instead (not shown). The coefficients on some of the explanatory variables alter slightly. However, the differences are extremely minor, indicating that the degree of bias associated with the choice of one as a scalar for this data is quite small.

considered to be fixed in the long run, since buying or selling of land is prohibited. Labour, capital, oxen, and trees are fixed in the short run. Although not commonly practised, labour may be hired-in by households during busy periods, and various types of work-party arrangements are found in most sites. Use of commercial fertiliser, manure and other inputs is considered to be variable in the short run. Land quality and slope are proxies for environmental conditions.[6] Site dummy variables are included to capture site-specific fixed effects, such as problems of inadequate rainfall or widespread pests. The site dummy variables also capture variations by site in infrastructure quality and the timely availability of inputs, such as fertiliser, among other site characteristics.

Several different variables are available to proxy education at the household level. We focus on average years of schooling obtained within the household. Education at the site- or neighbourhood-level is measured by aggregating years of schooling of household members within the cluster. There are several possibilities. For example, we do not know, a priori, whether it is necessary to provide a certain amount of schooling on average to members of the cluster for external effects to become apparent or whether it is necessary only that a few individuals or households within the community are well educated for externalities to emerge. Experimentation was necessary to determine the correct specification of the community education variable. Two different variables are tested in the results presented below: average years of schooling of all adults in the cluster and average years of schooling of the most educated household in the cluster.

There is a danger in using the education of only the most educated in the cluster to illuminate external returns: if highly educated households tend to live in highly productive areas, regardless of their own contribution to farm output, the correlation between site-level education and productivity may be spurious. For example, civil servants, doctors or teachers may congregate in areas which are prosperous. If members of their households are not themselves primarily engaged in agricultural production, there is no reason why highly educated households should be active in encouraging the diffusion of new technologies. Therefore, before proceeding, it was necessary to ensure that the households with the highest levels of schooling in each neighbourhood tended to be farmers. Fortunately, this was generally found to be the case for our sample.[7]

---

[6] Other possible environment proxies, such as dummy variables indicating self-reports that the farmer's crops suffered from disease, or unfavourable weather, animals or other pests, are potentially endogenous and therefore have been excluded from the analysis.

[7] Investigating the main activity of household heads in the most educated two or three households in each neighbourhood revealed that over 90 percent were primarily engaged in farming. Only a few households were headed by non-farmers and, of these, only one was headed by a teacher.

4. Results

4.1. Basic Production Function Results

Results of the estimation of standard Cobb-Douglas production functions (without education variables) are presented in equation 1 of Table 2. The first equation incorporates all basic regressors, including site dummy variables. A Hausman test was not able to reject the null hypothesis that land, labour, capital, and fertiliser are weakly exogenous. Individual coefficient results are discussed in detail in Weir (1999).

Land, labour and capital coefficients are each positive and significant.[8] Their sum is 0.70, which suggests decreasing returns to scale. This is consistent with expectations, since minimum efficient scale in rural agriculture in developing countries is usually found to be rather low. This may be partly explained in terms of increased transport costs faced by larger scale farmers (Croppenstedt and Muller 1998, using the same data set as this paper). It may also reflect increased costs of monitoring labour when farms become large enough to need to hire labour. Overall, the Cobb-Douglas production functions estimated are able to explain 55 percent of variation in cereal production.[9] This performance is good, given that production is influenced by a number of unobservable variables, such as farmer ability.

4.2. Effects of Education upon Cereal Crop Production

We now turn to an analysis of the internal (household-level) and external (site-level) effects of education upon cereal production. Equation 2 in Table 2 shows that an additional year of schooling on average in the household increases cereal output by four percent. This result is significant at the one percent level on a two-tailed t-test. It compares well with results for other countries, particularly given that Ethiopian agriculture is predominately traditional in nature.[10]

In Table 2, the coefficients on the site dummy variables included the effect of site-level education investments along with all other site-level effects upon output. We wish to investigate the effect of site-level education separately from that of other site-specific effects. It is necessary, therefore, to omit the site dummy variables included previously. Table 3 illustrates the internal versus external effects of schooling. Including average education in the site in a household production function, along with average years of schooling in the household, shows that the effect of education at the site-level (the external effect) is considerably greater than the internal effect of education acquired by household members. Previous research has revealed a threshold effect, whereby at least four years of education are

[8] An alternative specification of the labour variable, natural log of the number of adults aged 16 to 60, was tested. The coefficient was virtually identical to that given in Table 2, but insignificant.

[9] A restricted translog specification was also estimated to determine whether the assumption of unitary elasticity of substitution inherent in the Cobb-Douglas production function is too restrictive. However, it was not possible to reject the null hypothesis that extra variables included in the translog function are jointly equal to zero at the five percent level, suggesting that the Cobb-Douglas specification is sufficient to represent cereal output in the sites surveyed.

[10] Data for a selection of other countries is provided in Lockheed, Jamison and Lau (1980), who observe that returns tend to be higher for regions characterised by modern farm technology. Updated and expanded comparisons are provided in Phillips (1994) and Appleton and Balihuta (1996).

required for an impact of household-level education on farm productivity to be fully evident. The effect of four years of schooling for farmers was found to be 12 percent (Weir 1999). This is an impressive result, but still well below the external effect documented here of a 43 percent rise in output for an extra year of schooling on average at the site-level (equation 3) .

The coefficient on site-level education in equation 3 is implausibly high. It may be that the community education variable is correlated with other site-specific fixed effects and that the coefficient on education at the site-level is biased upwards. To control for some additional community fixed effects, a set of site-level variables, including average rainfall, distance to an all-weather road, the proportion of female-headed households, the percentage of households which use manure and average use of commercial fertiliser in the site (which may proxy extension service activities), are incorporated in equation 4. The coefficient on site-level investment in schooling falls to 17 percent and becomes less significant than in the previous equation. However, the site-level effect is still considerably larger than that of education at the household level. It is difficult to control for all important site characteristics owing to the small number of sites surveyed and consequent lack of degrees of freedom. Community fixed effects may be correlated with site-level schooling. Thus, the external effects may still be overstated owing to omitted variable bias.[11]

A possible remedy is to examine external effects at a different level of analysis, while controlling for fixed effects at the site-level. To do this, it was necessary to examine site maps and group households geographically. While such groupings may not fully capture social networks of households, this approach should provide a reasonable approximation, since we may expect farmers to copy the practices of those located nearby. There are 48 different neighbourhoods in total. The number of neighbourhoods in each site was determined by the particular geographical features of the site and the distribution of households within the site (shown in the enumerators' site maps with each surveyed household marked) and ranged from two to nine.[12] On average, there are 12 households per neighbourhood.

Two alternative neighbourhood-level education variables are employed in Table 4. The first (shown in equation 5) is the average years of schooling of all adults in the neighbourhood. The result is that there is no influence upon farm production of living in a neighbourhood where average education is high. The coefficient on neighbourhood education takes a negative coefficient and is not significantly different from zero when other site-specific sources of variation in output are taken into account. It is reasonable that only one household need be educated to encourage innovation in an area as small as a neighbourhood. To test this, the neighbourhood education variable used in equation 6 is average years of schooling of the most educated household.[13] The results indicate that there is a significant, positive effect upon cereal output of living in a neighbourhood where at least one household has invested in education. Household production increases by three percent for each additional year of schooling, on average, of the most educated household in the neighbourhood. Thus, it is not necessary to have high average levels of schooling across the neighbourhood to induce a

[11] There is an additional problem of robustness. The coefficient on site education is sensitive to changes in the set of other site variables included.

[12] There were two sites for which no maps were available, these have been treated as single neighbourhoods for the purpose of calculating neighbourhood-aggregated data.

[13] Maximum education of the most educated individual is not used here, because the most educated person in the neighbourhood may be a teacher or public administrator and not primarily engaged in farming even if they belong to a farm household.

positive effect upon productivity at the household level. It is sufficient that at least one household has invested significantly in schooling.

4.3. Basic Stochastic Frontier Production Function Results

Results of estimating one- and two-stage stochastic frontier production functions are presented in Table 5. Land, labour, capital and the site dummy variables are the most significant explanatory variables in the first stage. The factor coefficients sum to 0.56, once again indicating decreasing returns to scale.[14] Average farm-specific efficiency is estimated to be 54 percent in equation 7. This is in line with estimates by Croppenstedt and Muller (1998), which range from 51 to 76 percent depending on the assumed distributional form of the one-sided error, and Admassie and Asfaw (1997), who estimate mean profit efficiency of 54 percent for a sub-sample of farms in the ERHS.

In the two-stage estimation (equation 8), the variables included in the first stage are the same as those used in the one-stage model estimated earlier. However, farm-specific measures of efficiency - calculated following estimation of the frontier in the first stage - are converted to measured inefficiency, which becomes the dependent variable to be explained in the second stage using average years of schooling of all adults in the household.[15] The estimated coefficients on years of education of all adults in the second stage is significantly negative, indicating that a one year increase in average schooling attained in the household will reduce measured farm inefficiency in the production of cereal crops by 2.1 percentage points. Thus, if educational attainment is raised from zero to four years of primary schooling on average in the household, mean efficiency is expected to rise from 54 to 62 percent overall. This represents an efficiency increase of 15 percent for the sample as a whole. The benefit of providing a basic primary education to all household members in terms of creating more efficient use of agricultural resources appears to be substantial.

4.4. Effects of Schooling upon Efficiency and the Frontier

The external effect of schooling upon the production frontier and efficiency is investigated in Table 6. The first two equations estimate a one-stage frontier. That is, the shape of the frontier is determined by including all variables in a single-stage MLE equation. If education influences the position of the frontier, the coefficient on years of schooling will be positive and significant. However, household education is not significant. This may be because educated farmers, having innovated, have been copied by uneducated farmers, obscuring the original importance of education upon the placement of the frontier.

Equation 9 shows that average years of education in the site has a significant positive influence upon the placement of the production frontier. When other community variables are included in the equation, the size and significance of the coefficient on average education in the site diminishes somewhat, but is still large and significant (equation 10). These frontier results

[14] These findings are described in detail in Weir (1999).

[15] The coefficients on the first stage variables are differ slightly between equations 7 and 8, as the two-stage estimates produced by FRONTIER incorporate assumptions regarding the independence of the inefficiency effects in the two stages which improves the consistency of estimation. It is not possible to test for a preferred specification (one-stage versus two-stage), since the models are non-nested (Coelli 1994).

mirror the external effect of schooling documented previously for the average production function.

Equations 11 and 12 estimate a two-stage frontier. The shape of the frontier is estimated in the first stage. Then, deviations of each farm household's production from the frontier are calculated and used as the dependent variable in the second stage. Years of education are included as regressors in the second stage to help explain measured inefficiency. If education helps to reduce inefficiency, the coefficient on schooling in the second stage will be negative and significant. The coefficient on average education in the site is almost significant in equation 11, suggesting that there may be external benefits of schooling in terms of reduced farm inefficiency. However, when community variables are included in the first stage of equation 12, the effect entirely disappears. In this case, average education in the site appears to act as a proxy for other site characteristics and to have no influence of its own upon farm efficiency.

The results in Table 6 suggest that, whereas site-level education has a significant impact upon placement of the frontier, increasing average education in the site does not influence efficiency deviations from a given frontier. The reported strong external benefits of schooling apparently take the form of encouraging the spread of agricultural innovations, which push out the production frontier, rather than increasing efficiency given existing technology. That is, there is no evidence of external effects of schooling in terms of production efficiency. It may be that individual farmers must adapt use of inputs to their particular circumstances, and that information on the best-practice use of new inputs is not easily transferable between farms, for instance if they differ in land quality and suitability for use with modern inputs.

For purposes of comparison, it is also possible to re-do the frontier analysis using neighbourhood-level education variables and controlling for other site fixed effects in the first stage. The first stage results, presented in Table 7, provide evidence that average years of schooling of all adults in the neighbourhood has a significant influence upon the placement of the frontier, having controlled for other site-level fixed effects (equation 13). At 18 percent, it is slightly smaller than the effect of education at the site-level after other site variables have been included (see equation 10 in Table 6). However, living in a neighbourhood where at least one household has invested heavily in schooling for its members does not appear to be important in this case (equation 14). These findings contrast with those on neighbourhood education in the average production function shown in the previous section, where average years of schooling of all adults had no effect upon household productivity after controlling for other sources of site level variation but the education of the most educated household was found to play a role (see Table 4). This suggests that the relevant specification of the neighbourhood education variable may depend upon the type of equation being estimated. Average productivity is enhanced by living in a neighbourhood where at least one household has invested heavily in schooling, whereas high average investments in education in the neighbourhood appear to be necessary to shift out the production frontier.

Equations 15 and 16 show the results of the second stage analysis where household and neighbourhood education variables are used to explain changes in inefficiency. If neighbourhood education reduces farm inefficiency, the coefficient on education at the neighbourhood-level will be negative and significant. Whether the neighbourhood education variable is average years of schooling of all adults or average years of schooling in the most educated household, the results are the same. Neighbourhood education has no significant impact upon the efficiency with which households produce cereals crops. This confirms the results presented earlier for education at the site-level (see Table 6).

5. The Scale of Externality Benefits of Schooling by Site

The substantive effect of education at the site-level upon household production depends upon average investments in schooling in each site. Controlling for site-level variations in internal returns to schooling (e.g., the interaction between site and household education) and other farm and household variables, the effect of average education in the site upon production by site may be simulated. The effect of education by site is simulated by using the coefficients from a production function estimated across all sites to compute predicted farm output associated with the particular site-level investment in education for each site. Then, the percentage change in output in a given site compared to the mean for all sites indicates the strength of the external effect of schooling in that site. Table 8 provides information on average years of schooling in the site, the predicted natural log of cereal output for each site and the percentage difference between predicted cereal output in the site and mean output for the entire sample.

External benefits of education are found to vary from 15 percent above average in Site 10 to 7 percent below average in Site 2, as a result of different site investments in schooling. Also shown are simulated external effects of schooling in a fictional site with no investment in schooling whatsoever and one where the site has an average of six years of schooling complete. The difference is striking and suggests the possibilities for sites when there are large increases in educational attainment.

6. Alternative Explanations of the Results

The evidence presented above suggests that social benefits of schooling are substantial, particularly in comparison to private benefits. An additional year of formal schooling on average in the site has a much larger impact upon household farm productivity than an additional year of schooling on average within the household. This provides support for the hypothesis that there are strong externality effects of schooling even in a relatively traditional farming system. However, we must consider alternative explanations for our results.

It is important, but difficult, to demonstrate that the relationship between education and output at the site-level is causal. Appleton (1999) notes that high external effects may be due to the very large differences in productivity between the sites sampled. To illustrate, a simulation of predicted output based on coefficients estimated in a standard production function (excluding household- and site-level education), where only the effect of the site dummy variable is allowed to vary, indicates that households in the second most productive site produce five times more output than those in the second least productive site, ceteris paribus.[16] This may suggest that farmers in certain sites are in a better position to benefit from higher investments in schooling in the site because they are already using more modern technologies or are located near to large markets. However, it may also suggest that site-level educational investments are spuriously correlated with site-level output. That is, sites with high farm

[16] The standard deviation of the distribution is two-thirds of the mean.

output may also tend to be sites with good access to schooling because of a third factor, such as agro-economic conditions which have historically been conducive to high farm output and to investments in both education and the farm.

The possibility that investments in both education and productivity-enhancing farm inputs are caused jointly by a third factor, such as agro-economic conditions, is plausible only to the extent that households are able to vary their behaviour in response to changes in the third factor. The charge of spurious correlation will be weakened considerably if it can be shown that households are not able to respond to economic incentives to send their children to school. In fact, there is some evidence to suggest that households in our sample are prevented from investing in education even where there are strong incentives to do so owing to institutional constraints, such as limited supply of school places or restricted access to credit, and to technical constraints, such as children being needed for farm work and other household activities. Weir (1998) finds severe supply-side constraints (including transport and transaction costs) upon enrolment in many sites covered by the Ethiopia Rural Household Survey on account of the distances required to attend good quality - or any - schools. Poor timing of school fee demands and of the school year impose serious constraints upon schooling owing to the lack of credit and labour markets, which prevent farm households from borrowing to pay school fees during lean periods and from hiring farm labour from outside the household during busy periods. Furthermore, there is compelling descriptive evidence that opportunity costs of schooling are a critical determinant of school enrolment. This suggests that households in sites where farm productivity is high may actually be less likely to send their children to school than households in sites where productivity is low. Given the weight of the evidence, the argument in favour of spurious correlation is weak.

To take this further, we may test the hypothesis that investments in education are not highly sensitive to economic incentives using data from a sub-sample of four sites covered by the Ethiopia Rural Household Survey where an in-depth survey of human capital was conducted in 1995. Data from the Education Sub-Sample Survey is used to estimate the probability of school enrolment among school-age children (those aged 7 to 18 years who are not listed as head or spouse of the household). The marginal effects of each variable upon the probability of enrolment are presented in Table 9. For continuous variables, the number reported indicates the increase in the probability of enrolment associated with a unit change in the regressor. For dummy variables, the change in probability of enrolment associated with a change in the regressor from zero to one is reported.

Equation 17 is able to explain 30 percent of the variation in enrolment. Given the potentially large effects of unobserved characteristics, such as motivation, ability and health of the child, this success rate is good. However, it does suggest that the probability of enrolment is not highly sensitive to changes in the variables included. In particular, if credit is constrained and if households respond to rising income by investing in more schooling for their children, the coefficient on income in the probit is expected to be positive and significant. However, the variable used to proxy income - consumption per adult equivalent - has an insignificant, negative coefficient. Furthermore, of the variables indicating cost constraints, school fees as a percentage of consumption per adult equivalent is also negative and insignificant and land holding per adult equivalent is positive but not quite significant at the ten percent level on a two-tailed t-test. Instead, the results show that variables related to the intra-household allocation of resources and those representing parental traits are the most significant. Site fixed effects, which incorporate the availability and quality of schools in the local area, are also important. Thus, educational resources are rationed within the household according to the traits of the children, within the community according to the traits of parents, and between communities according to the provision of school resources. However, income elasticity of demand for schooling is rather low: it appears that households are prevented from responding to changes in income or prices in their human capital investments.

If these findings can also be extended back to previous decades, as they almost certainly can, it is less likely that the link between education and production is spurious. Since, historically, schools have been mainly supplied by the government and by missionary activity (Weir 1998), the link between productivity and investments in schooling in the past is probably weak in rural Ethiopia. Thus, the issue of causality is less likely to pose a problem than would be the case were demand for education found to be highly responsive to changes in income or were school facilities provided in response to community demand for education or rising average prosperity.

This question might be definitively settled if we could instrument the site education variable and test for endogeneity. Unfortunately, this is not possible given a lack of historical data on education and household income of the families of adults in the community. As a second best measure it was possible, in the econometric results presented above, to control for other site fixed effects by including several site-level variables, including education, in the production function. Including other site-level variables (in equation 4) resulted in a loss of explanatory power for site education. This indicates that the site-level education variable, previously used alone, was partly picking up the effects of other site-specific fixed effects and that the true external effect of schooling is not as high as the original estimate suggested. Without perfect information on each individual site fixed effect, it is not possible to know how much of the remaining effect attributed to site-level education is actually due to that variable and not to other unobserved site characteristics. Indeed, other community variables may themselves explain average school attainment in the site. However, if sites with more education are more likely to take the initiative in building better roads to transport goods to and from markets and if they are more likely to invest in land quality improvements, then it is reasonable to omit these variables and estimate the effect of average schooling without their inclusion in the equation.

The use of neighbourhood-level information on schooling represents a way to control for site fixed effects. However, the full externality effects of schooling may then be understated insofar as farmers' networks extend beyond their immediate geographic cluster. Networks may be based on religious, ethnic or kinship ties and the relative influence of one member of a network upon another may depend upon demographic similarities or differences between the two. Furthermore, credit constraints may prevent farmers from copying new technologies, even if the experience of their neighbours indicates that the innovation is likely to be profitable. Hence, the neighbourhood-level education effects reported in Table 4 may be understated.

7. Summary and Conclusion

Interest in the idea of knowledge spillover effects upon economic development is not new. However, strong quantitative microeconomic evidence that productivity is enhanced by the spread of knowledge in industry or agriculture has been lacking from the applied literature on knowledge spillovers (Behrman 1990). This may be partly because of a paucity of reliable data and partly because of the difficulties associated with measuring externality benefits.

This paper has attempted to provide evidence on externality benefits of schooling in terms of farm production and efficiency in rural Ethiopia. We have demonstrated that there may be substantial benefits to be gained, both in terms of average productivity and in terms of placement of the production frontier, by living in a place where at least some households have invested heavily in schooling. The size of the coefficient on education at the neighbourhood-level, controlling for community fixed effects, is lower than the coefficient on education at the site-level, where other site effects are not fully incorporated. However, the externality effect of schooling upon production is still found to be significantly positive and is comparable with the internal effect of education upon output. Moreover, there is a large, significant effect of neighbourhood-level schooling upon placement of the frontier, whereas no evidence of an internal effect is provided in the one-stage frontier estimation. This suggests that diffusion of innovations - which shifts the production frontier outwards - is facilitated by education in the neighbourhood.

Where the internal effect of education is clearly more important than the externality effect is in reducing inefficiency in the use of a given technology. It appears that efficient production requires that each household make its own investments in education to reduce the costs of adapting the new technology to its own particular situation. This may occur because although uneducated farmers can be influenced to adopt a new technology by observing the profits of educated farmers, new technologies must be adapted to the specific features of a particular farm (e.g., land availability, soil quality, pest problems, etc.). Thus, farmers may need to calculate their own quantities of the input to use and learn how to apply the input effectively, given their own farm characteristics. The process of adapting a new technology to a particular farm may be faster and less costly for farmers who have been to school. The externality benefits of education are limited to the extent that information on appropriate quantities and usage of the new input may not easily be shared among farms.

Our unique approach of applying both average and frontier production function methods to the measurement of internal and external returns to education provides evidence to suggest that the productivity gains associated with living in a highly-educated cluster may be attributed to the spread of innovations, rather than to efficiency gains associated with a given technology. This explanation is explored in Weir and Knight (2000), which examines the role of education in facilitating the adoption and diffusion of innovations and finds that early adopters of fertiliser tend to be more educated than those who copy them, and that the extent and speed of diffusion within a site depends upon average education in the site. Thus, there is evidence for two types of schooling externalities in the adoption of a new technology: firstly, less educated households copy the innovative behaviour of more educated households; and secondly, the speed and success of diffusion of the innovation depends upon site-level investments in schooling.

References

Admassie, Assefa, and Abay Asfaw (1997) The impact of education on allocative and technical efficiency of farmers: The case of Ethiopian small holders, paper presented at the Seventh Annual Conference on the Ethiopian Economy in Nazeret (Addis Ababa: Addis Ababa University Department of Economics, mimeo).

Ali, Mubarak, and Derek Byerlee (1991) Economic efficiency of small farmers in a changing world: A survey of recent evidence, *Journal of International Development*, **3**, 1-27.

Appleton, Simon (1999) Education and health at the household level in sub-Saharan Africa, paper prepared for the AERC collaborative research project on Explaining African Growth Performance (Oxford: Centre for the Study of African Economies, mimeo).

Appleton, S. and A. Balihuta (1996) Education and agricultural productivity: evidence from Uganda, *Journal of International Development*, **8**, 415-444.

Battese, George E. and Tim J. Coelli (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data, *Journal of Econometrics*, **38**, 387-99.

Behrman, Jere R. (1990) Human resource externalities on a micro level in rural Pakistan, report prepared for the Pakistan Rural Education Research Team (Philadelphia: University of Pennsylvania, mimeo).

Borjas, G.J. (1995) Ethnicity, neighbourhoods and human-capital externalities, *American Economic Review*, **85**, 365-389.

Coelli, Tim (1994) A guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation (Department of Econometrics, University of New England, Australia, disk copy).

Chaudhri, D. P. (1979) *Education, Innovations and Agricultural Development: A Study of Northern India* (London: Croom Helm Ltd. for the International Labour Organisation).

Croppenstedt, Andre, Mulat Demeke and Meloria M. Meschi (1998) Technology adoption in the presence of constraints: the case of fertiliser demand in Ethiopia, mimeo (Oxford: Centre for the Study of African Economies).

Croppenstedt, Andre and Christophe Muller (1998) The impact of health and nutritional status of farmers on their productivity and efficiency: Evidence from Ethiopia, mimeo (Oxford: Centre for the Study of African Economies).

Dercon, Stefan, and Pramila Krishnan (1994a) Ethiopia Rural Household Survey: Questionnaire for the First Round (Oxford: Centre for the Study of African Economies).

Dercon, Stefan, and Pramila Krishnan (1994b) Ethiopia Rural Household Survey: Questionnaire for the Second Round (Oxford: Centre for the Study of African Economies).

Dercon, Stefan, and Pramila Krishnan (1994c) Ethiopia Rural Household Survey: Questionnaire for the Third Round (Oxford: Centre for the Study of African Economies).

Foster, Andrew and Mark R. Rosenzweig (1995) Learning by doing and learning from others: human capital and technical change in agriculture, *Journal of Political Economy*, **103**, 1176-1209.

Glaeser, E.L., H.D. Kallal, J.A.Scheinkman and A. Schleifer (1992) Growth in cities, *Journal of Political Economy*, **100**, 1126-151.

Glaeser, E.L. and D.C. Mare (1994) Cities and Skills, NBER Working Paper **4728** (Cambridge, MA: National Bureau of Economic Research).

Grossman, Gene M., and Elhanan Helpman (1991) *Innovation and Growth in the Global Economy* (Cambridge, MA: M.I.T. Press).

Jacoby, H. (1992) Productivity of men and women and the sexual division of labour in peasant agriculture of the Peruvian Sierra, *Journal of Development Economics*, **37**, 265-87.

Jamison, Dean T. and Lawrence J. Lau (1982) *Farmer Education and Farm Efficiency* (Baltimore: Johns Hopkins University Press).

Johnson, S. R. and Gordon C. Rausser (1971) Effects of misspecifications of log-linear functions when sample values are zero or negative, *American Journal of Agricultural Economics*, **53**, 120-24.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt (1982) On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics*, **19**, 233-38.

Lall, Sanjaya (1987) *Learning to Industrialize: The Acquisition of Technological Capability by India* (Basingstoke: Macmillan).

Lockheed, Marlaine E., Dean T. Jamison and Lawrence J. Lau (1980) Farmer education and farm efficiency: A survey, *Economic Development and Cultural Change*, **29**, 37-76.

Lucas, Robert E. (1988) 'On the mechanics of economic development,' *Journal of Monetary Economics*, **21**, 3-42.

MaCurdy, Thomas E. and John H. Pencavel (1986) Testing between competing models of wage and employment determination in unionised markets, *Journal of Political Economy*, **94:3**, S3-S39.

Moretti, Enrico (1998) Social returns to education and human capital externalities: evidence from cities, Center for Labor Economics Working Paper No. 9 (Berkeley: UC Berkeley).

Phillips, Joseph M., and Robert P. Marble (1986) Farmer education and efficiency: a frontier production function approach, *Economics of Education Review*, **5:3**, 257-264.

Rauch, J.E. (1992) Productivity gains from geographic concentration of human capital: evidence from the cities, *Journal of Urban Economics*, **34**, 380-400.

Romer, Paul M. (1986) Increasing returns to long run growth, *Journal of Political Economy*, **94:5**, 1002-1035.

Romer, Paul M. (1990) Endogenous technical change, *Journal of Political Economy*, **98**, S71-S102.

Strauss, John and Duncan Thomas (1995) Health and labour productivity: sorting out the relationships, in G. H. Peters and Douglas D. Hedley, eds., *Agricultural Competitiveness: Market Forces and Policy Choice, Proceedings of the 22nd International Conference of Agricultural Economists held in Harare, Zimbabwe* (Aldershot: Dartmouth Publishing Company Limited), 570-90.

Weir, Sharada and John Knight (2000) Adoption and diffusion of agricultural innovations in Ethiopia: the role of education (Oxford: Centre for the Study of African Economies, mimeo).

Weir, Sharada (1999) The Effects of Education on Farmer Productivity in Rural Ethiopia, CSAE Working Paper Series 99-7 (Oxford: Centre for the Study of African Economies).

Weir, Sharada (1998) The Determinants of School Enrolment in Rural Ethiopia: Attitudes, Returns and Resources (Oxford: D. Phil. Thesis).

Table 1: Variable Definitions and Mean Values

Variable Name	Definition	Mean
Dependent Variable:		
LN_CEREAL	Natural log of the value of cereal crops (deflated)	5.77
Farm Variables:		
LN_LAND	Natural log of land cultivated - cereals (hectares)	0.19
LN_LABOUR	Natural log of person days worked (ploughing/harvesting)	4.28
LN_CAPITAL	Natural log of the value of hoes/ploughs (Birr)	2.58
LN_INPUTS	Natural log of value of other inputs (Birr)	1.95
OXEN	Dummy: 1 if household owns at least 2 oxen	0.27
BANANA	Dummy: 1 if number of banana trees>=10	0.08
COFFEE	Dummy: 1 if number of coffee plants>=5	0.11
CHAT	Dummy: 1 if number of *chat* plants>=50	0.10
EUCALYPTUS	Dummy: 1 if number of eucalyptus trees>=5	0.33
ENSET	Dummy: 1 if number of *enset* plants>=12	0.10
GESHU	Dummy: 1 if number of *geshu* plants>=12	0.05
LN_FERT	Natural log of quantity of fertiliser used (kg)	2.11
MANURE	Dummy: 1 if manure used on fields	0.49
LN_QUALITY	Natural log of cereal land quality (fertile - - - infertile)	0.44
LN_STEEPNESS	Natural log of cereal land slope (flat - - - steep)	0.19
Household Variables:		
AGE_HHH	Age of household head (years)	45.35
AGESQ_HHH	Square of age of household head (years)	2288
FEM_HEAD	Dummy: 1 if household head is female	0.17
NONAGR_HHH	Dummy: 1 if household head is not primarily a farmer	0.05
NO_ADULTS	Dummy: 1 if head is only adult in household	0.05
Village Variables:		
ROAD	Dummy: 1 if site has access to an all weather road	0.50
FEMHEAD_AVG	Average number of female-headed households in site	0.24
RAINFALL	Rainfall in the site in the period preceding Round 1	1.08
FERT_AVG	Average fertiliser usage in the site (kg)	48.43
MANURE_AVG	Percentage of households in site who use manure for fertiliser	0.44
Site Dummy Variables:		
SITE_2	Dummy: household lives in site 2	0.02
SITE_3	Dummy: household lives in site 3	0.09
SITE_4	Dummy: household lives in site 4	0.15
SITE_5	Dummy: household lives in site 5	0.06
SITE_6	Dummy: household lives in site 6	0.15
SITE_7	Dummy: household lives in site 7	0.11
SITE_8	Dummy: household lives in site 8	0.12
SITE_9	Dummy: household lives in site 9	0.13
SITE_10	Dummy: household lives in site 10	0.05
SITE_15	Dummy: household lives in site 15	0.09
SITE_16	Dummy: household lives in site 16	0.03
Education Variables:		
ED_AD	Avg. years of schooling of all adults in hh	1.25
ED_VILLAGE	Average years of schooling in village	1.18
ED_NEIGH_AVG	Average years of schooling in neighbourhood	1.10
ED_NEIGH_MAX	Maximum of average years of schooling in neighbourhood	5.30

Note: Means are based on the 593 observations (out of 1477 observations in total) used in the analysis. We are grateful to Andre Croppenstedt for providing his set of farm and household variables for use in this analysis.

Table 2: OLS Estimation of the Average Production Function
Dependent Variable: Natural Log of Cereal Crops Produced (LN_CEREAL)

	Eqn. 1	Eqn. 2
CONSTANT	5.158 ***	5.046 ***
LN_LAND	0.374 ***	0.377 ***
LN_LABOUR	0.181 ***	0.177 ***
LN_CAPITAL	0.145 ***	0.142 ***
LN_INPUTS	-0.009	-0.010
OXEN	0.245 ***	0.242 ***
BANANA	0.303 ***	0.308 ***
COFFEE	-0.131	-0.146
CHAT	0.391 ***	0.406 ***
EUCALPTUS	0.146	0.158
ENSET	0.621 ✔	0.632 ✔
GESHU	0.035	0.080
LN_FERT	0.053 **	0.052 **
MANURE	0.247 *	0.240 *
LN_QUALITY	-0.076	-0.097
LN_STEEPNESS	0.033	0.062
AGE_HHH	-0.032 *	-0.028 ✔
AGESQ_HHH	0.000 ✔	0.000 ✔
FEM_HEAD	-0.221	-0.232
NO_ADULTS	-0.167	-0.148
NONAGR_HHH	-0.581 *	-0.584 *
SITE_2	-0.839 ***	-0.828 ***
SITE_3	-0.616 ***	-0.600 ***
SITE_4	-0.058	-0.091
SITE_5	0.267 *	0.212 ✔
SITE_7	0.939 ***	0.903 ***
SITE_8	0.206	0.223
SITE_9	-0.734 ***	-0.743 ***
SITE_10	0.683 ***	0.549 **
SITE_15	-2.454 ***	-2.482 ***
SITE_16	0.449 ***	0.415 ***
ED_AD		0.042 ***
R^2 (unadj)	0.562	0.563
NUMBER OBS.	593	593

Note: Standard errors have been adjusted to account for the clustered nature of the data. Stars indicate significance using a two tailed t-test as follows: *** = 0.01; ** = 0.05; * = 0.10; ✔ = 0.20. It is unconventional to be concerned with significance at the 20 percent level. However, under certain hypotheses, only the upper (or lower) tail of the distribution is relevant (Maddala 1992; Pindyck and Rubinfeld 1991; Larsen and Marx 1986). Since the appropriate form of the hypothesis to test may vary according to particular circumstances, coefficients which are significant at the ten percent level on a one-tailed t-test (equivalent to the twenty percent level on a two-tailed test) have been flagged.

Table 3: Ordinary Least Squares Production Function
External versus Internal Effect of Schooling, All Adults
Dependent Variable: Natural Log of Cereal Crops Produced (LN_CEREAL)

	Eqn. 3	Eqn. 4
ED_AD	0.062 ***	0.034 **
ED_VILLAGE	0.432 **	0.175 *
FARM_VARIABLES	YES	YES
HH_VARIABLES	YES	YES
SITES	NO	NO
VILLAGE VARIABLES	NO	YES
R^2 (unadj)	0.477	0.551
NUMBER OBS.	593	593

Note: Standard errors have been adjusted to account for the clustered nature of the data. Stars indicate significance using a two tailed t-test as follows: *** = 0.01; ** = 0.05; * = 0.10; ✔ = 0.20.

Table 4: Ordinary Least Squares Production Function
External vs. Internal Effects using Neighbourhoods, All Adults
Dependent Variable: Natural Log of Cereal Crops Produced (LN_CEREAL)

	Eqn. 5	Eqn. 6
ED_AD	0.042 ***	0.044 ***
ED_NEIGH_AVG	-0.006	
ED_NEIGH_MAX		0.034 ***
FARM_VARIABLES	YES	YES
HH_VARIABLES	YES	YES
SITES	YES	YES
R^2 (unadj)	0.563	0.565
NUMBER OBS.	593	593

Note: Standard errors have been adjusted to account for the clustered nature of the data. Stars indicate significance using a two tailed t-test as follows: *** = 0.01; ** = 0.05; * = 0.10; ✔ = 0.20. For each household, own education is removed from the aggregated variables prior to calculation of ED_NEIGH_AVG and ED_NEIGH_MAX.

Table 5: Maximum Likelihood Estimation of the Stochastic Frontier Function Dependent Variable: Natural Log of Cereal Crops Produced (LN_CEREAL)

	ONE-STAGE FRONTIER	TWO-STAGE FRONTIER
	Eqn. 7	Eqn. 8
Stage 1		
CONSTANT	5.667 ***	5.705 ***
LN_LAND	0.294 ***	0.314 ***
LN_LABOUR	0.164 ***	0.156 ***
LN_CAPITAL	0.109 ***	0.103 ***
LN_INPUTS	0.007	0.014
OXEN	0.160 **	0.150 **
BANANA	0.186 ✔	0.185 ✔
COFFEE	0.017	0.059
CHAT	0.045	0.073
EUCALPTUS	0.140 *	0.157 **
ENSET	-0.021	0.029
GESHU	0.110	0.137
LN_FERT	0.020	0.021
MANURE	0.077	0.070
LN_QUALITY	-0.091	-0.117 ✔
LN_STEEPNESS	-0.125	-0.089
AGE_HHH	-0.007	-0.008
AGESQ_HHH	0.000	0.000
FEM_HEAD	-0.109	-0.123 ✔
NO_ADULTS	-0.009	-0.011
NONAGR_HHH	-0.066	-0.086
SITE_2	-1.066 ***	-1.074 ***
SITE_3	-0.503 ***	-0.475 ***
SITE_4	0.131	0.128
SITE_5	0.290	0.249
SITE_7	0.994 ***	0.971 ***
SITE_8	0.503 **	0.519 **
SITE_9	-0.605 ***	-0.623 ***
SITE_10	0.958 ***	0.826 ***
SITE_15	-1.153 ***	-1.273 ***
SITE_16	0.289 ✔	0.269 ✔
$SIGMA^2$	6.865 ***	17.371 ***
GAMMA	0.983 ***	0.993 ***
MU	-5.196 ***	
Stage 2		
CONSTANT		-16.008 ***
ED_AD		-2.135 ***
MEAN EFF.	0.540	0.557
LOG-LIKELIHOOD	-729.1	-709.7
NUMBER OBS.	593	593

Note: Stars indicate significance using a two tailed t-test: *** = 0.01; ** = 0.05; * = 0.10; ✔ = 0.20. The one-sided error term is assumed to have a truncated normal distribution. The statistics, $SIGMA^2$, GAMMA and MU, provide information on the appropriateness of using a stochastic frontier specification and the distributional form of the two-part error term (see Coelli 1994).

Table 6: Maximum Likelihood Estimation of the Frontier Production Function
Average Village Schooling, External vs. Internal Effect
Dependent Variable: Natural Log of Cereal Crops Produced

	ONE-STAGE FRONTIER		TWO-STAGE FRONTIER	
	Eqn. 9	Eqn. 10	Eqn. 11	Eqn. 12
ED_AD	0.009	0.001	-2.266 ***	-2.169 ***
ED_VILLAGE	0.584 ***	0.218 ***	-0.689 ✔	0.382
FARM_VARIABLES	YES	YES	YES	YES
HH_VARIABLES	YES	YES	YES	YES
SITES	NO	NO	NO	NO
VILLAGE VARIABLES	NO	YES	NO	YES
MEAN EFF.	0.509	0.534	0.518	0.550
LOG-LIKELIHOOD	-814.7	-740.6	-830.6	-726.2
NUMBER OBS.	593	593	593	593

Note: Stars indicate significance using a two tailed t-test: *** = 0.01; ** = 0.05; * = 0.10; ✔ = 0.20. In the Two-Stage Frontier, education variables are included in the second stage, and all other variables are included in the first stage. In the One-Stage Frontier, all variables are included together. The one-sided error term is assumed to have a truncated normal distribution.

Table 7: Maximum Likelihood Estimation of the Frontier Production Function
External vs. Internal Effects using Neighbourhoods, All Adults
Dependent Variable: Natural Log of Cereal Crops Produced

	ONE-STAGE FRONTIER		TWO-STAGE FRONTIER	
	Eqn. 13	Eqn. 14	Eqn. 15	Eqn. 16
ED_AD	0.010	0.009	-1.663 ***	-1.286 **
ED_NEIGH_AVG	0.178 *		0.019	
ED_NEIGH_MAX		0.015		0.007
FARM_VARIABLES	YES	YES	YES	YES
HH_VARIABLES	YES	YES	YES	YES
SITES	YES	YES	YES	YES
MEAN EFF.	0.540	0.540	0.554	0.555
LOG-LIKELIHOOD	-728.6	-728.2	-707.4	-707.6
NUMBER OBS.	593	593	593	593

Note: Stars indicate significance using a two tailed t-test as follows: *** = 0.01; ** = 0.05; * = 0.10; ✔ = 0.20. In the Two-Stage Frontier, education variables are included in the second stage, and all other variables are included in the first stage. In the One-Stage Frontier, all variables are included together. The one-sided error term is assumed to have a truncated normal distribution. For each household, own education is removed from the aggregated variables prior to calculation of ED_NEIGH_AVG and ED_NEIGH_MAX.

Table 8: Simulation (Based on an OLS Production Function)
Effect of Education in the Village, by Site
Dependent Variable: Natural Log of Cereal Crops Produced

	Average Education in the Village (Years per Adult)	Effect of Education in Village*	Percent Difference: Effect of Education in Site vs. Mean Effect for All Sites
SITE 2	0.17	5.35	-6.98
SITE 3	0.32	5.42	-5.84
SITE 4 - VILLAGE 1	1.23	5.80	+0.79
SITE 4 - VILLAGE 2	1.31	5.84	+1.39
SITE 4 - VILLAGE 3	1.97	6.11	+6.24
SITE 4 - VILLAGE 4	1.61	5.96	+3.61
SITE 5	1.13	5.76	+0.08
SITE 6	0.72	5.59	-2.90
SITE 7	1.94	6.10	+6.01
SITE 8	0.37	5.44	-5.50
SITE 9	1.11	5.75	-0.04
SITE 10	3.16	6.61	+14.93
SITE 15	1.30	5.83	+1.34
SITE 16	1.23	5.80	+0.84
SIM: ED_VILLAGE = 0	0	5.28	-8.20
SIM: ED_VILLAGE = 6	6	7.81	+35.72

Note: This analysis employs all 616 observations for whom the relevant information is available - slightly larger than the sample used previously since no information is needed on neighbourhood in the present context. Hence, there are fewer missing observations.
* The effect of education at the site-level is calculated as the predicted log of farm output with education at the site-level set to average years of schooling in each site in turn while all other values of variables are set to the sample means for the entire sample.

Table 9
Probit Estimation: Probability of Current Enrolment (marginal effects)
Dependent Variable: ENROLLED (1 if currently in school; else 0)

	Eqn. 17	Mean
Intra-household Allocation Variables - Child Attributes		
AGE_YRS - Age of child (years)	0.093	12.21
AGE_YRS2 - Age squared	-0.004	149
FEMALE - Dummy: 1 if child is female	-0.143 **	0.46
CHILD_HEAD - Dummy: 1 if child of household head	0.223 ***	0.83
BIRTH_ORDER - Sibling order: Oldest = 1, …, youngest = n	-0.035 ✔	2.44
LIVE_PARENTS - Dummy: 1 if parents plan to live with child later	-0.133 *	0.18
Household Variables		
CONSAE - Consumption per adult equivalent (Birr)	-0.000	97.07
FEMALE*CONSAE - Interaction term: sex and income	0.001 **	44.85
HH_COST - School fees as a percentage of consae	-0.083	0.21
LAND_PA - Land per adult household member (hectares)	0.574 ✔	0.28
NUM_LT7 - Number of children aged under 7 in the household	-0.057 *	1.68
NUM_GT59 - Number of adults aged over 59 in the household	-0.062	0.28
REL_MUSLIM - Dummy: 1 if Muslim	-0.292 ***	0.21
REL_XIAN - Dummy: 1 if other Christian religion	0.045	0.43
Parental Attributes		
EDUC_H - Dummy: 1 if head has any education	-0.209 *	0.33
EDUC_S - Dummy: 1 if spouse has any education	0.008	0.23
COGNITIVE_H - Total cognitive skills score (0-17) - Head	0.011 ✔	4.27
COGNITIVE_S - Total cognitive skills score (0-17) - Spouse	-0.005	1.62
RAVEN_H - Score on Raven's test of reasoning ability (5-36) - Head	0.008 **	19.09
RAVEN_S - Score on Raven's test of reasoning ability (5-36) - Spouse	0.009	18.04
QUALITY_1_H - Dummy: 1 if school quality rating equals 1 - Head	0.067	0.05
QUALITY_2_H - Dummy: 1 if school quality rating equals 2 - Head	-0.280 ***	0.07
QUALITY_3_H - Dummy: 1 if school quality rating equals 3 - Head	-0.275 ***	0.05
QUALITY_4_H - Dummy: 1 if school quality rating equals 4 - Head	-0.340 ***	0.29
QUALITY_MISS_H - Dummy: 1 if missing school quality rating - Head	-0.269 **	0.30
MODERN_AGR_H - Dummy: 1 if head not traditional in farming views	0.214 ***	0.53
BENEFIT_AGR_H - Dummy: 1 if believe educ improves farming - head	0.181 ***	0.59
LIK_ATT_SCH_H - Score on likert attitude index - Head	0.022 ***	12.36
LIK_ATT_SCH_S - Score on likert attitude index - Spouse	0.001	11.65
Site Fixed Effects Variables		
SITE B - Dummy: 1 if household in site B	0.738***	0.36
SITE C - Dummy: 1 if household in site C	0.480 *	0.20
SITE D - Dummy: 1 if household in site D	0.186	0.27
Pseudo R^2	0.29	
Chi-Squared (2)	2.03	
Log Likelihood	-179.1	
Number of Observations	387	

Note: Standard errors have been adjusted to account for the clustered nature of the data. Stars indicate significance using a two tailed t-test as follows: *** = 0.01; ** = 0.05; * = 0.10; ✔ = 0.20.