Whose Education Matters in the Determination of Household Income? Evidence from a Developing Country*

Dean Jolliffe
Economic Research Service, U.S. Department of Agriculture

I. Introduction

In developed countries, where the majority of workers are wage earners, the returns to human capital are typically measured by regressing an individual's wage on their years of schooling. The human capital literature for developing countries is similarly focused on measuring the returns to education for wage earners, in spite of the fact that in most of these countries wage earners are a relatively small fraction of the labor force.[1] A predominant feature of many developing countries is that the largest share of the labor force is engaged in self-employed activities that generate income for households—either as farm households or as small enterprises.[2] The different composition of labor forces in developed and developing countries has important implications for the way income data are collected in both types of countries. Income generated from farming or other household enterprises is almost always measured at the household level, whereas wage income is available at the individual level.

This difference in the ways income data are collected makes it difficult to extend the wage regression model to the developing country context because in developing countries income is largely measured at the household level; the data on education attainment, however, are available at the individual level. As survey data rarely allow for the decomposition of household income to the individual level, it is not possible to map an individual's education attainment to their contribution to household income. A seemingly natural extension of the wage regression model for this situation would be to regress the household's income on the household's education level. This extension, though, leads to the difficulty of how to model the household's education level and poses the question, whose education matters?

The answer to this question depends on the way in which education affects a household's production. In pioneering work, T. W. Schultz and F.

0013-0079/2002/5002-0004$10.00

Welch both address this question.[3] Schultz proposes that education improves a household's ability to efficiently adjust production decisions during periods of disequilibria or change. Along similar lines, Welch suggests that education may have two distinct effects. First, education may enhance a worker's ability to produce more with the given resources, which he refers to as the "worker effect." Second, education may improve the worker's ability to select the mix of inputs, which Welch refers to as the "allocative effect" (p. 42). In more recent work, M. Kremer models heterogeneity in the quality of labor inputs in production functions, and he proposes a weak-link production function in which workers of similar skill are matched together.[4]

The model of how education affects production has implications for the appropriate empirical specification of education in the production process. D. T. Yang proposes that using the maximum value of education in the household will serve as a reasonable proxy for the allocative effect, while the average level of education will proxy the worker (or productivity) effect.[5] The allocative effect will likely be captured in managerial decisions, which are presumed to be made by the best-educated individual in the household. The natural implication of Kremer's weak-link model is that the minimum value of education in the household is the appropriate measure of household education, as it is this weakest link that will ultimately determine the value of the output.

Except for the work of Yang, this connection between a theoretical model of education in the production process and the empirical specification is generally not discussed. The large majority of the empirical literature on household income and education use either the education level of one individual in the household or the average level of education in the household. D. T. Jamison and L. J. Lau's survey of the literature on schooling and household farm income discusses the results of 37 studies from Asia, Africa, and Latin America.[6] Of these studies, 14 use the average level of education, 20 use the education level of the household head or the farm operator, and one uses the education level of the wife of the household head (in two the measure is unspecified). Similarly, studies by G. Fane, C. C. Wu, and Jamison and P. R. Moock use the education level of the head of household to represent the school attainment for the household, whereas W. E. Huffman extends this slightly and uses the education levels of the head of household and the spouse.[7] J. Y. Lin examines whether the education level of the head of household and the average level of household education affect adoption of new farm technologies, while A. Foster and M. Rosenzweig show that whether anyone in the household has primary education is an important predictor of adopting new farming technology.[8] Regardless of the measure of education that is used, most studies do not discuss the motivation for the choice, leaving the impression that the decision is ad hoc.

The purpose of this article is to empirically consider three existing models of how household education affects household income. I use the minimum, maximum, and average level of household schooling to test whether education affects the production of household income through a weak-link, allocative, or worker effect. I decompose total household income into farm and nonfarm income to examine whether the effect of education varies when considering source of income.

The plan of this article is as follows: Section II discusses the data used and presents a few descriptive statistics on headship in Ghana and labor activities of Ghanaian households. Section III presents three simple paradigms about whose education matters in regression models of household income. This section also describes the composition of household income and proposes the appropriate explanatory variables for estimation. Section IV describes the estimation methodology. Consistent estimates for the farm and off-farm income functions are obtained by using J. L. Powell's censored least absolute deviations (CLAD) estimators.[9] Standard errors are obtained by using a bootstrap methodology that mimics the two-stage sample design. Section V contains a summary of the estimation results. Most notably, the results suggest that the allocative effect is the largest effect for total income, while the productivity effect appears to dominate when considering farming and nonfarming activities separately. All specifications of household income and its components reject the use of the minimum level of household schooling as well as the schooling level of the head of the household. Section VI provides some concluding comments.

II. Data and Descriptive Statistics

The data used in this article are from the Ghana Living Standards Survey (GLSS), a nationwide household survey carried out by the Ghana Statistical Service with technical assistance from the World Bank. The survey, administered from October 1988 to September 1989, covers 3,200 households and contains detailed information on formal and informal labor activities, household farm activities, expenditures, education status of household members, and many other determinants of household welfare.[10] A supplemental education module was also administered to a nationally representative subsample of 1,585 households. This article uses the data from the subsample of 1,585 households. I dropped 38 of these households because of missing data, which resulted in a sample of 1,547 households.[11]

Because the majority of previous studies use the education level of the head of household to model the impact of education on household-level income, it is worth considering this choice. Presumably the education level of the head is selected, as it is assumed both that headship is positively correlated with higher levels of education within the household and that the head is the person making the decisions affecting household income. In the case of the GLSS data neither of these assumptions appears to be supported.

In the GLSS data, as with many household surveys, the "head of household" title is a self-ascribed characteristic. The household members are asked to identify a head of the household but are not provided with guidelines to do so. In 91% of the GLSS households, headship is ascribed to the oldest

TABLE 1

HEAD OF HOUSEHOLD'S EDUCATION ATTAINMENT, INTRAHOUSEHOLD COMPARISONS BY HOUSEHOLD SIZE

Number of Adults in Household	Head Is Most Educated Adult (%)	Head Is Least Educated Adult (%)	Gap Max	Gap Min	Number of Households
1	100	100			488
2	73	60	1.6	2.4	543
3	59	55	2.8	3.2	265
4	50	63	3.8	2.7	121
5	39	61	4.6	3.4	72
> 5	29	66	6.2	2.3	58

NOTE.—An adult is defined as 15 years of age or older. If the head of household is tied with another member for the most or least educated in the household, then the head is treated as most or least educated adult in this table. Gap max is the average difference in years between the head of household's education attainment and the maximum education attainment in the household. Gap min is the average difference in years between the head of household's education attainment and the minimum education attainment in the household.

member in the house, not necessarily the person making the household business decisions. The GLSS data also show that while the education level of the household head is often the highest in the household, it is also frequently the lowest. Table 1 presents the percentage of households in which the head of household has the highest and the lowest level of education in the household. Table 1 shows that as household size increases, the percentage of households where the head has the most education declines and the gap between the education level of the head and the most-educated household member increases.[12]

In Ghana there is a great amount of diversity in labor activities. There are 3,698 laborers in this sample of 1,547 households. Of these individuals, 77% spend some time working on a farm and 56% spend some time engaged in some other form of labor. Of these individuals, 34% are engaged in both some farming and nonfarming activities. The extent of labor diversification increases when considering all labor within the household. In 55% of the households, at least one person has spent some time working on a farm and at least one person has spent some time in nonfarm labor.

The human capital literature for developing countries estimates the returns to education by either estimating a Mincer-type wage equation or by estimating some form of a farm production or profit function. The literature typically assumes that individuals or farm households are engaged in only one income-generating activity. Because so many households and individuals are engaged in more than one income-generating activity, estimating the returns to education by focusing on either strictly farm income or wage income may provide an incomplete picture of the role of education. The hypothesis I examine in this article is that in terms of total income, education may improve the choices made in allocating household resources to each income-generating activity;

however, when examining each activity separately, the primary benefit of education may be the improved productivity effect. For this reason, this article examines total household income, farm household income, and off-farm household income.

III. Empirical Specification

A. Household School Attainment

In order to answer the question of how education affects household income, I consider a basic specification that includes the minimum, average, and maximum value of school attainment within each household. The three models, tested alternatively, assume that one of these terms is a determinant of household income, while the other two have no effect on household income. I then extend the model in two ways. First, I also include the education level of the head of household (along with minimum, average, and maximum schooling in the household) in the set of school regressors. I examine the assumption that only the head's education level matters by testing both whether the head's education level is a significant explanatory variable of income and whether the other variables are jointly equal to zero. The second extension to the model considers whether the results differ when the median level of education is used in lieu of the household average level of school attainment.

The basic model of the log of total income for household j is[13]

$$\ln(Y_j) = \alpha_0 + \alpha_1 \mathbf{X}_j + \sum_i \beta_i S_{ij} + \varepsilon_j \quad i \in \{\min, \text{avg}, \max\}. \tag{1}$$

In this expression, $S_{\min,\ j}$, $S_{\text{avg},\ j}$, and $S_{\max,\ j}$ are, respectively, the minimum, average, and maximum level of schooling within household j; $\mathbf{X}$ is a vector of the other explanatory variables.[14] The elements of $\mathbf{X}$ are discussed later in this section.

The weak link—household minimum. The specification of household income given in equation (1) nests a few simple paradigms of how school attainment within a household affects the determination of household income. The first considered is the paradigm that only the household's minimum value of education attainment matters in the determination of household income. This model is motivated by the popular management aphorism that a production process is only as good as its weakest link, and it is most appropriate for production processes where one bad input or one mistake significantly lowers the value of the output. If the weak-link paradigm is the correct model, there are two testable implications from estimating equation (1). The first is that the parameter estimates on the average and maximum level of schooling, β_{avg} and $\beta_{\max}$, are jointly equal to zero. The second implication is that the parameter on the minimum level of schooling, $\beta_{\min}$, is positive.

Of the three paradigms considered in this article, the weak link most closely resembles the head-of-household model, at least in the case of Ghana, where the head of household has the lowest level of education in 60% of the households with two or more adult members (or 72% of all households when including those households with only one member). In Ghana, the elderly have the least amount of education, and the oldest person is often designated the head by the household.[15]

Allocative effect—household maximum. Another paradigm considered here is that education improves the decisions of how to allocate household resources to the various activities, and this is modeled by using the household's maximum value of education. The model is motivated by the notion that it is the choices a household makes that critically require education, and not necessarily the quality of the implementation of those choices.

If the allocative effect of household education is the correct paradigm, there are two testable implications that are very similar to those for the weak-link model. The implications are that the parameter estimates from estimating equation (1) on the minimum and average level of schooling, β_{min} and β_{avg}, are jointly equal to zero, while the term for the maximum level of schooling, β_{max}, is positive.

Productivity effect—household median and household average. The final paradigm of how household schooling affects household income is that schooling improves the productivity of each person working in carrying out his or her tasks. For this model the skills of all workers are important for the creation of household income, and one good manager or one weak link is not the driving force. To capture this effect empirically, I use the average and the median level of schooling within each household to represent the central tendency of the household's distribution of schooling. While the median and the average values of school attainment are very similar, the estimation results differ slightly, and I present a summary of both. If the productivity effect of schooling is the correct paradigm, then the parameter estimates on the minimum and maximum level of schooling are jointly equal to zero, while the parameter on the average or median level of schooling is positive.

One important assumption placed on the way that schooling is specified in all three models is that households cannot hire educated laborers. It is assumed that the market for educated laborers is imperfect, and households are unable to purchase the profit-maximizing level of schooling by hiring educated managers. There is some evidence in the GLSS data suggesting both that markets for educated laborers are not very active and that this assumption is reasonable. For example, the average Ghanaian farm household spends less than 5% of farm income on hiring in labor. Of the labor that is hired in, 70% of the wages go for clearing land, which is not an education-intensive job.

One problem in estimating the models of school attainment is that they require regressing household income on three measures of school attainment, which are highly correlated. This is primarily the case because there are numerous households with only one or two adult members. In households with only one adult member, the minimum, average, and maximum level of school attainment will all be the same value. In order to reduce the level of correlation across the variables, all models are estimated over three samples:

(1) all households, (2) all households with two or more adult members, and (3) all households with three or more adult members.[16]

It is also important to note that human capital is a complex, multidimensional characteristic, and school levels will capture certain aspects of it but are also likely to confound human capital with other characteristics such as wealth or innate ability. These issues are only dealt with here to a limited extent. For example, to control for some forms of omitted variable bias, the regression models include household composition variables and some information on household assets. The household composition variables will control for differences in education levels, which are correlated with income levels but result from gender differences.[17]

The difficulties associated with using school attainment to measure human capital are substantial, so it is worthwhile to note that the primary purpose of this article is not to measure the returns to human capital. Rather, I compare different measures of household school attainment to determine which measures best explain income and to determine if the standard practice of using the head of household's schooling as the measure of the household's total level of schooling is valid.

B. Household Income

To test the three paradigms discussed above, I estimate three separate household income functions—total income, farm income, and off-farm income. This strategy explicitly acknowledges that many households are engaged in numerous income-generating activities, and while a household member's education level may not matter for the determination of income in one activity, it may matter for another. Below is a description of the components of total household income.

Farm profits. I measure farm output as the value of all crops and animal products marketed in the last 12 months plus the value of crops kept for seed and given away as gifts.[18] Because many farmers cover their subsistence needs from their own production, the estimated value of home consumption of food and animal products is also included in the measure of total farm output.[19] Subtracted from this measure of farm output are expenditures on seed, fertilizer, insecticide, pesticide, livestock, storage, transportation, rented-in land, and hired-in labor. Crops given as payments for other inputs are also subtracted from the value of farm output. The resulting figure is a measure of profit, conditional on the quantity of land and labor. (The value of one is added to farm profit to allow the log transformation of farm profit.) The log of farm profit is modeled as a type 1 Tobit model.[20] In this model,

$$\begin{aligned} \ln(Y_f^* + 1) &= \ln Y_f^*(A_f,\ L_f,\ p_f,\ S,\ \mu) \\ \ln(Y_f + 1) &= \ln(Y_f^* + 1) \quad \text{if} \quad Y_f^* + 1 > 1 \\ &= 0 \ \text{if} \quad Y_f^* + 1 \leq 1, \end{aligned} \tag{2}$$

where the *f* subscript denotes a farm variable; A_f is the log of acres of land cultivated and years of farm experience, both of which are treated as fixed inputs; L_f is the log of household farm labor hours; p_f is a vector of prices for farm products and farm inputs; S is the vector of household school attainment variables discussed above; and μ is an error term.[21] It is assumed that Y_f is observed for all households, but Y_f^* is only observed if $Y_f^* > 0$. About 70% of the households engage in some farming activities, and the average restricted profit of those farming households is 144,604 Cedis.

Off-farm and total income. The measure of off-farm income, Y_o, aggregates wage income and self-employment income. The decision to aggregate these two loses some information but helps focus on the difference between farm and nonfarm income. The measure of wage income adjusts the wage rate by including all pecuniary remuneration for the labor supplied, including commissions, bonuses, tips, allowances, and gratuities. Wage income is also adjusted to reflect the value of all nonpecuniary payments, including remuneration in the form of food, crops, animals, housing, clothing, transportation, or any other form.

The measure of off-farm, self-employed income is recommended by W. Vijverberg and is the reported amount of money left over from self-employed business activities after expenses have been incurred.[22] This measure has the advantage of resulting in strictly nonnegative values. The log of off-farm income is also modeled as a type 1 Tobit:

$$\begin{aligned} \ln(Y_o^* + 1) &= \ln Y_o^*(A_o, L_o, S, \varepsilon) \\ \ln(Y_o + 1) &= \ln(Y_o^* + 1) \quad \text{if} \quad Y_o^* + 1 > 1 \\ &= 0 \ \text{if} \quad Y_o^* + 1 \leq 1, \end{aligned} \tag{3}$$

where the *o* subscript denotes an off-farm variable, A_o is the log of business assets and years of work experience, L_o is the log of household off-farm labor hours, S is the vector of household school attainment variables discussed above, and ε is an error term. It is assumed that Y_o is observed for all households, but Y_o^* is only observed if $Y_o^* > 0$. (The value of one is also added to off-farm income to allow the log transformation of farm profit.) About 67% of the households have at least one household member who engages in some form of off-farm work, and the average off-farm income for these households is 373,143 Cedis. Total household income is modeled simply as the sum of farm and off-farm income. The average value of total household income is 326,743 Cedis.

Farm and off-farm labor. Both the farm and off-farm income functions include hours of household labor, the levels of which are chosen by the household. To correct for the likely case that this endogenous variable will bias the estimated school effect, I model farm and off-farm labor as functions of household size, gender composition of the household, wages, and the relative productivity of labor in farm and off-farm activities.[23] Household size and gender composition enter the labor functions primarily because the total household level of labor supply, not the individual level of labor supply, is being modeled.[24] This model of labor supply assumes that labor markets are not perfect and that the principle of separation does not hold. As I have already noted, the average farm household spends less than 5% of the farm income on hiring in outside labor, which suggests that formal labor markets are not very active. This model of labor supply allows equations (2) and (3) to be rewritten as, for farm income:

$$\begin{aligned}\ln(Y_f^* + 1) &= \ln Y_f^*[A_f,\ L_f(Y_f^*,\ Y_o^*,\ X_h,\ \omega_f,\ \omega_o),\ p,\ S,\ \mu] \\ &= \ln Y_f^*(A,\ X_h,\ \omega,\ p,\ S,\ \mu); \qquad (4)\end{aligned}$$

for off-farm income:

$$\begin{aligned}\ln(Y_o^* + 1) &= \ln Y_o^*[A_o,\ L_o(Y_f^*,\ Y_o^*,\ X_h,\ \omega_o,\ \omega_f),\ S,\ \varepsilon] \\ &= \ln Y_o^*(A,\ X_h,\ \omega,\ p,\ S,\ \varepsilon), \qquad (5)\end{aligned}$$

where A is the vector of fixed farm and off-farm inputs, X_h represents household characteristics, ω is a vector of farm and off-farm wages, p is a vector of farm input and output prices, S is the vector of household school attainment variables discussed above, and μ and ε are error terms. It is again assumed that Y_f and Y_o are observed for all households, but Y_f^* and Y_o^* are only observed if positive.

IV. Estimation

Most household survey data are fraught with violations of the assumptions made for the classical linear regression model. The GLSS data are no exception. Three important sources of the violations are censored observations, outlier values, and two-stage sample design. The estimation methods I use are sensitive to these three factors. The regression estimates presented here are either least absolute deviations (LAD) estimators or censored least absolute deviations (CLAD) estimators. The standard errors used for these estimators are bootstrap estimates, which are derived by replicating the two-stage sample design.

A. Censored Dependent Variables

Roughly 70% of all Ghanaian households are engaged in farming activities, and, similarly, about 67% of the households generate some of their household income from off-farm activities. I avoid modeling the selection rule determining who farms and who does not, and I simply treat farm and off-farm income as data that are censored at zero.[25]

The problem introduced by censoring is that OLS results in biased estimators, and the standard Tobit or Heckman estimators for censored models rely heavily on the assumption of normality. A. Arabmazar and P. Schmidt show that the bias resulting from the Tobit estimator in the presence of heteroscedastic residuals can be large.[26] W. Vijverberg presents similar results showing that the bias of the Tobit estimator is also large when kurtosis and skewness are nonnormal.[27] Powell's CLAD estimator results in consistent estimates for the limited dependent variable model in the presence of many violations of normality, including heteroscedasticity and nonnormal kurtosis.

B. Outliers and Other Violations of Normality

A standard feature of many household data sets is the presence of unusually large or small values. When the dependent variable in a model contains outlier values, it is likely that the resulting residuals will be nonnormally distributed. Consider, for example, the total income variable used in this article. The mean value for total household income is 326,743 Cedis, the standard error of total income is 10 times this size, and the maximum value is 140 million Cedis. The residuals from estimating the log of total household income exhibit large values of kurtosis and strongly violate the assumption of normality.[28] The advantage of the LAD and CLAD estimators used here is that they are less sensitive to outliers than OLS and are robust to violations of kurtosis.[29] For this reason I exclude no extreme data points from the sample and avoid arbitrary selection rules.

C. Complex Sample Design and the IID Assumption

As with essentially all nationwide household surveys, the design of the GLSS sample is not a simple random draw of households. The GLSS sample is a two-stage design in which the first stage entails randomly selecting numerous clusters or geographic regions; in the second stage a fixed number of households are randomly selected within each cluster. The primary concern with data from a two-stage sample design is that they are likely to result in residuals that are neither identically nor independently distributed (IID). This is because households within a specific cluster are likely to be more similar to each other than to households in other clusters. The result of this is that intracluster variation is likely to be significantly different from intercluster variation of the residuals, and this clustering of the residuals will lead to a violation of the assumption of homoscedasticity. Numerous papers illustrate this point and show that the violation of IID residuals can have large effects on estimated parameters and standard errors.[30]

I use the Breusch-Pagan test to check the assumption of homoscedastic residuals from the total income regression model, and it strongly rejects the assumption of homoscedasticity.[31] The Breusch-Pagan test cannot be used directly to examine the assumption of homoscedasticity for the farm and off-farm income models because a necessary condition of the test is that the vector of errors has an expected value of zero. Since both farm and off-farm income are censored at zero, the residuals from OLS estimation will not have an expected value of zero in the presence of heteroscedasticity. A. Pagan and F. Vella propose a modified version of the Breusch-Pagan test that first constructs "generalized" residuals by using any consistent estimator and then employs the standard Breusch-Pagan test.[32] The Pagan-Vella test statistics from the household farm and off-farm income regressions strongly reject the assumption of homoscedasticity in these models.[33]

W. Rogers suggests the use of bootstrap standard errors to correct for heteroscedasticity, but this does not address the problem of correlated residuals.[34] The standard bootstrap fails to correct for dependent residuals, because the bootstrap method typically redraws samples using purely random selection. Since the sample design is not a pure random sample, this bootstrap method will not accurately reflect the characteristics of the data. In order to correct for both heteroscedasticity and the dependence of the residuals, I use a bootstrap procedure that replicates the sample design. In the first stage of bootstrap resampling, clusters are randomly selected; in the second stage, households are drawn in each of the selected clusters.[35]

Using this method, each household does not have an equal probability of being chosen; rather, there is a dependence created in the resampling such that if one household is selected in a cluster, then the probability of selection for the other households in that cluster increases. By following this method, the bootstrap samples exhibit the same characteristics as the initial sample and the estimated standard errors are robust to violations of the assumptions that the residuals are identically and independently distributed. In contrast, the bootstrap that redraws from the sample following a pure simple random draw will be robust to violations of identically distributed residuals but not to violations where the residuals are independently distributed. In the specifications I estimate, the standard error for the estimated marginal impact of schooling on total household income increases by 69% when comparing a bootstrap following a pure random draw with a bootstrap procedure that replicates the sample design.[36] See B. Efron and R. Tibshirani for a general discussion of the bootstrap and A. Deaton for a brief discussion on using the bootstrap to replicate sample design.[37]

D. LAD and CLAD Estimators

To estimate total household income, I use the LAD estimator. The properties of this estimator are presented in R. Koenker and G. Bassett.[38] The LAD estimator is found by minimizing:

$$\sum |y_i - x_i'\beta|. \tag{6}$$

To estimate farm and off-farm income, both of which are censored at zero, I use Powell's CLAD estimator. This estimator provides consistent estimates for the censored model when heteroscedasticity is present as well as other violations of normality and is robust to outliers.[39]

The CLAD estimator is found by minimizing:

$$\sum |y_i - \max(0,\ x_i'\ \beta)|. \tag{7}$$

The consistency of this estimator rests on the fact that medians are preserved by monotone transformations of the data, and equation (7) is a monotone transformation of equation (6), the standard median regression.

The estimation technique I use for the CLAD estimator is M. Buchinsky's iterative linear programming algorithm (ILPA).[40] The first step of the ILPA is to estimate a quantile regression for the full sample, then to delete the observations for which the predicted value of the dependent variable is less than zero.[41] Another quantile regression is estimated on the new sample, and again negative predicted values are dropped. Buchinsky shows that if the process converges, then a local minimum is obtained.[42]

V. Results

This section first summarizes the results from testing the weak-link, allocative, and productivity paradigms of schooling as well as from examining the use of the schooling of the household head to predict income. Each of the models is tested for the three measures of household income (total, farm, and off-farm).[43] In this section, I then summarize results from estimating household income separately using the maximum, average, and household head's level of school attainment. In conclusion, I discuss an extension to the model of schooling, which incorporates gender.

A. Tests of Household School Attainment Models

The two conditions tested for the weakest link model state that the minimum value of schooling is the only school variable that has a statistically significant effect on household income. The null hypothesis tested is that the minimum level of schooling has no effect, and rejection of this is taken as evidence supporting the weak-link model.[44] The other hypothesis is that the average and maximum levels of schooling have no effect on income, and failing to reject this joint hypothesis supports the weak-link model. In discussing the results, rejecting a model means that one of the two tests are not supportive of the model, strongly rejecting ("reject both") means that both tests fail to support the model, and failing to reject means that the results from testing both hypotheses support the model.

Table A1 presents the full results from estimating household income using minimum, average, and maximum levels of household schooling. Table 2 summarizes table A1 by presenting the *p*-values from testing each of the school models. (This table also shows the estimated schooling parameters.) Table 3 presents the summary test statistics for each of the schooling models using the minimum, median, and maximum level of schooling.[45]

The test results presented in these tables present a strong argument against the weak-link and household-head models.[46] In all cases of estimating total, farm, or off-farm income and in all cases over the three samples, both models are rejected. As a large percentage of household heads are the least educated household member, it is perhaps not too surprising that the results from testing these two models are similar.

The results from testing the maximum and average values of schooling

TABLE 2

HOUSEHOLD INCOME AND SCHOOLING—TESTS OF MINIMUM, AVERAGE, AND MAXIMUM SCHOOLING

	TOTAL INCOME		FARM INCOME		OFF-FARM INCOME	
	LAD Estimate	SE	CLAD Estimate	SE	CLAD Estimate	SE
Household minimum level of schooling	.016	(.0340)	−.071	(.0852)	−.042	(.1323)
Household average years of schooling	−.016	(.0608)	.285*	(.1613)	.434*	(.2457)
Household maximum level of schooling	.036	(.0306)	−.084	(.0798)	−.158	(.1340)
	Joint Tests of Schooling Models, *p*-Values					
Household minimum:	Reject		Reject both		Reject	
Condition 1 (average and maximum = 0)	.13		.10		.11	
Condition 2 (minimum = 0)	.63		.41		.75	
Household maximum:	Reject		Reject both		Reject both	
Condition 1 (minimum and average = 0)	.83		.10		.00	
Condition 2 (average = 0)	.24		.29		.24	
Household average:	Reject		Fail to reject		Fail to reject	
Condition 1 (minimum and maximum = 0)	.42		.57		.36	
Condition 2 (average = 0)	.79		.08		.08	
Head of household:	Reject		Reject		Reject both	
Condition 1 (minimum average and maximum = 0)	.73		.54		.07	
Condition 2 (head = 0)	.39		.47		.64	

NOTE.—The first three rows are the regression parameters and are presented in full in table A1. The remaining statistics are *p*-values from testing the described conditions. Evidence supporting a model appears as a large *p*-value for condition 1 and a small *p*-value for condition 2. “Reject” means that one of these two conditions fails to support the model. “Reject both” means both conditions fail. “Fail to reject” means that both conditions provide evidence supporting the model. The sample is all households with two or more members 15 years of age or older.

* Parameter estimates if the *p*-value is less than .1.

are somewhat mixed. When the minimum, average, and maximum levels of schooling are used, the data reject the allocative model of education (see table 2). This contrasts with the results presented in table 3, which support the allocative model for predicting total household income. The only difference between these sets of tables is that I use the median level of schooling in table 3 in place of the average level of schooling. The difference between these two sets of results comes from the difference in the standard errors of the parameter estimates.[47] When the median level of schooling is used instead of the average level of schooling, the estimated effect of the maximum level of schooling is much more precisely estimated. This difference may be partially due to the fact that even though the median and the average levels of schooling are very similar, the median level of schooling is not as highly

TABLE 3

HOUSEHOLD INCOME AND SCHOOLING—TESTS OF MINIMUM, MEDIAN, AND MAXIMUM SCHOOLING

	TOTAL INCOME		FARM INCOME		OFF-FARM INCOME	
	LAD Estimate	SE	CLAD Estimate	SE	CLAD Estimate	SE
Household minimum level of schooling	.011	(.0207)	.010	(.0506)	.088	(.0756)
Household median years of schooling	−.007	(.0276)	.121*	(.0646)	.173	(.1116)
Household maximum level of schooling	.032*	(.0175)	−.005	(.0412)	−.028	(.0764)
	p-Values					
Household minimum:	Reject		Reject both		Reject	
Condition 1 (median and maximum = 0)	.11		.07		.15	
Condition 2 (minimum = 0)	.59		.85		.24	
Household maximum:	Fail to reject		Reject both		Reject both	
Condition 1 (minimum and median = 0)	.86		.10		.00	
Condition 2 (maximum = 0)	.07		.90		.71	
Household median:	Reject		Fail to reject		Reject	
Condition 1 (minimum and maximum = 0)	.19		.97		.42	
Condition 2 (median = 0)	.81		.06		.12	
Head of household:	Reject		Reject		Reject	
Condition 1 (minimum, median, and maximum = 0)	.72		.37		.12	
Condition 2 (head = 0)	.42		.42		.60	

NOTE.—The first three rows are the regression parameters and the remaining statistics are *p*-values. The only difference between the model in table A1 and this model is that the median level of schooling is used here instead of the average. See table 2 for description of "Reject," "Reject both," and "Fail to reject." The sample is all households with two or more members 15 years of age or older.

* Parameter estimates if the *p*-value is less than .1.

correlated (as the average level of schooling) with the maximum level of schooling.

The summary results presented in table 2 show some support for using the average level of schooling when predicting farm and off-farm income. Similarly, the results presented in table 3 show some support for using the median level of schooling when estimating farm and off-farm income. The data fail to reject the productivity-effect paradigm when estimating farm income using all households, and when using all households with two or more adult members. Similarly, the data fail to reject this paradigm when estimating off-farm income using the sample of households with two or more adults, and when using households with three or more adults. These results indicate that education appears to affect the components of income (farm and off-farm) through a productivity effect (as proxied by household average and

TABLE 4

HOUSEHOLD INCOME AND SCHOOLING—MINIMUM, AVERAGE, MEDIAN, AND MAXIMUM
(Households with Three or More Adults)

	TOTAL INCOME		FARM INCOME		OFF-FARM INCOME	
	LAD Estimate	SE	CLAD Estimate	SE	CLAD Estimate	SE
Specification 1:						
Household minimum level of schooling	.030	(.0312)	−.033	(.0849)	.014	(.0993)
Household average: years of schooling	−.005	(.0264)	.062	(.0684)	.187*	(.1110)
Household maximum level of schooling	.042*	(.0229)	.012	(.0502)	.090	(.1027)
Specification 2:						
Household minimum level of schooling	.026	(.0306)	−.048	(.0841)	.027	(.0946)
Household median years of schooling	.000	(.0253)	.062	(.0593)	.176*	(.1010)
Household maximum level of schooling	.041*	(.0223)	.019	(.0469)	.104	(.1027)

NOTE.—This table contains summary results from two specifications of household income. In both specifications, total, farm, and nonfarm income are regressed on a set of explanatory variables including three measures of education. The only difference between the two specifications is that in one, average schooling is used and in the other, the median level of schooling is used. The sample is all households with three or more members 15 years of age or older.

* Parameter estimates if the *p*-value is less than .1.

median schooling), while education appears to improve total income through an allocative effect (as proxied by the household maximum value of schooling).

In addition to estimating the three sources of income on the sample of all households with two or more adults, table 4 presents a summary of the regression results from using the sample of all households with three or more adults. The household measures of average and median level of schooling are the median and average of all members who are neither the maximum nor the minimum. So, in a three-adult household, the school attainment of the individual who has neither the highest nor the lowest level of schooling will represent the household average and median. The advantage of constructing measures in this fashion and in using only those households with three or more adults is reducing the level of collinearity of the schooling variables. The disadvantage of using this sample is that the sample size drops considerably, the precision of the estimates declines accordingly, and the ability to distinguish across the paradigms is diminished. Nonetheless, table 4 illustrates some similarities to tables 2 and 3 in that maximum level of schooling is a significant determinant of total income, but neither of farm income nor of off-farm income. Similarly, table 4 shows that the average and median level of schooling is an important determinant of off-farm income but not of total income. The result that average schooling is a significant determinant of farm

TABLE 5

Comparison of Parameter Estimates—Household School Attainment Measures and Household Income

	Total Household Income		Farm Income		Off-Farm Income	
	LAD Estimate	SE	CLAD Estimate	SE	CLAD Estimate	SE
Household head: schooling	.037***	(.0114)	.097***	(.0330)	.194***	(.0390)
Household minimum: schooling	.038***	(.0121)	.096**	(.0423)	.215***	(.0442)
Household maximum: schooling	.047***	(.0125)	.069**	(.0316)	.185***	(.0395)
Household average: schooling	.052***	(.0140)	.118**	(.0467)	.250***	(.0465)

Note.—Each parameter in the table results from separately estimating the school effects. The full regression results are presented in D. Jolliffe, "Cognitive Skills, Schooling, and Household Income: An Econometric Analysis Using Data for Ghana" (Ph.D. diss., Princeton University, 1996). The standard errors are estimated with 500 replications of the two-step, bootstrap procedure described in the article. The sample includes all households with at least one member 15 years of age or older.

** Parameter estimates if the *p*-value is less than .05.

*** Parameter estimates if the *p*-value is less than .01.

income does not follow through in table 4 and suggests that this result is sensitive to the smaller sample size used in this exercise.

B. Comparison of Head, Minimum, Average, and Maximum Estimates

Table 5 presents a summary of results from estimating the log of household income (total, farm, and off-farm income) separately using the household minimum, average, maximum, and head of household's school level. For example, the parameter estimate for the head's schooling in the total income column results from regressing total income on only the head's schooling and the other nonschool explanatory variables. The results from tables 2 and 3 suggest that either the average or maximum values of schooling serves as a better measure of household school levels (depending on whether total income or its components are being estimated). By examining each measure of schooling separately, table 5 presents evidence that the estimated return to schooling varies fairly dramatically depending on which measure of schooling is used.

The estimated return to schooling from using the maximum level of schooling to predict total income is 27% higher than from using the head's schooling. The estimated return to schooling from using the average household level to predict farm and off-farm income is 22% and 29% higher than from using the head's schooling.[48] It is important to note, though, that raising average level of schooling in the household by 1 year requires increasing total school attainment by more than would be required to raise just the education level of the most-educated individual by 1 year. Table 5 also shows that the

TABLE 6

TESTS OF JOINT SIGNIFICANCE FOR THE GENDER MODEL (*p*-Values Reported)

Tested Hypotheses	Total Income	Farm Profit	Off-Farm Income
Condition 1a: the effect of the female's minimum schooling is the same as the male's	.45	.64	.99
Condition 1b: the effect of the female's maximum schooling is the same as the male's	.25	.30	.52
Condition 1a and 1b jointly true	.52	.53	.68
Condition 2: All four school terms (female minimum and maximum, male minimum and maximum) are jointly equal to zero	.16	.25	.00

NOTE.—The statistics reported are the *p*-values from four separate joint conditions estimated over total income, farm profit, and off-farm income. The sample contains all households with one adult male and one adult female member. The *p*-values are based on the regression results reported in D. Jolliffe, "Cognitive Skills, Schooling, and Household Income: An Econometric Analysis Using Data for Ghana" (Ph.D. diss., Princeton University, 1996), table 17. Total income is estimated by LAD, while farm and off-farm income are estimated by CLAD.

estimates from using the head of household are very similar to the estimates from using the minimum value of schooling. While the differences across parameters are not statistically significant, they show further support for rejecting the use of the education level of the head of household in favor of using the household maximum or household average level of education.

C. Gender and Household Schooling

One important issue I have ignored so far is that gender may play an important role in whether income can be explained by a specific individual's school attainment. For example, one hypothesis could be that the maximum school level only matters if it is held by a male or a female. I will briefly explore this issue by considering a null hypothesis that the gender of the individual with the minimum or maximum level of schooling has no effect on the determination of household income.

To test this hypothesis, total income, farm profit, and off-farm income are estimated using four variables for school measures: years of schooling of the least educated male and female, and years of schooling of the most educated male and female.[49] Under the null hypothesis, the parameters on the minimum level of schooling will be equal for men and women, as will be the parameter estimates for the maximum level of schooling. In addition, the null hypothesis is credible only if the four school variables are jointly significant.

The results summarized in table 6 are fairly ambiguous. For total income, farm profit, and off-farm income, the data fail to reject the hypothesis that the school effects are the same across genders.[50] This finding supports the null hypothesis that there are not gender effects of this type. The ambiguity results from noting that for the total income and farm functions, the school terms are not jointly significant. One interpretation of the results is that the data are not well suited for exploring in greater detail how the allocation of schooling across gender affects income.

VI. Conclusion

My focus here is on finding an appropriate statistic or variable to measure a household's level of education attainment when trying to model the importance of schooling in the determination of household income. The motivation comes from noting that in developing countries, income is primarily earned (or, more important, measured) at the household level and not the individual level. A significant portion of the literature assumes that the education attainment of the head of household measures the entire household's level of education attainment. On the basis of only the basic descriptive statistics I present, this assumption appears dubious.

This article presents further evidence against using the head of household by testing three competing models of school attainment against each other and against the head-of-household model. The most unambiguous result in this article is the robust rejection of the weak-link and head-of-household models.[51] While the estimation results reject using either the minimum level or the head's level of education to measure household school attainment, they show support for using the maximum level of school attainment when estimating total household income. The data also show support for using the average or median level of schooling when estimating farm and off-farm income.

The proposed interpretation for these empirical results is that education improves the allocation of household resources into farm and off-farm activities, and the allocative effect dominates when considering total income. When considering the farm and off-farm income separately, though, the results suggest that the allocative effect is no longer important; rather, the productivity effect is the primary force through which education improves these components of income.

I also assert that it is important to examine estimates that are robust to violations of normality. Many data sets, particularly data resulting from household surveys in developing countries, are fraught with numerous outliers and other violations of normality. The GLSS data are shown to be no exception to this statement. Similarly, the large majority of nationally representative, household data sets are based on complex sample designs, and this information needs to be incorporated into the estimation strategy. In this article, I note that the differences between OLS and LAD estimates are large and that the effect of correcting estimated standard errors for violations of the independent and identically distributed assumption is also large.

This article attempts to explore a richer model of household school attainment incorporating the possibility that how schooling is distributed across the male and female members may also be an important determinant of income. The GLSS data do not support the hypothesis that it is important to capture the gender differences in school attainment when estimating household income. This rejection, though, is weak and primarily suggests that the hypothesis is not well tested with the GLSS data.

Appendix

TABLE A1

HOUSEHOLD INCOME AND HOUSEHOLD SCHOOLING ESTIMATES OF EQUATION (1) USING ALL HOUSEHOLDS WITH TWO OR MORE ADULTS

Independent Variables	Total Household Income		Farm Profit		Off-Farm Income	
	LAD Estimate	SE	CLAD Estimate	SE	CLAD Estimate	SE
Household minimum level of schooling	.016	(.0340)	−.071	(.0852)	−.042	(.1323)
Household average years of schooling	−.016	(.0608)	.285*	(.1613)	.434*	(.2457)
Household maximum level of schooling	.036	(.0306)	−.084	(.0798)	−.158	(.1340)
Log (acres of land farmed)	.172***	(.0501)	.818***	(.1445)	.015	(.1730)
Household average log of farm experience	−.062	(.0711)	2.720***	(.5371)	−1.246***	(.2937)
Cluster/zone average: day farm wage	−.040	(.0824)	−.025	(.1442)	.083	(.3249)
Region average: ln(price fertilizer)	−1.106**	(.4936)	.084	(1.9166)	.469	(2.2422)
Region average: ln(price insecticide)	−.140	(.3861)	−.488	(1.4769)	.138	(1.8850)
Region average: price of maize	.002	(.0032)	.017*	(.0093)	.006	(.0114)
Region average: price of okra	−.003	(.0211)	.019	(.0565)	−.038	(.0968)
Region average: price of cassava	.013	(.0097)	.004	(.0274)	.039	(.0471)
Region average: price of pepper	.008	(.0066)	.009	(.0157)	.000	(.0312)
Log: business assets	.025**	(.0126)	.056**	(.0282)	.237***	(.0518)

Household maximum: ln(off-farm experience)	.062	(.0581)	−.247**	(.1184)	2.319***	(.3220)
Area average: off-farm wage, type 1	4.189	(5.8086)	−9.827	(15.642)	−1.971	(22.944)
Area average: off-farm wage, type 2	−1.043	(1.6276)	2.869	(4.2740)	2.312	(6.9801)
Area average: off-farm wage, type 3	−1.439	(2.1959)	5.088	(5.5675)	2.044	(7.9940)
Area average: off-farm wage, type 4	−1.007	(.7465)	.950	(2.3684)	.059	(3.1652)
Area average: off-farm wage, type 5	−.083	(.1567)	−.953	(.9674)	−.436	(.6817)
Number of males: 15–24 years old	.064	(.0742)	.184	(.1676)	−.112	(.2957)
Number of males: 25–34 years old	.286*	(.1477)	.342	(.2727)	.718	(.5264)
Number of males: 35–44 years old	.323**	(.1331)	.127	(.3007)	.173	(.5369)
Number of males: 45–55 years old	.161	(.1424)	.239	(.2853)	.433	(.6274)
Number of females: 15–24 years old	.189***	(.0714)	.362**	(.1807)	.067	(.3502)
Number of females: 25–34 years old	.261**	(.1054)	.412	(.2692)	.604	(.4890)
Number of females: 35–44 years old	.174	(.1108)	.456	(.3406)	.078	(.5577)
Number of females: 45–55 years old	.000	(.1132)	−.533**	(.2707)	−.140	(.5533)
Intercept	21.016***	(5.0576)	1.491	(15.250)	−2.547	(23.344)
Pseudo R^2	.105		.333		.224	
Number of observations	1,059		989		1,027	

Note.—Household total, farm, and off-farm income are all estimated in log form. The standard errors are estimated with 500 replications of the two-step, bootstrap procedure described in the text. The sample includes all households with at least two members 15 years of age or older. The total income is estimated by least absolute deviations, while farm and off-farm income are both estimated using J. L. Powell’s censored least absolute deviations estimator (“Least Absolute Deviations Estimation for the Censored Regression Model,” *Journal of Econometrics* 25 [1984]: 303–25).

* Parameter estimates if the p-value is less than .1.

** Parameter estimates if the p-value is less than .05.

*** Parameter estimates if the p-value is less than .01.

Notes

* I would like to thank Chris Paxson for helpful conversations and advice while I wrote this article. I also thank Bo Honoré, Cecilia Rouse, Hanan Jacoby, Paul Glewwe, Andrew Foster, an anonymous referee, seminar participants at the International Food Policy Research Institute, and conference participants at the Econometric Society European meetings for their valuable comments and suggestions. In addition, I thank Bohdan Krushelnytskyy and Anastassia Semykina for excellent research assistance. The views and opinions expressed in this article do not reflect the views of the Economic Research Service of the U.S. Department of Agriculture.

1. Good examples of work on the vastness of the literature on the returns to human capital for wage earners in developing countries include George Psacharopoulos, "Returns to Education: An Updated International Comparison," *Comparative Education* 17 (1981): 321–41, "Returns of Education: A Further International Update and Implications," *Journal of Human Resources* 20 (1985): 583–604, and "Returns to Investment in Education: A Global Update," *World Development* 22 (1994): 1325–43. For a discussion of the relative focus of the human capital literature in developing countries, see Dean Jolliffe, "Cognitive Skills, Schooling, and Household Income: An Econometric Analysis Using Data for Ghana" (Ph.D. diss., Princeton University, 1996).

2. See World Bank, *World Development Report 1995: Workers in an Integrating World* (New York: Oxford University Press, 1995) for a detailed discussion of the typical composition of labor forces in developing countries.

3. Theodore W. Schultz, "The Value of the Ability to Deal with Disequilibria," *Journal of Economic Literature* 13 (1975): 827–46; Finis Welch, "Education in Production," *Journal of Political Economy* 78 (1970): 35–59.

4. Michael Kremer, "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics* 108 (1993): 551–75.

5. Dennis T. Yang, "Education and Off-Farm Work," *Economic Development and Cultural Change* 45, no. 3 (1997): 613–32.

6. Dean T. Jamison and Lawrence J. Lau, *Farmer Education and Farm Efficiency* (Washington, D.C.: World Bank, 1982).

7. George Fane, "Education and the Managerial Efficiency of Farmers," *Review of Economics and Statistics* 57 (1975): 452–61; C. C. Wu, "Education in Farm Production: The Case of Taiwan," *American Journal of Agricultural Economics* 59 (1977): 699–709; Dean T. Jamison and Peter R. Moock, "Farmer Education and Farm Efficiency in Nepal: The Role of Schooling, Extension Services, and Cognitive Skills," *World Development* 12 (1984): 67–86; and Wallace E. Huffman, "Decision Making: The Role of Education," *American Journal of Agricultural Economics* 56 (1974): 85–97.

8. Justin Y. Lin, "Education and Innovation Adoption: Evidence from Hybrid Rice in China," *American Journal of Agricultural Economics* 73 (1991): 713–23; Andrew Foster and Mark Rosenzweig, "Technical Change and Human Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review* 86 (1996): 931–53.

9. James L. Powell, "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics* 25 (1984): 303–25.

10. See Paul Glewwe and Kwaku A. Twum-Baah, "The Distribution of Welfare in Ghana in 1987–1988," Living Standards Measurement Study Working Paper no. 75 (World Bank, Washington, D.C., 1991).

11. Five households are missing school data, and 33 households are missing agricultural price data. For basic descriptive statistics of the variables used in this article, see Jolliffe (n. 1 above).

12. Table 1 actually lists households by the number of household members over

the age of 15. Throughout this article, all education attainment variables are only for individuals over the age of 15. This is to ensure that the majority of continuing students are not included in the analysis.

13. I test the assumption that the school terms enter the income function linearly by estimating income with the squares and interactions of the school terms. In all of the models estimated, the assumption that the squared and interaction terms are jointly zero cannot be rejected.

14. The model is also tested with $S_{\text{med},\ j}$, which is the median instead of the average.

15. For example, those individuals between the ages of 15 and 44 have on average 6.6 years of schooling, while those 45 years of age and older have on average 2.2 years of education. (The *t*-statistic for whether these averages are different is 26.4.) See Jolliffe (n. 1 above) for a breakdown of education attainment by age and sex.

16. Only the estimation results from using the samples of households with two or more adults and households with three or more adults are presented in this article. The results from using all households are qualitatively similar and are presented in full in Jolliffe.

17. Households with more female members have on average lower levels of schooling because females attain less education than males on average. Households with more females are also likely to engage in different types of labor, which will affect farm and off-farm income differently. In particular, females are more likely to engage in work that does not generate income, as measured by the GLSS survey, such as housework. These gender effects are likely to be both correlated with income levels and schooling, and unless controlled for will bias the estimated effect of schooling on income.

18. To correct for a 24% rate of inflation during fieldwork (Ghana Statistical Service, *Statistical News Letter*, no. A11/91, Republic of Ghana, 1991), all values are converted to constant Cedis, with the base month as October 1988. The average exchange rate during 1988 was 200 Cedis to the U.S. dollar (Ghana Statistical Service, *Quarterly Digest of Statistics*, no. 9, Republic of Ghana, 1991).

19. In Ghana, the GLSS data indicate that the value of home-consumed crops constitutes 62% of the total value of farm output. Crops sold on the market constitute 35% of the total farm output, and the remaining 3% comes from the sale and home consumption of animal products.

20. In the type 1 Tobit model, the zeros are typically explained by an optimization problem, which results in a negative value for the desired level of the dependent variable. This estimation problem perhaps more naturally falls into the type 2 Tobit, or selection model framework. In the type 2 framework, the zeros exist because some households choose not to be farmers. I choose the type 1 Tobit framework due to a lack of an empirically credible model defining the selection process into (or out of) farming.

21. Throughout this article the subscript f will denote a farm variable and the subscript o will denote an off-farm variable. These subscripts will be used only on variables that could pertain to either farm or off-farm activities. The subscripts denoting household and individual-level characteristics previously used are dropped here for clarity. Land is treated throughout as a fixed input. In an economy where land markets function well, treating land as fixed would be inappropriate. The more appropriate strategy would be to treat land like any other input: subtract the rental value of land from total output and include the rental price of land in the set of regressors. In the case of Ghana, though, it is difficult to establish a reasonable rental value for the land. Only 17% of the farmers rent any land in or out, and land is rarely sold, both of which mean that land rental prices are not well defined. The fact that land rental markets are not very active suggests, though, that treating land as a fixed input may not be too egregious an assumption. The p_f vector contains cluster average prices for maize, okra,

cassava, and pepper crops, as well as the cluster average input prices for fertilizer and insecticide.

22. Wim Vijverberg, “Measuring Income from Family Enterprises with Household Surveys,” Living Standards Measurement Study Working Paper no. 84 (World Bank, Washington, D.C., 1991).

23. The relative productivity in the farm and off-farm activities is measured by wages in the two sectors. The measure of farm wages used are the wages for an adult male day laborer. The measures of off-farm wages are representative of the wages faced by the sample, yet they are drawn from an independent sample. The supplemental education module was included for only a randomly selected half of the total sample. From the half that did not receive the supplemental module, an hourly wage for all off-farm work was calculated and then grouped by occupation types. From these occupation groupings, which are representative of the off-farm work of the tested sample, mean wages by regions are calculated and then used as estimates of the off-farm wages faced by the sample.

24. Gender composition is also included, as there are cultural norms dictating that men, women, and children will typically perform different work activities; oftentimes the activities of the women and children will not be picked up in the measure of off-farm work. For example, the time spent collecting firewood or food preparation is not included in the measure of total hours worked.

25. I made this decision because it is difficult to find variables that explain why a household engages in farming, but that have no effect on the household’s farming abilities. This is particularly true in a country like Ghana where movement between farm and off-farm activities is fairly fluid, so many households are engaged in both activities.

26. Abbas Arabmazar and Peter Schmidt, “Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity,” *Journal of Econometrics* 17 (1981): 253–58.

27. Wim Vijverberg, “Non-Normality as Distributional Misspecification in Single-Equation Limited Dependent Variable Models,” *Oxford Bulletin of Economics and Statistics* 49 (1987): 417–30.

28. The value of kurtosis is 10.4 for the residuals. The Shapiro-Francia test of normality (see Samuel Shapiro and R. S. Francia, “An Approximate Analysis of Variance Test for Normality,” *Journal of the American Statistical Association* 67 [1972]: 215–16) for the residuals from predicting the log of total income results is a z-statistic of 9.6, which has a p-value less than 0.01.

29. The LAD estimators are less sensitive to outliers than OLS because it is the distance from the median and not the square of the distance (from the average) that determines the parameter estimates. This is analogous to the fact that medians are less sensitive to outliers than are means.

30. For a discussion of the impact of sample design on the correction required for the OLS standard errors, see Andrew J. Scott and Tim Holt, “The Effect of Two-Stage Sampling on Ordinary Least Squares Methods,” *Journal of American Statistical Association* 77 (1982): 848–54. Arabmazar and Schmidt show that when complex sample designs result in heteroscedasticity, standard limited dependent variable estimators (including the Tobit and Heckman’s two-step procedure) are significantly biased.

31. See Trevor S. Breusch and Adrian R. Pagan, “A Simple Test for Heteroscedasticity and Random Coefficient Variation,” *Review of Economic Studies* 47 (1979): 1287–94. The p-value of the test statistic is less than 0.001. The value of the test statistic is 560 and is distributed as a χ^2 with 60 degrees of freedom.

32. Adrian R. Pagan and Frank Vella, “Diagnostic Tests for Models Based on Individual Data: A Survey,” *Journal of Applied Econometrics* 4 (1989): S29–S59.

33. The p-values of both of these test statistics are less than 0.001. The test

statistics are 1,152.5 and 108.6 for the farm and off-farm functions, respectively. Both statistics are distributed as a χ^2 with 60 degrees of freedom.

34. William Rogers, "Calculation of Quantile Regression Standard Errors," *Stata Technical Bulletin* 13 (1993): 18–19.

35. The number of households selected in each bootstrap sample is the number of households in the original sample cluster. Similarly, the number of clusters chosen is equal to the number of clusters in the original sample. For this article, I used 500 bootstrap samples for each bootstrap estimate.

36. For this example I use the maximum level of household schooling. When considering the average level of schooling the household, the standard error of the estimated marginal effect increases by 62%.

37. Bradley Efron and Robert Tibshirani, *An Introduction to the Bootstrap* (New York: Chapman Hall, 1993); Angus Deaton, *The Analysis of Household Survey* (Baltimore: Johns Hopkins University Press, 1997).

38. Roger Koenker and Gilbert Bassett, Jr., "Regression Quantiles," *Econometrica* 46 (1978): 30–50. The LAD estimates presented in this article come from Stata's *qreg* command, which follows Koenker and Bassett in deriving its estimates. Rogers shows that the standard errors reported by Stata are not robust to violations of homoscedasticity or independence. The standard errors presented are derived from the two-step, bootstrap procedure described in the text.

39. The importance of this characteristic is highlighted by comparing OLS and LAD estimates. In the case of estimating the return to total household income, using the maximum level of schooling, the OLS estimates are 94% greater than the LAD estimates, which is consistent with the notion that the OLS is more heavily influenced by outlier values.

40. Moshe Buchinsky, "Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression," *Econometrica* 62 (1994): 405–59.

41. More generally, I drop observations if the predicted value is less than the censoring value when the left tail of the distribution is censored. Similarly, I drop observations if the predicted value is greater than the censoring value when the right tail of the distribution is censored.

42. Moshe Buchinsky, "The Theory and Practice of Quantile Regression" (Ph.D. diss., Harvard University, 1991). Convergence occurs when there are no negative predicted values in two consecutive iterations. All of the models estimated in this article converged and typically converged in fewer than 15 iterations.

43. The models are tested over three samples—all households, households with two or more adults, and households with three or more adults. I present only the summary results from using the samples with two or more adults and three or more adults. The estimation results from the sample of all households are presented in Jolliffe (n. 1 above).

44. Throughout this article a hypothesis is rejected if the *p*-value of the test statistic is less than .10.

45. The full set of estimation results when median schooling is used instead of average schooling are available from the author. They are essentially the same as the full estimation results provided in table A1. (The primary differences are in the schooling estimates, which are listed.)

46. The household head model is tested by reestimating eq. (1), with the head's education level included with the minimum, average, and maximum schooling levels. For the sake of brevity I report only the *p*-values from the relevant *F*-tests; the regression results are omitted. The results are qualitatively similar to the regression results reported in table 6.

47. The point estimates for maximum schooling are almost identical whether the median or average level of schooling is used.

48. Similarly, the estimated return to schooling from using the household median

level to predict farm and off-farm income is 21% and 23% higher, respectively, than from using the head's schooling.

49. The model is tested on the sample of all households with at least one male and female adult member.

50. These tests are listed as condition 1a and 1b separately, as well as the joint test of condition 1a and 1b together.

51. The rejection of both of these models is robust to whether total income, farm profit, or off-farm income are estimated using the full sample of all households, the sample of households with two or more adults, or the sample with three or more adults.