Linkages between Farm and Non-Farm Sectors at the Household Level in Rural Ghana.
A consistent stochastic distance function approach

Gustavo Anríquez and Silvio Daidone

ESA Working Paper No. 08-01

March 2008

Agricultural Development Economics Division

The Food and Agriculture Organization

www.fao.org/es/esa

ESA Working Paper No. 08-01

www.fao.org/es/esa

Linkages between Farm and Non-Farm Sectors at the Household Level in Rural Ghana.

A consistent stochastic distance function approach

March 2008

Gustavo Anríquez*
Agricultural Development
Economics Division
Food and Agriculture Organization,
Italy
e-mail: Gustavo.Anriquez@fao.org

Silvio Daidone
Agricultural Development
Economics Division
Food and Agriculture Organization,
Italy
e-mail: Silvio.Daidone@fao.org

Abstract

In the light of an expanding rural non-farm (RNF) sector in developing rural economies, this paper explores the effects of this expansion within the household. Using rural Ghana as a case study this paper explores if the RNF economy allows for economies of diversification within farms; how input demands, agricultural-specific and shared, are transformed by the expansion of this sector; and if this expansion has measurable effects in overall household production efficiency. We first explore the characteristic of the intra-household linkages (technological and welfare driven) between the agricultural and RNF sectors both assuming perfectly working input and output markets, and assuming market failures, in particular missing labor and credit markets. We then try to measure the identified linkages by estimating a household level input distance function. This function is estimated consistently without making log-transformations as has been previously done in the literature. Our empirical analysis suggests that there are high levels of inefficiency in Ghanaian farms. Also, there are cost-complementarities between the RNF sector and the agricultural sector, particularly with food crops in which the poorest tend to specialize. The expansion of the RNF sector increases demand for most inputs including agricultural land. Finally, we show that smaller farms tend to be more efficient, and that RNF output is helping the farm household to become more efficient, but the latter result is not robust.

Key Words: Rural non-farm sector, input distance function, cost complementarities, technical efficiency; Ghana.

JEL: D13, Q12.

* Corresponding author.

Table of Contents

I.- Introduction

A well accepted characteristic of the process of development of national economies is a relative contraction of the agricultural sector with respect to the rest of the economy as countries become wealthier. This stylized fact probably first formally documented by Kuznets (1957) is by now an accepted feature of the process of economic development. A similar and related process can be observed in the rural economies. As countries grow, agriculture becomes less important in their rural economies as the rural non-farm (RNF) sector of manufactures and services grows faster than farm output. This transformation in the rural world, should also be accepted as a feature of economic development, and has been formally documented by Reardon et al. (1998), and more recently by Davis et al. (2007), among others. The first macroeconomic transformation has important implications for the role of agriculture as an engine of growth and development, and is an issue that has been amply studied in the literature since very early. The second transformation in the rural economy also has sectoral implications that have been studied, less extensively, mainly from a sectoral perspective using Social Accounting Matrices (SAM) and Computed General Equilibrium Models (CGEs) (see for example Haggblade et al. (1989), Vogel (1994), and references contained therein). However, the growth of the RNF economy also has important microeconomic consequences in the economic behavior of rural households that comparatively has been largely ignored in the literature.

In this document we study the linkages between the agricultural and non-farm sector, and take a microeconomic view at the relationship between sectors within the household. Using Ghana as a case study, we try to determine if there are productive linkages within the household, or equivalently a household level multiplier, which would make diversification beneficial for rural households. The existence of this type of linkages would warrant the policy promotion of the RNF sector in case of barriers to entry (like education or access to credit), or other market failures that could hinder its development.

The sectoral relationship between agriculture and the rest of the economy has been chief among concerns of economists since early development economists. Given that agriculture is the most important sector of an economy at early stages of development, economists like Hirschman (1958) explored the input-output linkages between agriculture and the rest of the economy. In what amounts to an historical mistake, according to Anríquez and Stamoulis (2007), Hirschman argued that agriculture's backward linkages, i.e. the capacity of the sector to "pull" the rest of the economy by increasing intermediate input demand when it expands, were very low, therefore it was not a sector worth promoting. This became the common understanding, and agricultural economists who wished to promote the sector started focusing in consumption linkages. Work like that of Haggblade et al. (1989) began showing that the sector's household demand multiplier was very high. This means that when agriculture grows, rural household income grows and the additional household demand caused by agricultural expansion has a very high multiplier effect across the rest of the economy, particularly in closed economies, which is in practice the case of many developing rural economies due to high transaction costs. Additionally, agricultural economists focused in forward linkages, or how agriculture act as an input for downstream activities like the food processing industry or the hospitality services industry[1].

The conclusions of this sectoral and macro view is that productive linkages both forward and backward linkages are more important at early stages of developments. Forward linkages tend to fall relatively less rapidly, as this type of linkages also grow with development. If one adds the household accounts to this multiplier analysis, as done with a Social Accounting Matrix (SAM), one discovers that demand multipliers are high at early stages of development (see Vogel (1994)). Not all of these linkages are accounted by agents, some of them are externalities, for example the expansion of the non-farm sector causing cheaper input supply for farmers. There are of course many nuances in these sectoral linkages, some agricultural activities by their nature have higher forward linkages; for example when they are marketed as processed food, while other have inherently less linkage potential (see for example case studies in Davis et al. (2002)).

The Ghanaian Rural Non-Farm Sector

During the 1990's the Ghanaian economy experienced positive per capita growth which manifested in an important reduction of the national poverty rate by roughly ¼ from 51.7% in 1991/92 to 39.5% in 1998/99 (see Table 1). The picture in the rural economy, the focus of our study, is less clear. There is an inconsistency between on the one side national accounts figures, which indicate very little growth in agricultural value added per capita; and on the other side, the big gains in household expenditure (and consequently reduction in poverty), and the growth of agricultural production per capita reported by FAO. Part of this inconsistency of *maybe* slow agricultural growth and fast rural poverty reduction is explained by a fast expansion of the rural non-farm sector. For example, Table 1 shows that non-farm self-employment income grew in rural Ghana from 11.9% of total household income in

[1] See for example Valdés and Foster (2003).

1991/92 to 23.6% in 1998/99. These big micro-economic changes in rural households in Ghana are likely to have important impacts in farm production as well.

In this study we precisely explore the changes brought about by this transformation of the rural economy where the RNF sector is increasingly more important. The first question regarding specifically the farm sector is whether an expansion of non-farm output is hindering the expansion of the farm economy by competing for scarce household inputs, or instead households are able to benefit from economies of diversification. Furthermore, we explore what type of transformations in the composition of input demand can be expected from this transformation of the Ghanaian rural economy. For example, is RNF output helping households fund input purchases in the absence of working credit markets? Another important question addressed in this paper is whether the expansion of the non-farm economy within households is increasing the technical efficiency of farmers.

The next two sections present the theoretical framework of our analysis: a microeconomic analysis of household level linkages between sectors, and the input distance function used in our empirical study. The fourth section discusses the econometric and empirical issues associated with the estimation of a stochastic distance function. Results of our empirical analysis are presented and discussed in the fifth section, followed by concluding remarks.

II.- Household level linkages between farm and non-farm production.

As it has been established before, when input and output markets are working efficiently, price-taking households behave as a 2-part decision making unit: choosing consumption bundles that maximize welfare given income; while at the same time choosing input and output sets that maximize profits or minimize costs (which are equivalent in this case). However, when market failures are present (cash-constraints, missing markets, information failures, etc.) as agricultural economists like to argue is the case with poor rural households, then production and consumption decisions are taken jointly. In this latter case we expect linkages between farm and non-farm production to be more pronounced; however, even in the case of the "separable" decision-making households, linkages at the production "technology" level may have welfare implications. We try to identify more transparently the underlying relationships.

Households are assumed to maximize a quasi-concave utility function: $U(\mathbf{c}, T - L^S)$, which depends positively on the consumption vector **c**, and the consumption of leisure time, which is the remainder of the total available time T minus worked time (labor supply) L^S. This maximization is bounded by the budget constraint:

$$\sum_j p_j c_j + w(T - L^S) \leq wT + p_0 Q_F + Q_N - wL^D - \sum_i w_i x_i + E \qquad (1).$$

In this constraint the total consumption is valued at market prices p_j, and the time allotment is valued at the opportunity cost of time, that is, the available market wage rate (note that the price of non farm output is used as the numeraire). Total consumption can not be higher than the income generated by the household, which is equal to the value of available time plus exogenous income *E*, plus the rents of producing farm output Q_F and non-farm output Q_N (note that these same outputs are measured as consumption quantities in the **c** vector). These rents are net of the costs of inputs, purchased variable inputs x_i with unit cost w_i, and labor used L^D. Note that the labor supplied L^S and labor employed L^D, need not be equal; if the former is larger labor is offered outside the household, if the inequality is reversed, the household hires external labor.

The welfare maximization is also bounded by technology, which we manifest here with the aid of an implicit production function as:

$$G(Q_F, Q_N; L, \mathbf{x}; K) = 0 \qquad (2),$$

where *K* represents fixed (in the medium term) assets and household characteristics like land, capital, and human capital. Finally, the solution to the household welfare maximization is also bounded by non-negativity of the variables, all inputs, outputs, and consumption goods must be non-negative.

We refer the reader to the still excellent Singh et al. (1986) for details on the separability of this household model when households are price takers in the inputs and goods markets. Further, if we are willing to make the additional assumption that households minimize costs of producing a given amount of output, we can merge the budget constraint and the technical constraint into one:

$$\sum_j p_j c_j + w(T - L) \leq wT + p_0 Q_F + Q_N - C(Q_F, Q_N; \mathbf{w}; K) + E, \qquad (1)'$$

where the function $C(\cdot)$ is the cost function defined as:

$$C(Q_F, Q_N; \mathbf{w}; K) \equiv \min_{L,\mathbf{x}} \left\{ wL + \sum_i w_i x_i \;\middle|\; G(Q_F, Q_N; L, \mathbf{x}; K) = 0 \right\} \qquad (3).$$

Let us isolate consumption and production effects by assuming first that households are price takers in perfectly working inputs and output markets. In this case production decisions are taken separately from consumption choices. We consequently have that production sets are chosen to maximize the income necessary for consumption (even though part of production may/should be consumed):

$$\underset{Q_F, Q_N}{\text{Max}}\ wT + p_0 Q_F + Q_N - C(Q_F, Q_N; \mathbf{w}; K) + E \tag{4}$$

The first-order conditions for this income maximization problem are simply $p_0 = C_{Q_F}(\cdot)$, and $1 = C_{Q_N}(\cdot)$. Differentiating the first condition with respect to the output vector (and not with respect to prices as the household is a price taker) and rearranging we obtain:

$$\frac{dQ_F}{dQ_N} = -\frac{C_{Q_N,Q_F}}{C_{Q_F,Q_F}} \tag{5}$$

This means that farm output can actually increase after an exogenous increase in non-farm output if there are cost complementarities, i.e. $C_{Q_N,Q_F} < 0$ [2].

Cost Complementarities and Economies of Scope

The concepts of cost complementarities and economies of scope are related, but are not the same. Economies of scope refer to the case when it is cheaper to produce goods jointly than to produce them separately, formally:

$$ES \equiv \frac{C(Q_N = 0, Q_F; w, \mathbf{q}; K) + C(Q_N, Q_F = 0; w, \mathbf{q}; K) - C(Q_N, Q_F; w, \mathbf{q}; K)}{C(Q_N, Q_F; w, \mathbf{q}; K)}.$$

There are two separate sources for economies of scope. One is the savings in fixed-costs if the fixed costs of joint production are lower than the separate fixed costs, in our case: $F(Q_F \cup Q_N) < F(Q_F) + F(Q_N)$. The second source for economies of scope is cost complementarities which are caused by the joint usage in production of variable inputs, or other cost saving mechanisms implicit in joint production. Note that these sources (fixed costs economies and cost complementarities) can act in opposite directions and still observe economies of scope; in particular if there are anti-cost complementarities, there could still be economies of scope if these anti-cost complementarities are not large enough to completely eliminate the fixed costs economies:

$$C_{Q_F,Q_N} < \frac{F(Q_F) + F(Q_N) - F(Q_F \cup Q_N)}{Q_F Q_N}\ ^{3}.$$

[2] We are implicitly assuming that C_{Q_F,Q_F} is positive. In a strict technological sense, this derivative could be negative; however, the economic area of this function, that is where rents are positive is defined by increasing marginal costs, i.e. $C_{Q_F,Q_F} \geq 0$.

[3] See Gorman (1985).

If at market prices the household can produce positive amounts of both outputs, in the presence of economies of scope the household would clearly be better off by producing both outputs.

Do economies of scope make sense for poor rural households? This is of course an empirical question that we try to address in this study, however we can hypothesize that they are likely important even in poor rural households:

- Distribution of fixed costs: It may be thought that only high value fixed costs, like expensive machinery and equipment could cause this type of scope economies; however, what is important is not the nominal value, but the value of these fixed costs relative to variable costs. In this sense, the housing infrastructure is usually the largest fixed asset of a poor household and is necessary for farm and non-farm operations.
- Distribution of variable inputs that are useful in both operations. Even in the poorest household there is this type of complementarity as food is an input for labor productivity in both sectors, and labor effort itself can be shared across outputs. For example marketing efforts for one type of goods can be use to market the other type of good.
- Cost complementarities caused by externalities. For example, a non-farm activity could be human capital forming (i.e. book keeping, budget management); skills that could be useful for more efficient management of the agricultural operation. This type of complementarities are externalities because they may not be internalized by the household in their decision-making process.
- Inputs for the other operation are produced below the market price. Likely, this type of cost-complementarities arise when a farm input or by-product is used in the non-farm operation; and this input is produced at a shadow price lower than the market price. These economies are expected, particularly in food processing activities and animal and plant based textiles.

Thus, we expect that even in very poor rural settings to observe important and measurable economies of scope and cost-complementarities.

Non Separable Household Model – Missing Labor Market

In the case of the Ghanaian rural economy, the assumption of a working labor market, with a given wage rate at which labor is freely traded, seems very unlikely. In a cross-country study of 15 developing countries, Davis et al. (2007), showed that the share of wage income (agricultural plus non-farm wages) in household income in rural Ghana is the lowest in their sample (9-11%), and only comparable to Nigeria. These shares indicate quite starkly that access to paid labor markets is very limited in rural Ghana. Before entering into the details of the model, it is straightforward to understand that when only family labor is available, then there is no market wage rate, and the implicit shadow wage rate has to be calculated in the equilibrium between the disutility of the effort and the productivity of the same effort in generating welfare improving income. Thus, utility and technology jointly, not separately determine the effort, the income, and the consumption choices of the household.

With missing labor markets, assuming that non-negativity constraints are not binding, we can define the household equilibrium with two equations:

$$e(\mathbf{p}, w; \bar{U}) = wT + pQ_F + Q_N - C(Q_F, Q_N; w, \mathbf{q}; K) + E \text{ , and} \tag{6}$$

$$T - e_w = C_w \,. \tag{7}$$

Where we are using the expenditure function $e(\cdot)$ to value consumption, which is defined as:

$$e(\mathbf{p}, w; \bar{U}) = \min_{\mathbf{x}, T-L} \left\{ \sum_j p_j c_j + w(T-L) \;\; | \;\; U(\mathbf{c}, T-L) \;\; \geq \bar{U} \right\}$$

Equation (7) defines the intra household labor market equilibrium, it solves for the shadow wage rate, w^*, at which the supply of labor (i.e. the residual of the demand for leisure), is equal to the labor demand schedule from the productive side of the household. As there is no external labor to be hired, or external demand for labor, the implicit shadow wage rate is determined within the household. The amount of labor employed in the household can be obtained by evaluating either side of (7) at the shadow wage rate which solves (6) and (7) jointly.

Totally differentiating (7) and re-arranging terms we get the effect of increasing output on the shadow wage rate:

$$\frac{dw^*}{dQ_F} = \frac{-C_{w,Q_F}}{(e_{w,w} + C_{w,w})} = \frac{-\dfrac{\partial L^D}{\partial Q_F}}{\left(\dfrac{\partial (T - L^S)}{\partial w^*} + \dfrac{\partial L^D}{\partial w^*} \right)} > 0 \,. \tag{8}$$

The denominator of (8) is unambiguously negative because both the expenditure function and the cost functions are concave in prices; a result that only relies in the quasi-concavity of the welfare function and convexity of technology; while the numerator would be also negative if labor is a normal input (a safe assumption). This means that an exogenous increase in farm (or non-farm output), increases the marginal product of labor which is what the shadow price w^* is.

Now we can totally differentiate the production first-order conditions, noting that now w is variable:

$$\frac{dQ_N}{dQ_F} = -\frac{C_{Q_N,Q_F}}{C_{Q_N,Q_N}} - \frac{C_{Q_N,w}}{C_{Q_N,Q_N}} \cdot \frac{dw^*}{dQ_F} = -\frac{C_{Q_N,Q_F}}{C_{Q_N,Q_N}} - \frac{\partial L^D / \partial Q_N}{C_{Q_N,Q_N}} \cdot \frac{dw^*}{dQ_F}. \tag{9}$$

Therefore, in the absence of a working labor market, the production linkage between outputs is lower. The cost complementarity effect is reduced because an exogenous increase of farm output generates an increase in the marginal productivity of labor, an increase in the shadow wage rate, which reduces the demand for labor in the production of the non-farm sector. We also note that the pure labor effect of expanding the output in one sector is negative for the output of the other.

Cash Constraints

In the literature the non-farm economy is many times seen as an escape route when agriculture is failing, but also as an alternative source of cash as working capital when financial markets are not present or working (see for example Reardon et al. (2007)). To analyze the intra-household effect of this type of behavior, we assume a separable household model again, to isolate the cash-constraint effect. In this case the household maximizes income as described in (4), but subject to the constraint:

$$Q_N > \overline{Q}_N, \tag{10}$$

where $\overline{Q}_N$ is the minimum level of farm output to guarantee cash for food and inputs. The first order conditions are simply $p_0 = C_{Q_F}(\cdot)$ and:

$$1 + \mu = C_{Q_N}, \tag{11}$$

where μ is the lagrangian multiplier associated with minimum non-farm output constraint. In this case we can not calculate dQ_N / dQ_F, because the household is at a corner solution, however several conclusions can be derived from (11).

First, the household producing more Q_N than optimal, which means that the non-farm sector is producing at a loss as the marginal costs ($1+\mu$) exceeds marginal revenues at only 1. In the presence of cost complementarities, an exogenous increase in farm output reduces the efficiency losses. From (11) it can be shown that:

$$\frac{\partial \mu}{\partial Q_F} = C_{Q_N,Q_F} < 0.$$

Thus when cost complementarities exist, even for the cash constrained household there is an income effect larger than dQ_F when there is an exogenous expansion of farm output.

Furthermore, the existence of cash constraints has different testable manifestations. As we explain above the non-farm sector is operating at a loss. This arises from the fact that, as opposed to the unconstrained household, all shared variable inputs including labor are producing in the non-farm sector at a lower value of marginal product than in farm production:

$$VMP, L_F = w > \frac{w}{1+\mu} = VMP, L_N \, .$$

Another testable implication of the constrained household is that it is not operating at the optimal point of its production possibility frontier:

$$p_0 = \frac{C_{Q_F}}{C_{Q_N}} > \frac{p_0}{(1+\mu)} \, . \tag{12}$$

Another important reason argued in the literature for rural household diversification into the non-farm sector is a risk management strategy that may help consumption smoothing through low agricultural output years, or through the low season. We do not explore this type of linkages here; however, a theoretical and empirical assessment of this type of linkages is necessary. This analysis requires the consideration of the intertemporal dimension of household welfare maximization, as well as risk.

III.- The Input (Shepherd's) Distance Function

To have an empirical assessment of the technological linkages between farm and non-farm production of rural Ghanaian households we estimate an input distance function. An input distance function, which is defined as:

$$D(\mathbf{Q},\mathbf{x}) \equiv \sup_{\lambda}\left\{\lambda^{-1}\mathbf{x} \;\middle|\, (\mathbf{Q},\mathbf{x}) \in T\right\} = \sup_{\lambda}\left\{\lambda^{-1}\mathbf{x} \;\middle|\, \mathbf{x} \in L(\mathbf{Q})\right\}, \tag{13}$$

is a complete representation of the technology. In (13) T represents the technology, i.e. the technologically feasible set, $L(\mathbf{Q})$ represents the input requirement set, i.e. all input combinations that can produce the output bundle **Q**, and **x** represents an inputs vector. The function describes the largest radial contraction of inputs that leaves the production of a certain output bundle **Q** still technologically feasible. This radial contraction is special in the sense that it contracts all inputs by the same proportional amount. This radial contraction is described for a two-input example in Figure 1, where the input set A is producing the output bundle **Q**, but this input set could be proportionally contracted up to point B and still produce the same output bundle. In this example the value of the distance function is OA/OB. The figure also shows how the input distance function has to be greater or equal to one, and with strict equality when technical efficiency is achieved, i.e. the chosen input set lies exactly within the isoquant. As the input distance function fully represents the technology there is a direct correspondence: $(\mathbf{x},\mathbf{Q}) \in T \Leftrightarrow D(\mathbf{x},\mathbf{Q}) \geq 1$.

We choose to estimate a distance function, because there is a direct relationship between the cost minimization hypothesis, the cost function, and the input distance function; therefore all the cost function properties discussed in the previous section are obtained in a straightforward fashion from the distance function. Furthermore, we prefer the input distance function over the cost function in our case because it does not need reliable price information in order to estimate. What are largely the two most important inputs of agricultural production, land and labor, have extremely underdeveloped markets in rural Ghana. This means that each household has its own shadow labor price which we ignore, and with very few land trades it is very difficult, even for farmers themselves to get an accurate value and price for land. Furthermore, even food crops are not always traded, which means that the market price is not always the relevant shadow price in household production.

Since the inputs and output bundles that are technically feasible are represented by $D(\mathbf{Q},\mathbf{x}) \geq 1$, we can express the cost function as:

$$C(\mathbf{w},\mathbf{Q}) = \min_{\mathbf{x}} \left\{ \mathbf{w}'\mathbf{x} \;\middle|\; D(\mathbf{x},\mathbf{Q}) \geq 1 \right\} \tag{14}$$

Applying the envelope theorem to the maximization problem associated with (14), we get:

$$\frac{\partial C(\mathbf{w},\mathbf{Q})}{\partial Q_i} = -\eta \cdot \frac{\partial D(\mathbf{x},\mathbf{Q})}{\partial Q_i}, \text{ and } \frac{\partial^2 C(\mathbf{w},\mathbf{Q})}{\partial Q_i \partial Q_j} = -\eta \cdot \frac{\partial^2 D(\mathbf{x},\mathbf{Q})}{\partial Q_i \partial Q_j} \tag{15}$$

where η is the multiplier of the optimization problem associated with (14). We can uncover η, using the first-order condition of (14) with respect to inputs we get,

$$w_i - \eta \cdot \partial D(\mathbf{x},\mathbf{Q}) / \partial x_i = 0, \tag{16}$$

if we further multiply by the input and sum over all input first order conditions we get: $\sum_i w_i x_i = \eta \sum_i x_i \partial D(\mathbf{x},\mathbf{Q}) / \partial x_i$. Since the distance function is homogeneous of degree 1 in inputs, which can be observed by inspecting (13), we have on the left hand that the summation over all input derivatives times the input is equal to the original distance function, and as we are evaluating at the optimum the distance function is equal to 1; hence

$$\eta = \mathbf{w}'\mathbf{x} = C(\mathbf{w},\mathbf{Q}).$$

Therefore, the marginal cost is equal to the partial derivative of the input distance function with respect to the same output, but with the opposite sign and multiplied by the value, which is the total cost.

Another important derivative property of the distance function is the returns to scale measure. In general the scale measure indicates the proportion that output changes given a change in inputs. Thus if we have $D(\lambda\mathbf{x},\mu\mathbf{Q})=1$, the returns to scale measure would be: $\varepsilon(\mathbf{x},\mathbf{Q})=\left.\frac{d\ln\mu}{d\ln\lambda}\right|_{\lambda=\mu=1}$. Applying the implicit function rule we have:

$$\varepsilon=\frac{-\sum_i\frac{\partial D(\mathbf{x},\mathbf{Q})}{\partial x_i}\frac{x_i}{D(\mathbf{x},\mathbf{Q})}}{\sum_j\frac{\partial D(\mathbf{x},\mathbf{Q})}{\partial Q_j}\frac{Q_j}{D(\mathbf{x},\mathbf{Q})}}=\frac{-1}{\sum_j\frac{\partial D(\mathbf{x},\mathbf{Q})}{\partial Q_j}\frac{Q_j}{D(\mathbf{x},\mathbf{Q})}}=\frac{-1}{\sum_j\frac{\partial\ln D(\mathbf{x},\mathbf{Q})}{\partial\ln Q_j}} \quad (17),$$

where we apply Euler's theorem in the numerator in moving from the second to the third equality, as we use the linear homogeneity in inputs property of the distance function again.

IV.- Empirical Implementation: A consistent stochastic distance function approach

Probably the first empirical attempt to estimate a stochastic distance function can be found in Grosskopf and Hayes (1993). The authors estimate $1=D(\mathbf{x},\mathbf{Q})-u+v$, where $D(\cdot)$ is approximated by a flexible functional form, in this case a Generalized Leontief (GL), and the residual is composed of a one sided error $u\geq 0$, and v which is a mean zero random noise; therefore $(1+u)$ is λ as defined in (13). The residual u (although with a wrong sign in this paper) is estimated using the third moments of the OLS residual as suggested by Aigner et al. (1977), the seminal paper of the stochastic frontier literature. This approach was not further pursued, first because x's under the cost minimization hypothesis are endogenous (we expand on the endogeneity issue below), and the left hand side variable is a constant.

Following attempts at econometrically estimating a distance function exploited the linear homogeneity of the distance function and started from $x_0\cdot D(\mathbf{x}/x_0,\mathbf{Q})=\lambda$, which is another expression for (13). Applying logarithms, rearranging, and adding the unbiased noise (e^v), we get:

$$-\ln x_0=\ln D(\mathbf{x}/x_0,\mathbf{Q})-\ln\lambda+v \quad (18).$$

In this expression $\ln\lambda$ is u the one sided error term, and $\ln D(\cdot)$ is approximated by the Translog flexible functional form. Expression (18) can be estimated with stochastic frontier methods (see Kumbhakar and Lovell (2000)) which maximize the joint likelihood of the one sided error (assumed to be distributed half normal, truncated normal, exponential or gamma) and a normally distributed random noise *v*. This is the methodology applied by most studies which attempt to estimate a stochastic distance function[4].

In this study this approach can not be followed because the different outputs contain zero as a value, and in particular one output, the non-farm sector will contain zeroes in many rural households. There are alternatives like replacing zeroes with arbitrarily small units, or replacing the logs with arbitrarily large negative numbers. The problem with this type of solutions is that they are arbitrary, and the choice affects the estimated properties of the technology. Thus, we have to use a functional form which does not apply a log transformation, like the Generalized Leontief, the Generalized Quadratic, or the Generalized McFadden. In this case we could estimate:

$$1/x_0 = D(\mathbf{x}/x_0, \mathbf{Q}) - u + v \qquad (19).$$

However, the "true" model as defined by (13) is $\lambda/x_0 = D(\mathbf{x}/x_0, \mathbf{Q})$, therefore the one sided error term in this case is:

$$u = (\lambda - 1)/x_0 \qquad (20).$$

From (20) we see that as expected the one sided error term is positive as the value of the distance function $\lambda \geq 1$. However, (19) violates a key assumption of the stochastic frontier model and the classical linear regression model in general: the one sided error *u* is not uncorrelated with the regressor.

Assessing the Inconsistency

To assess the magnitude and direction of this asymptotic bias let us consider the simplest possible linear distance function with two inputs and one output.

$$(1/x_1) = \beta_0 + \beta_1(x_2/x_1) + \beta_2 Q + e \qquad (21)$$

Here the error *e* is composed of a mean zero i.i.d. error *v* and the input inefficiency *–u*, defined in (20). Assuming for simplicity first that $Cov\left[(x_2/x_1), (\lambda-1)/x_1\right] = 0$, then it is a well established result that the asymptotic bias of the OLS estimation of β_2 is:

$$\text{plim}\ \hat{\beta}_2 = \beta_2 - \frac{Cov\left[Q, (\lambda-1)/x_1\right]}{Var(Q)}. \qquad (22)^5$$

Outputs and inputs are positively correlated, which is guaranteed by positive marginal product of inputs, therefore output and the inverse of an input is negatively correlated. Thus, unless

[4] A good survey of the different nuances of this approach may be found in Coelli et al. (2007).
[5] See for example Wooldridge (2001) pp. 61-65.

there is a very high positive correlation between scale and input inefficiency, we expect that the covariance in (22) to be negative, and hence the bias of the output coefficient to be positive. This means that both OLS and the stochastic frontier would be overestimating both output elasticities and scale economies (recall (17)).

Our earlier assumption that $Cov\left[(x_2 / x_1), (\lambda - 1) / x_1\right] = 0$, is not unreasonable. Input ratios should depend on the relevant input price ratios; so unless there is a high positive or negative correlation between input and allocative efficiency, we should expect that $Cov\left[(x_2 / x_1), \lambda\right] \approx 0$. Also, unless the underlying technology manifests a high degree of non-homotheticity or there a high correlation between scale of production and allocative inefficiency, we should expect $Cov\left[(x_2 / x_1), 1/ x_1\right]$ to be small.

In conclusion we expect the OLS and the stochastic frontier estimation of a linear input distance function to be inconsistent, with output elasticities being the most unreliable.

A possible way out of this inconsistency is to estimate the derivatives of (19), which would not depend on λ. This approach involves estimating forms of (16), or ratios of these first order conditions, as suggested with different implementations by Atkinson and Primont (2002) and Coelli et al. (2007), and is equivalent to estimating input demands or cost shares which are first derivatives of a cost function. In the case of the distance function the derivatives are price shares; however, this approach requires knowledge of the prices of all relevant inputs to deflate prices by total cost. Again, this is an approach not available to us, we have prices of marketed inputs, but not all inputs have working markets, as is pointedly the case of the labor input.

We hence propose to estimate the distance function with the following equation:

$$\hat{\lambda} / x_0 = D(\mathbf{x} / x_0, \mathbf{Q}) + v \qquad (23),$$

which follows directly from (13) after normalizing by an input and applying the linear homogeneity property, and where *v* is obviously the mean zero random noise. We estimate $\hat{\lambda}$, by calculating the Farrell input oriented technical efficiency (Farrell (1957)), using Data Envelopment Analysis (DEA) techniques[6]. The DEA method is a mathematical programming approach to measuring relative technical efficiency. Using Figure 1 again, we can define DEA as the method that uses linear programming to provide an answer to input oriented technical efficiency, in this case the ratio OB/OA, that is $1/\lambda$.

[6] A good manual to DEA methods is Färe et al. (1994).

Assume that we have J, $j = 1,...,J$ decision making units (in our case households), that produce M outputs, $m = 1,...,M$, using N inputs, $n = 1,...,N$, then we can provide piece-wise linear approximation of the input set **x** that can produce **Q**:

$$L(\mathbf{Q}) = \left\{ \begin{array}{ll} \mathbf{x} \;\middle|\; Q_m \le \sum_j z_j Q_{mj} & \forall m = 1,...,M \\ \qquad \sum_j z_j x_{nj} \le x_n & \forall n = 1,...,N \\ \qquad z_j \ge 0 & \forall j = 1,...,J \end{array} \right\} \qquad (24),$$

where we are implicitly assuming constant returns to scale. It is easy to implement alternative hypotheses like non-increasing returns to scale, non-decreasing returns to scale, or variable returns to scale. Given this definition of the input set, the technical efficiency measure of decision making unit *j* is reduced to finding the minimal contraction $\theta \in [0,1]$ that will proportionally reduce all inputs of the decision making unit while still being within the piece-wise representation of the input set. The linear programming representation of this problem is:

$$\begin{aligned} & TE_j(\mathbf{x}_j, \mathbf{Q}_j) = \min_{\theta, \mathbf{z}} \theta \\ & \text{s.t. } Q_m \le \sum_j z_j Q_{mj} \quad \forall m = 1,...,M; \\ & \sum_j z_j x_{nj} \le \theta x_{nj} \quad \forall n = 1,...,N; \\ & z_j \ge 0 \qquad\quad \forall j = 1,...,J. \end{aligned} \qquad (25)$$

which can be solved using the simplex method. The DEA method has evolved healthily since its early simple formulation as we use here, to calculate other economic relations beyond technical efficiency like technical change, allocative efficiency, etc.; but has also suffered criticism. The two main criticism that are exclusive to DEA methods and do not apply to the stochastic frontier approach is that it is not a statistical approach and therefore the battery of hypothesis testing tools can not be applied. The other big criticism is that its results are too sensitive to outliers. This is a contested criticism, because outliers also affect standard regression analysis as well; however here the frontier and the relative efficiency of all the rest of the observations may be affected by one bad observation, while in regression analysis the bias of the outliers is mitigated by the rest of observations.

The method we propose to estimate an input distance function as described in (23) is not ideal. An ideal method would allow for the direct econometric identification of λ. The other drawback of the proposed method is that it is computationally intensive. The benefits of the proposed approach are that it does not rely on an assumption about the distribution of technical inefficiency and the other hand is consistent. The stochastic frontier methods are not the ideal either, because lambda is econometrically identified, but indirectly, and only after making a distributional assumption about it, that may or may not hold. We will return to the drawbacks of distributional assumptions when we benchmark our results, but we can postulate that the higher the inefficiency λ, the higher the effects of assuming the wrong distribution.

Endogeneity of Regressors

If the cost minimization hypothesis holds, then as can be seen in (14) outputs can be taken as exogenous, but inputs are endogenous. This assertion may sound a bit controversial, when one thinks that production units choose both the inputs and outputs. This is true, however, under the cost minimization hypothesis, production units choose (i.e. are endogenous) input (demand) schedules for any positive output bundle. It is in this sense that outputs are exogenous under the cost minimizing hypothesis. As inputs are endogenous, early attempts at estimating a stochastic distance function used instrumental variable (IV) methods. However, Coelli (2000) showed that under the cost minimization hypothesis, the distance function estimated as (18) provides a consistent estimation of the underlying technology (he assumes Cobb-Douglas Technology), even under allocative inefficiency. We do not know if this result can be extrapolated to every possible underlying technology, and functional form of the estimated distance function; however this results relies in the fact that the distance function estimated is a function of input ratios, not inputs, and these are uncorrelated with the technical efficiency residual $\ln\lambda$. This results is further generalized in Coelli et al. (2007), where different types of errors in the observation of *x*'s are present, like technical inefficiency, measurement error and other *multiplicative* errors. In this case the distance function normalized by an input will provide consistent estimates of the technology. This conclusion relies in the definition of the distance function, if we define the technically efficient level $x_1^t \equiv x_1 / \lambda$, as defined by (13), then the ratios of observed inputs, $x_1 / x_2 = x_1^t / x_2^t$ is also technically efficient by definition. This is the reason why input ratios as regressors allow for consistent estimates of the technology, and not because production units choose inputs but not ratios as some have argued.

There is a cost, however, to choosing the input with which to normalize the distance function. Although in theory it does not matter which input is used to normalize the function, in practice it matters and results vary. This is a topic that the econometrician usually chooses to ignore, but has been discussed in the context of cost function estimation where it has been recognized that the chosen normalizing input (when linear homogeneity in prices of the cost function is imposed) significantly affects the estimated technology[7]. This is why Kumbhakar and Lovell (2000) suggest normalizing the distance function by $\|\mathbf{x}\|$, i.e. the Euclidian norm of the input vector. Although this is a reasonable choice that eliminates the arbitrariness of the normalizing input choice, it is not clear that this normalization allows for a consistent estimation of the parameters of the underlying technology. Thus, in order to eliminate this arbitrariness, while at the same time exploiting the full variability of our data set, we propose to estimate the distance function in a system of equations:

$$\begin{aligned} \hat{\lambda}/x_0 &= D(\mathbf{x}/x_0, \mathbf{Q}) + v_0 \\ &\vdots \\ \hat{\lambda}/x_n &= D(\mathbf{x}/x_n, \mathbf{Q}) + v_n \end{aligned} \tag{26}$$

in which we obviously impose the restrictions that all parameters of the linear approximation of the distance function are equal across equations, and we exploit the cross equation error correlation in a maximum likelihood System of Unrelated Regressions (SUR).

In this study we approximate the distance function (with four outputs and four inputs as described below) with the following flexible functional form, which is a form of Generalized Leontief:

$$\begin{aligned} D(\mathbf{x},\mathbf{Q}) \equiv {} & \sum_{ij} a_{ij}(x_i x_j)^{1/2} + (Q_1Q_2)^{1/2}\sum_i b_i x_i + (Q_1Q_3)^{1/2}\sum_i c_i x_i + (Q_1Q_4)^{1/2}\sum_i d_i x_i + \\ & (Q_2Q_3)^{1/2}\sum_i e_i x_i + (Q_2Q_4)^{1/2}\sum_i f_i x_i + (Q_3Q_4)^{1/2}\sum_i g_i x_i + Q_1\sum_{ij} m_{ij}(x_i x_j)^{1/2} + \\ & Q_2\sum_{ij} n_{ij}(x_i x_j)^{1/2} + Q_3\sum_{ij} p_{ij}(x_i x_j)^{1/2} + Q_4\sum_{ij} q_{ij}(x_i x_j)^{1/2} + Q_1^{1/2}\sum_i r_i x_i + Q_2^{1/2}\sum_i s_i x_i + \\ & Q_3^{1/2}\sum_i t_i x_i + Q_4^{1/2}\sum_i z_i x_i \end{aligned} \qquad (27).$$

The specification described in (27) only imposes linear homogeneity in inputs, which is property imposed by theory, i.e. the definition of a distance function. All the rest of the properties of the technology, are flexible, in the sense that even second derivatives are not constant and depend on the data. We highlight coefficients **b**, **c**, **d**, **e**, **f**, **g**, which are used to estimate output jointness, they allow for these cross-output effects to be scale dependent, as the cross derivatives and elasticities will depend on both output and input level. This scale dependence is a desirable property, because one would expect that cost complementarities if they exist would probably be more important at lower scales of production.

[7] See Maietta (2002) and Kumbhakar and Karagiannis (2004) for example.

V.- Data and Results

We use household level data coming from the Ghana Living Standard Survey Round 4 (GLSS4), a nationally representative multi-purpose household survey. In order to ensure national representativity, the survey uses the 1984 Demographic Census enumeration areas (EA) as primary sampling units, which in total sum to 300 EAs. A fixed number of 20 households were selected as secondary sampling units[8]. As it has been pointed out by the authors of the survey, this sampling frame, though quite old and inadequate, is the only available in the country (Ghana Statistical Service (1999)).

Of the 6,000 observations, 3,799 households correspond to rural areas. In this study we focus on farm households (with positive owned or operational landholdings) which reduces the sample to 3,165 observations. However, not all these observations could be used. A large amount of households did not report the level of key inputs and/or outputs or other important control variables. Further, we had to deal with outliers, i.e. observations for which there was likely a problem of misreporting (see details of our treatment of outliers in the Appendix). Our final sample then consists of a cross-section of 2,138 rural households.

Farms in our dataset undertake several activities, producing both farm and non-farm income. We computed three farm output measures, cash-crops, food crops, and livestock and other crops. These three farm outputs are measured as indexes: total value of the household output divided by the cross-section median[9]. Livestock output has been computed as the sum of in-cash and in-kind incomes from livestock produce (eggs, milk, dairy products, etc.), plus sales and rents of livestock and the value of own consumption of livestock and its produce. Off-farm output is measured as the sum of all non-farm revenues. These include all household non-farm enterprises revenues, incomes produced by selling water and renting/sharecropping land, and wages from employment (including agricultural employment).

Table 2 provides a brief overview of the structure of farmers' production in Ghana. Food crops prevail in most regions, except in those characterized by urban agglomerates, where not surprisingly off-farm incomes make up approximately 40% of the total value of production. Livestock does not seem to be of great importance in the country, since on average it accounts for 6.5% of the value of production. The components of non-farm income

[8] Stratification was done according to ecological zones and then further dividing locality into urban/rural. An EA is considered to be urban if it had a population greater than 1,500 people during the 1984 population census.

[9] We constructed pure quantity indexes as well. However, in Ghana many traditional and non-standard units (in the sense that they vary by region, like “box”) are reported in the survey, not all of them with known conversion factors to standard volume or weight units. Thus the construction of these indexes required many assumptions, and estimations, which is why we feel much comfortable about using value as a proxy of quantity. Further, a sensitivity analysis below explores the effects of this choice.

are detailed in Table 3. The non-farm enterprises category considers any business or trade not related to agriculture operated by household members, including self-employed professionals or craftsmen. This category largely accounts for most of non-farm income, independently of the region of residence. A high contribution is given also by wages from employment, particularly in the southern regions where it reaches 20% of non-farm income. Revenues generated from the sale of water and from renting out and sharecropping-out land appear to be very marginal. In some regions though, it seems that sharecropping has a tangible impact on off-farm incomes. Nonetheless both (water sales and land-leasing income) show very high variability coefficients, indicating, that although overall the importance of these income sources is low, for a limited number of household these are important sources of income. Finally, remittances are the main source of non-farm income in the Northern and poorest regions of Ghana.

With respect to inputs, we have constructed four indexes by dividing the measures of land, labor, livestock and operating expenses by their respective sample median. Land is given by the number of acres operated by the household members and may include any plot which has been owned, rented in and sharecropped in, all of which amount to land that is used as an agricultural input. We also add to this measure the land that has been rented-out or sharecropped out which is the land that is used as an input of non-farm income.

Labor employed is proxied by the family members 12 years or older. Obviously we would like to use effective labor employed, i.e. hours used in the farm and non-farm activities, but this information was only recorded for the household non-farm enterprise. In absence of effective labor we use family labor supply as a proxy. Livestock units (as an input stock) are evaluated at the sale price (see details in the Data Appendix). Lastly, we account operating expenses, by adding all purchased inputs, which include expenditures on inputs such as energy, fertilizers, seeds and the like.

In Table 4 we describe input usage by Ghanaian farms. We can see that there is not much variation in the labor input, which mean is concentrated around 2. Further it can be pointed out that plots are on average very small in almost every region and that their variability is also quite low. This means that in our study we are considering a sample of relatively homogeneous small farms[10]. Much more variability is present instead in the two other input measures, although some of this variability is price driven. We also observe, as

[10] We remind the reader that 1 acre is approximately 0.4 hectares of land. This means that the average land size is roughly 3.5 hectares.

expected, a positive correlation between regions with high livestock input usage and higher livestock output.

In addition to inputs and outputs, the distance function is estimated with several control variables. We used regional dummies to control for unobserved region-level differences and dummy variables related to land, to check whether owning and/or renting out plots affect the efficiency of the productive process. Further we inserted household characteristics, such as a dummy whether a female is the head of the household, the age of the head of the household in linear and squared terms, a housing index to control for the quality of the house, the highest level of education within the household, the minimum distance to a school and a dummy variable whether the household has a formal loan.

The Results

Table 5 presents the estimated coefficients by maximum likelihood of the system represented by (26) and (27). The system has a very good fit as reflected by the fact that most technology parameters are highly significant. We first note that most controls, with the exception of age have the expected sign. In this case female headship and distance to markets (proxied by distance to schools) are associated with higher technical inefficiency, while education, housing quality, and formal loans are all associated with higher technical efficiency. However, only education, female headship, and land ownership are statistically significant. Age is surprisingly correlated (although not significantly) with more technical inefficiency, but at a decreasing rate. We also acknowledge there could be reverse causality, i.e. households have members achieving higher education because they are more efficient.

Input technical efficiency as estimated in a first stage by the DEA method yielded rather surprising results. We expected, due to low general levels of education, missing and imperfect markets to find high inefficiency, but the average level of 0.18, as shown in Table 6, was rather surprising. These estimates signify that on average input sets could be proportionally contracted to 18% of their original levels and still produce the same amount of output. At the same time the high variance of technical efficiency shows that households were distributed all over the feasible range (0,1] as shown in Figure 2. Furthermore, efficiency is not correlated with farm size, as could be conjectured, as a matter of fact smaller farms are more efficient than larger farms as we explain in more detail below.

The output elasticities, presented in the first column of Table 6 indicate that as expected all marginal costs are positive. Also, the figures reveal increasing returns to scale, however, as shown, a test of constant returns to scale can not be rejected. A separate DEA analysis also indicated increasing returns to scale, but the hypothesis can not be statistically tested. These results may seem surprising, because in the context of working input and output markets increasing returns to scale is consistent with some sort of externality, like agglomeration economies, knowledge externalities, and the like; all of which are unlikely to be present in a developing country agricultural sector. However, when there are market failures, it is possible that units can not adjust to levels consistent with constant or decreasing returns to scale, which is likely the case of rural Ghanaian households. In Table 8 where we calculate elasticities by farm size, we see that the increasing returns to scale result is strongly driven by farms smaller than five acres, while farms larger than 10 acres actually show decreasing returns to scale.

Given the high technical inefficiency it is hard to interpret adequately the input elasticities. If households were technically efficient, then, as shown by (16), the input elasticity would be exactly equal to the input cost-share. In this case what we recover from the input elasticities is the cost share under shadow prices, as shown by price ratio, w_2^s / w_1^s in Figure 1. The difference between, this technically efficient set, point B in the figure, and point C, the cost minimizing set, is called allocative inefficiency. It is hard to talk about allocative inefficiency when input markets are clearly not fully working, which is why we do not try to estimate it. Under the shadow cost prices, operating expenses and labor account for an equal share of total output costs around 37%, while land amounts to 23%, and livestock account for less than 1% of total costs. Livestock however is an important input for smaller farms, up to 5 acres, accounting for 6-10% of shadow costs of small farms of less than 2.5 acres (as shown in Table 8).

The cross output elasticities presented in Table 7, indicate that cost complementarities are present among all outputs. This indicates the opportunity for important economies to diversification for rural household. The most important cost complementarities are among, not surprisingly, food crops and livestock, the activities in which the diversified rural households are mostly employed (i.e. specialization in cash crops is more likely). Next in importance are the cost complementarities between food and cash crops. Non-farm complementarities are the third in importance, particularly with food crops, and livestock production.

Another important consequence of participating in non-farm activities to consider is its effect on input demand. It is frequently argued, see Katz and Stark (1986) and Haggblade et al. (2007) for example, that income from non-farm activities, including remittances from migration, serves to alleviate credit constraints. If this was the case, we would expect to find that expansions in non-farm output would cause increases in farm input use. In Table 8 we show the implicit input partial demand elasticities obtained from the distance function[11]. Both purchased inputs and workers, which are shared inputs, expand with non-farm production; but more importantly, land, which is an exclusively agricultural input, also expands with non-farm production. This observation is *consistent* with the hypothesis of cash constraints. Similarly, we also find important differences in the ratio of marginal costs between farm sizes. If we take the ratio of marginal costs of food crops with respect to non-farm activities, or cash crops with respect to non-farm output, we find that in both cases this ratio is higher for the overall sample than for smaller farms under 5 acres, implying that smaller farms get involved in non-farm activities at lower levels of relative productivity. This result as (12) suggests is *consistent* with the hypothesis that small farm households face cash-constraints. Both observations, positive partial input demand elasticity with respect to non-farm output, and the lower relative productivity of non-farm activities in smaller farms are not a *proof* of cash-constraints, but observations that are empirically coherent with this hypothesis.

Finally, we explore the effects of non-farm production in overall input efficiency. We estimate a closed-form regression trying to explain the determinants of input-oriented technical efficiency. We use control variables, and output measures of farm and non-farm production, on which we focus. In the first column of Table 9, we show as a benchmark the results of estimating the closed form model with standard OLS procedure. These results are not reliable, because there is a serious problem of endogeneity in the sense that more input-efficient farms, *ceteris paribus*, produce more and vice versa. Thus, the statistically positive partial correlation reported in the first column for both output measures is actually expected. When we control for the endogeneity of output, estimating in a first stage a simple Cobb-Douglas production function to use predicted output levels (column 2) we find that farm output is actually negatively and significantly correlated with efficiency (note that this is consistent with Table 8, which shows that on average small farmers are more efficient than larger farmers). Non-farm output, on the other hand is marginally positively correlated with more efficiency, however this is not a robust result as it depend on the estimating procedure, as column 3 shows, where non-farm output is not significant when estimated with a 3-stage

[11] The partial input demand elasticities can be obtained by totally differentiating first order condition (16), and can be shown is equal to: $\partial \ln x_i / \partial \ln Q_4 = Q_4 / x_i \cdot \left(\partial D / \partial Q_4 \cdot \partial D / \partial x_i - \partial^2 D / \partial x_i \partial Q_4 \right) / \partial^2 D / \partial x_i^2$. This expression is highly non-linear in the regression coefficients estimated, which prompted us into using bootstrapping techniques for hypothesis testing throughout this paper.

procedure[12]. Chavas et al. (2005), found comparable results in The Gambia, non-farm earnings in their 1993 sample did not significantly affect household technical efficiency, in their OLS model.

Sensitivity Analysis

An important issue to explore is if the land elasticity is mis-calculated due to differences in land productivity. We try to control for this unobserved characteristic by estimating the system with enumeration area dummies, which will capture cluster level differences, among them, variations in land productivity. In unreported regressions we find that when cluster dummies are used surprisingly the land elasticity does not change significantly, but it is the livestock output elasticity the only one which is significantly reduced.

Another, issue that requires further examination is our use of production values instead of value-free physical measures of output. If prices where constant throughout Ghana, this would be an innocuous choice; however, prices are likely to vary a lot, particularly by distance to markets. We expect that for isolated farms the per unit value of agricultural output to be much lower than for farms close to markets. We explore the consequences of this price differences assuming that for farm *i* the gate price of output, p_i^g, is equal to the market price, p^m, times a deflating function that depends on distance to market $g(d)$, i.e. $p_i^g = p^m g(d_i)$, with $g(d_i) < 1$ and $g'(d_i) < 0$. For simplicity we assumed $g(d_i) \equiv 1/(1+\alpha d_i)$, and explored the sensitivity of the estimated results to different levels of α. Distance to markets was proxied by distance to schools, and we deflated crop and livestock values (input stocks and output), and inflated the cost of purchased inputs. In unreported regressions we found that most elasticities reported in Table 6, are surprisingly robust, except for the output elasticity of food-crops (Q_2), which is very likely underestimated. This result is sensible, as precisely those more isolated farms will specialize in food crops, not is cash crops, because of distance to markets. If the gate value of food crops of these distant farms is lower, we are underestimating their output level, and consequently overestimating the marginal cost of their production. Consequently, this overestimation of the marginal cost of food-crop production together with the robustness of the other elasticities means that we are likely underestimating scale economies, and as we increased the α, the scale-economies became larger, and statistically larger than 1.

[12] We also explored the Tobit model, given that efficiency is censored at 1, but these models are not reported as they are not significantly different from the 2-stage estimation.

Benchmarking Results with Stochastic Frontier Estimation

The second column of Table 6 shows the elasticities of distance function (27) estimated with a half-normal stochastic frontier model, as described in (19). The results indicate that the half-normal stochastic frontier model clearly fails to estimate the underlying technology. The implicit negative marginal costs and implausibly high implicit returns to scale estimated are violations of basic economic behavior. We believe that the linear stochastic model fails for three reasons. First, as shown above, the linear stochastic distance function estimation is inconsistent, and this inconsistency is proportional to the level of inefficiency, which in our case is high; therefore estimates are highly inconsistent. In particular, as predicted in section IV, all output elasticities are overestimated (compared to our consistent SUR estimates), leading to negative marginal costs and unfeasibly high scale economies.

Second, *a posteriori* we can see that the half normal distribution of the input distance is an inadequate assumption. The DEA technical inefficiency measure could be questioned regarding levels given the treatment of outliers. However, the underlying distribution of efficiency calculated and shown in Figure 2 is harder to question. This distribution can not be fit by a half-normal or exponential distribution, as shown by a simulated half-normal distribution in the same figure. The third reason for the failure of the stochastic frontier model is that the input distance is too high. Stochastic frontier methods are an elegant way to ask the residual for the level of technical inefficiency. If the level of technical inefficiency is too high, the method may be asking too much to the residual. The econometric lessons learned in this study call for care when using stochastic frontier methods in the context of microeconomic development analysis.

VI.- Conclusions

Perhaps the most important finding of this study is not related to non-farm activities, its original focus, but the surprisingly low overall efficiency of Ghanaian farms. This has important implications for the discussion of agricultural technology for Sub-Saharan Africa. The numbers presented in this paper suggest that the challenge is not to develop new technologies for Africa, but rather to enable the adoption of existing technologies. Another important finding regarding the overall Ghanaian rural economy is the likely presence of increasing returns to scale in household production. On the one hand it means that there are important gains to be attained in the economy by increasing scale of production, but it is also an indicator that markets are not working, and that there are obstacles that hinder households, farms and small business from achieving their optimal scale of production.

With regards to the focus of our study, the non-farm activities, we found that overall the sector allows for significant economies of diversification for rural households. However, we do not know how large these linkages are relative to cost complementarities present in the households of other developing rural economies. Furthermore, the other important question is to compare these micro level linkages with other macro/sectoral level linkages. It would be important from the policy perspective to know which types of linkages are more important within an economy, to properly target development policy. We also found marginally significant effects of non-farm production on overall household efficiency; but at the same time, we are sure that non-farm production is not negatively correlated with efficiency as is the case with farm production. Also, this study presented diverse evidence consistent with the hypothesis of the non-farm sector easing household cash-constraints. Although this hypothesis requires further examination, it provides yet another argument to provide an enabling environment for the development of the non-farm sector.

Finally, the different estimation techniques explored in this investigation call for the attention of the practitioner when using stochastic frontier models in the presence of high levels of technical inefficiency. When inefficiency is high, the effects of making wrong assumptions about the distribution of technical efficiency may have, as we showed, serious effects on the estimated parameters.

Figures

Figure 1. Production and efficiency

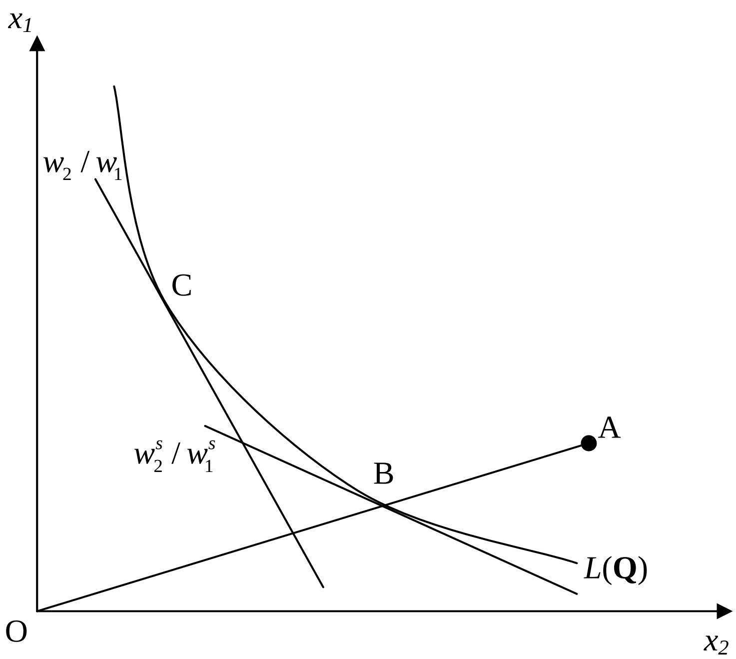
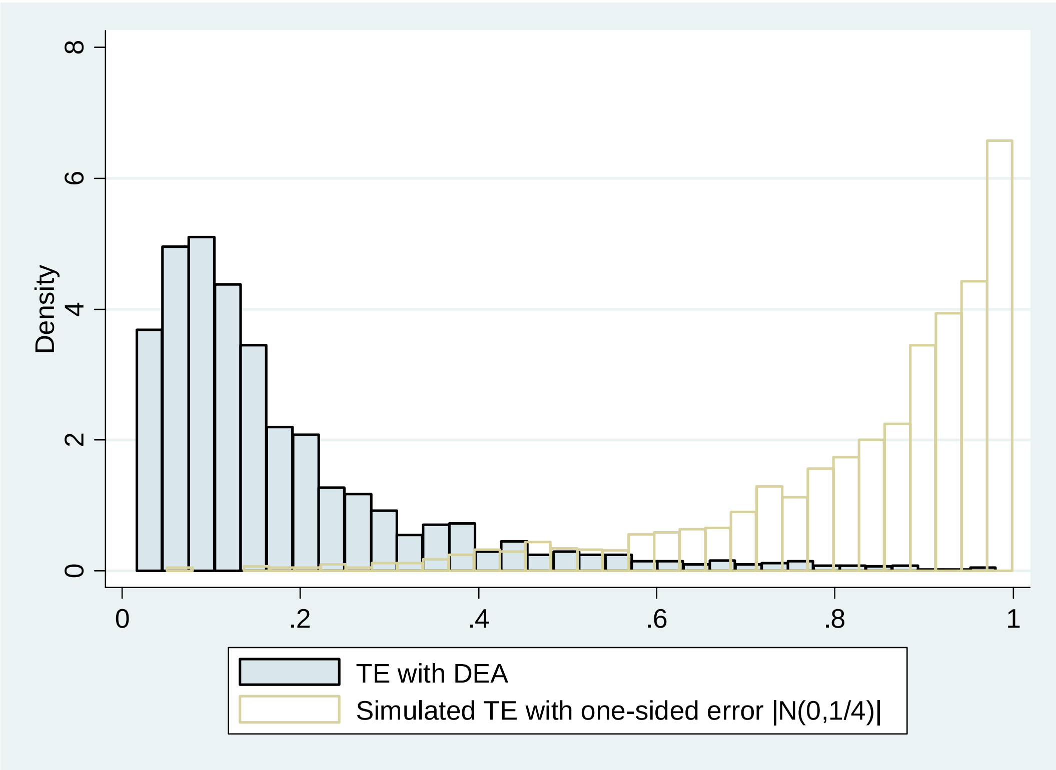


Figure 2. Calculated and Simulated Technical Efficiency



Tables

Table 1. Key Economic and Social Indicators from Ghana

	1987/88		1991/92		1998/99
Per capita GDP[1]	202.56		216.91		244.17
Mean yearly growth rate		1.73		1.71	
Agriculture, value added (% GDP)	50.5		45.5		36
Per capita agricultural GDP[1]	84.49		82.58		87.59
Mean yearly growth rate		-0.57		0.85	
Per capita agricultural production[2]	70.05		84.20		93.70
Mean yearly growth rate		4.71		1.54	
Population, total	14,439,140		16,145,312		19,221,380
Rural population[3]	65.5		62.5		57.5
Households income shares[4] (%)					
Farm	66.4[6]		60.88		42.08
Non-farm self-employment	16.1[6]		15.49		28.67
Wage employment (including agr.)	17.5[6]		23.62		29.25
Rural households income shares[4] (%)					
Farm			77.00		58.31
Non-farm self-employment			11.93		23.64
Wage employment (including agr.)			11.07		18.05
Per capita expenditure[5] – National			798,594		993,897
Mean yearly growth rate				3.17	
Per capita expenditure[5] – Rural			658,882		773,093
Mean yearly growth rate				2.31	
Poverty incidence – National (%)			51.7		39.5
Poverty incidence – Rural (%)			63.6		49.5

Notes: 1) Constant 2000 US $. 2) Production index. 3) % of total population. 4) Calculated as shares of aggregate household income, excluding transfers and miscellaneous sources of income. 5) In 1999 local currency (cedi). 6) Income shares from Newman et al. (2000), not exactly comparable to 1991/92 and 1998/99 income shares.

Sources: World Development Indicators from the World Bank.91/92 and 98/99 income shares, poverty indexes, per capita expenditures from GLSS3 & GLSS4 and Ghana Statistical Service (2000). Agricultural production indexes from FAOSTAT.

Table 2. Output composition: mean values and shares (coefficients of variation in parentheses)

Region (Observations)	**Cash crop (Q_1)** Value	Share	**Food crop (Q_2)** Value	Share	**Livestock (Q_3)** Value	Share	**Off-farm (Q_4)** Value	Share
Western	1,090,539	28.0	1,512,527	41.5	41,923	1.9	1,756,757	28.6
(222)	(1.3)	(1.0)	(1.5)	(0.7)	(2.6)	(3.7)	(2.5)	(1.2)
Central	390,636	14.9	965,923	48.8	68,408	5.7	1,360,021	30.6
(274)	(1.8)	(1.4)	(1.0)	(0.6)	(1.9)	(2.3)	(2.9)	(1.1)
Greater Accra	48,667	0.9	1,031,508	48.1	11,677	0.6	2,367,639	50.4
(24)	(4.0)	(2.9)	(1.5)	(0.8)	(3.3)	(3.1)	(1.9)	(0.8)
Eastern	126,339	6.1	720,464	42.3	79,467	7.4	2,061,004	44.3
(244)	(2.8)	(2.2)	(1.4)	(0.8)	(2.5)	(2.3)	(2.5)	(0.9)
Volta	252,790	10.2	1,292,040	52.3	86,643	5.4	1,490,641	32.2
(323)	(2.2)	(1.7)	(1.3)	(0.6)	(2.1)	(2.2)	(2.0)	(1.1)
Ashanti	383,897	9.7	1,622,635	60.3	47,503	2.8	1,531,279	27.2
(374)	(2.5)	(1.8)	(1.2)	(0.5)	(4.0)	(2.9)	(2.8)	(1.2)
Brong Ahafo	406,276	9.1	2,019,749	69.3	35,411	1.6	766,439	20.0
(243)	(3.1)	(1.9)	(1.1)	(0.4)	(2.8)	(2.6)	(2.5)	(1.3)
Northern	243,255	18.5	724,586	54.0	148,351	10.6	443,130	16.9
(182)	(1.3)	(1.0)	(1.0)	(0.5)	(1.5)	(1.2)	(2.5)	(1.5)
Upper East	122,174	17.4	389,624	60.6	86,974	10.8	975,544	11.2
(53)	(1.3)	(1.0)	(0.8)	(0.4)	(1.6)	(1.1)	(6.4)	(2.2)
Upper West	165,066	15.0	770,498	51.8	136,959	11.9	669,321	21.3
(199)	(1.0)	(0.9)	(1.9)	(0.5)	(1.3)	(1.1)	(2.9)	(1.4)
Total	368,886	13.2	1,226,305	53.2	75,277	5.6	1,322,882	28.1
(2138)	(2.4)	(1.5)	(1.4)	(0.6)	(2.3)	(2.1)	(2.8)	(1.2)

Note: Values in local currency (Ghanaian cedi). Shares in percentages.

Table 3. Non-farm activities composition: mean shares (coefficients of variation in parentheses)

Region (Observations)	Non-farm income	Non-farm enterprises	Wages	Water sold	Land Rental	Remittances	Other
Western	1,756,757	44.15	19.82	0.00	2.98	29.22	3.82
(222)	(2.53)	(1.09)	(1.92)	(.)	(5.48)	(1.52)	(4.59)
Central	1,360,021	42.79	8.42	0.49	3.99	37.54	6.77
(274)	(2.94)	(1.09)	(3.02)	(14.25)	(4.53)	(1.20)	(3.26)
Greater Accra	2,367,639	45.83	15.13	0.00	9.40	24.95	4.69
(24)	(1.88)	(1.01)	(2.33)	(.)	(2.35)	(1.52)	(2.91)
Eastern	2,061,004	58.37	8.92	0.00	0.34	26.17	6.20
(244)	(2.54)	(0.79)	(2.91)	(.)	(9.52)	(1.57)	(3.56)
Volta	1,490,641	39.24	13.29	0.22	5.88	36.75	4.62
(323)	(1.98)	(1.17)	(2.40)	(12.49)	(3.68)	(1.24)	(4.06)
Ashanti	1,531,279	26.80	8.83	0.02	5.56	40.22	18.56
(374)	(2.85)	(1.57)	(2.98)	(16.70)	(3.60)	(1.12)	(1.97)
Brong Ahafo	766,439	26.74	14.37	0.00	4.80	52.06	2.04
(243)	(2.50)	(1.57)	(2.34)	(.)	(3.79)	(0.90)	(6.37)
Northern	443,130	33.68	7.51	0.00	0.00	45.81	13.00
(182)	(2.53)	(1.36)	(3.27)	(.)	(.)	(1.06)	(2.35)
Upper East	975,544	14.29	7.14	0.00	0.00	47.05	31.52
(53)	(6.40)	(2.54)	(3.74)	(.)	(.)	(1.06)	(1.47)
Upper West	669,321	29.39	6.92	0.00	0.00	47.35	16.34
(199)	(2.93)	(1.48)	(3.49)	(.)	(.)	(0.98)	(2.02)
Total	1,322,882	37.53	11.03	0.11	3.53	38.53	9.28
(2138)	(2.84)	(1.22)	(2.66)	(26.36)	(4.67)	(1.19)	(2.86)

Notes: Off-farm incomes in local currency (Ghanaian cedi). Other columns represent off-farm income shares (percentages).

Table 4. Average input values (coefficient of variation in parentheses)

Region (Observations)	**Land size[1] (x_1)**	**Purchased inputs[2](x_2)**	**Workers[3] (x_3)**	**Livestock[2] (x_4)**
Western	10.97	326,445	2.01	439,018
(222)	(0.97)	(1.93)	(0.42)	(1.54)
Central	10.49	158,302	1.78	353,189
(274)	(1.86)	(2.05)	(0.41)	(1.40)
Greater Accra	6.65	130,833	2.00	1,166,667
(24)	(2.29)	(1.24)	(0.42)	(1.25)
Eastern	4.56	247,635	2.48	972,325
(244)	(1.85)	(2.09)	(0.55)	(3.52)
Volta	7.85	246,940	2.23	357,061
(323)	(1.82)	(2.07)	(0.48)	(1.15)
Ashanti	9.29	263,532	2.08	600,321
(374)	(1.09)	(3.39)	(0.48)	(4.68)
Brong Ahafo	13.86	189,045	1.77	555,428
(243)	(1.78)	(1.62)	(0.47)	(1.75)
Northern	7.63	171,897	2.52	1,667,577
(182)	(0.78)	(1.08)	(0.44)	(1.17)
Upper East	6.55	71,192	2.81	1,146,244
(53)	(0.52)	(3.93)	(0.47)	(1.07)
Upper West	5.11	92,140	2.39	1,109,244
(199)	(0.69)	(2.38)	(0.44)	(1.32)
Total	8.75	213,781	2.15	762,746
(2,138)	(1.62)	(2.51)	(0.49)	(2.50)

Notes: 1) in acres; 2) in local currency (Ghanaian cedi); 3) household members12 or older.

Table 5. Maximum Likelihood SUR technology parameters estimates (2138 observations, standard errors in parentheses)

Western	1.819	a24	-3.017***	g1	0.300*	p11	-0.526***	s1	-3.210***
	(2.120)		(0.446)		(0.171)		(0.192)		(0.361)
Central	2.838	a33	22.234***	g2	-0.212***	p12	-0.214	s2	0.434
	(2.015)		(0.739)		(0.061)		(0.240)		(0.284)
Greater Accra	9.325**	a34	9.930***	g3	4.605***	p13	0.487	s3	-24.186***
	(4.582)		(0.757)		(0.351)		(0.331)		(1.008)
Eastern	11.212***	a44	-3.980***	g4	-0.403***	p14	0.206*	s4	1.618***
	(2.067)		(0.487)		(0.109)		(0.125)		(0.316)
Volta	3.558*	b1	-0.143*	m11	0.334***	p22	0.335***	t1	0.337**
	(1.990)		(0.078)		(0.076)		(0.058)		(0.453)
Ashanti	6.779***	b2	-0.271***	m12	-0.274**	p23	-1.981***	t2	0.236**
	(1.972)		(0.072)		(0.085)		(0.217)		(0.213)
Brong Ahafo	5.349***	b3	5.513***	m13	-0.596**	p24	0.269***	t3	-22.872***
	(2.037)		(0.438)		(0.258)		(0.067)		(0.788)
Northern	1.764	b4	-0.263**	m14	-0.003	p33	3.732***	t4	1.365***
	(2.094)		(0.117)		(0.131)		(0.544)		(0.207)
Upper East	4.084	c1	0.149	m22	0.169***	p34	-0.487*	z1	-2.301***
	(3.235)		(0.111)		(0.027)		(0.250)		(0.269)
Land-owner	-1.846*	c2	-0.112***	m23	-0.792***	p44	-0.149	z2	0.658***
	(0.986)		(0.041)		(0.241)		(0.047)		(0.209)
Land Rent-out	-1.277	c3	2.326***	m24	0.286***	q11	0.198***	z3	-15.326***
	(3.848)		(0.431)		(0.092)		(0.062)		(0.704)
Female Head	1.577	c4	-0.290***	m33	1.126**	q12	-0.234**	z4	2.226***
	(1.105)		(0.076)		(0.531)		(0.112)		(0.356)
Age of Head	0.034	d1	0.278**	m34	-0.310	q13	-0.840***		
	(0.082)		(0.109)		(0.331)		(0.285)		
$(\text{Age of Head})^2$	-0.001	d2	-0.133***	m44	-0.019	q14	0.152		
	(0.001)		(0.042)		(0.113)		(0.173)		
Highest Educ-	-0.387***	d3	2.879***	n11	0.968***	q22	0.195***		
ation Attained	(0.106)		(0.443)		(0.103)		(0.027)		
Loan	-1.026	d4	-0.263**	n12	-1.042***	q23	-1.251***		
	(1.001)		(0.121)		(0.147)		(0.202)		
Min. distance	0.001	e1	0.156	n13	-0.681**	q24	0.513***		
to school	(0.004)		(0.162)		(0.280)		(0.090)		
Housing Index	-0.217	e2	-0.156*	n14	0.343**	q33	2.358***		
	(0.454)		(0.093)		(0.134)		(0.352)		
a11	1.611***	e3	7.353***	n22	0.762**	q34	-1.308***		
	(0.470)		(0.449)		(0.117)		(0.260)		
a12	8.762***	e4	-0.235***	n23	-2.454***	q44	-0.178***		
	(0.560)		(0.067)		(0.312)		(0.064)		
a13	6.345***	f1	0.704***	n24	0.347**	r1	-0.800***		
	(0.803)		(0.109)		(0.122)		(0.209)		
a14	-2.714***	f2	-0.336***	n33	5.299***	r2	0.580***		
	(0.591)		(0.093)		(0.555)		(0.145)		
a22	-5.853***	f3	5.443***	n34	-1.558***	r3	-13.587***		
	(0.438)		(0.413)		(0.269)		(0.764)		
a23	19.579***	f4	-0.717***	n44	-0.075	r4	1.096***		
	(0.762)		(0.141)		(0.091)		(0.249)		

Note: ***99% , **95% , *90% confidence level.

Table 6. Input – output elasticities (full sample, 2138 observations)

Elasticities	SUR estimates		Half Normal Stochastic Frontier estimates	
	Value	Std. Error	Value	Std. Error
ε_{D,Q_1}	-0.117*	0.091	0.048	0.037
ε_{D,Q_2}	-0.524***	0.185	-0.143***	0.049
ε_{D,Q_3}	-0.102*	0.106	-0.039	0.045
ε_{D,Q_4}	-0.122**	0.066	0.029	0.030
$\sum_i \varepsilon_{D,Q_i} = -1$	0.136	0.285	0.892***	0.081
ε_{D,x_1}	0.239***	0.049	0.594***	0.128
ε_{D,x_2}	0.383***	0.051	-0.016	0.028
ε_{D,x_3}	0.373***	0.060	0.429***	0.036
ε_{D,x_4}	0.005	0.044	-0.008	0.022
Technical Efficiency[1]	0.166	0.146	0.997	0.004

Notes: Q_1 = cash crops, Q_2 = food crops, Q_3 = livestock, Q_4 = Off-farm x_1 = land, x_2 = Operating expenses, x_3 = workers, x_4 = livestock.
Standard errors and hypothesis testing on the first column is based on the bootstrapped empirical distribution of each statistic (B = 7000).
***99%, **95%, *90% confidence level
1) In the SUR columns we report DEA estimates; in the frontier column, technical efficiency is calculated with the estimated one-sided error of the stochastic frontier model.

Table 7 .

A. Cross-output elasticities: Cost Complementarities

	Cash Crops	Food Crops	Livestock
Food Crops	0.059*** (0.017)		
Livestock	0.018** (0.012)	0.081*** (0.022)	
Non-Farm	0.021*** (0.010)	0.047*** (0.013)	0.032*** (0.011)

B. Cross-term elasticities: Input responses to non-farm output expansion

	Land size (x_1)	Purchased inputs(x_2)	Workers (x_3)	Livestock (x_4)
$\partial \ln x_i / \partial \ln Q_4$	7.458** (3.563)	7.651** (3.702)	7.391** (8.191)	2.295 (57.189)

Note: Standard errors and hypothesis testing based on a bootstrapped empirical distribution of each statistic (B = 7000). ***99%, **95%, *90% confidence level

Table 8. Elasticities by farm size.

Elasticities	Less than 1 acre 188 obs.		Between 1 and 2.5 acres 369 obs.		Between 2.5 and 5 acres 542 obs.	
	Value	Std. Error	Value	Std. Error	Value	Std. Error
ε_{D,Q_1}	-0.048***	0.020	-0.043**	0.024	-0.044*	0.028
ε_{D,Q_2}	-0.202***	0.041	-0.268***	0.063	-0.345***	0.093
ε_{D,Q_3}	-0.161***	0.049	-0.149***	0.051	-0.182***	0.072
ε_{D,Q_4}	-0.271***	0.054	-0.299***	0.066	-0.249***	0.057
$\sum_i \varepsilon_{D,Q_i} = -1$	0.318*	0.126	0.241	0.147	0.181	0.176
ε_{D,x_1}	0.148***	0.035	0.188***	0.040	0.221***	0.043
ε_{D,x_2}	0.400***	0.033	0.394***	0.031	0.375***	0.033
ε_{D,x_3}	0.357***	0.038	0.356***	0.040	0.361***	0.044
ε_{D,x_4}	0.096**	0.041	0.062*	0.035	0.043	0.033
TE[1]	0.2493	0.2426	0.1909	0.2008	0.1575	0.1635
	Between 5 and 10 acres 558 obs.		Greater than 10 acres 481 obs.			
ε_{D,Q_1}	-0.098	0.097	-0.727	13.666		
ε_{D,Q_2}	-0.573***	0.266	-2.295	64.338		
ε_{D,Q_3}	-0.264**	0.188	-0.685	11.940		
ε_{D,Q_4}	-0.328***	0.120	-0.553	13.479		
$\sum_i \varepsilon_{D,Q_i} = -1$	-0.262	0.509	-3.260	98.389		
ε_{D,x_1}	0.239***	0.066	0.210	4.867		
ε_{D,x_2}	0.346***	0.063	0.350	3.105		
ε_{D,x_3}	0.405***	0.073	0.513	1.726		
ε_{D,x_4}	0.010	0.057	-0.073	3.670		
TE[1]	0.1632	0.1555	0.1578	0.1362		

Note: Hypothesis testing based on the bootstrapped empirical distribution of each statistic (B = 7000).***99%, **95%, *90% confidence level

1) DEA estimates of technical efficiency

Table 9. Determinants of Input Oriented Technical Efficiency

DEA Input Technical Efficiency	OLS		2 - Stage		3 - Stage	
	Value	Std. Error	Value	Std. Error	Value	Std. Error
Farm Output	0.043***	0.003	-0.074***	0.013	-0.037***	0.006
Off-farm Output	0.002***	0.000	0.022*	0.012	-0.001	0.001
Western	0.007	0.014	0.079***	0.019	0.053***	0.018
Central	-0.003	0.014	0.035**	0.017	0.019	0.016
Greater Accra	-0.018	0.030	-0.082**	0.032	-0.047	0.036
Eastern	-0.002	0.014	-0.042***	0.015	-0.030*	0.017
Volta	0.032**	0.013	0.072***	0.017	0.054***	0.016
Ashanti	-0.031**	0.014	0.031*	0.019	0.008	0.017
Brong Ahafo	-0.046***	0.014	0.067*	0.026	0.010	0.018
Northern	-0.037***	0.021	-0.005	0.017	-0.031*	0.017
Upper East	0.016	0.006	-0.042*	0.025	-0.051*	0.026
Land-owner	-0.009	0.025	0.024***	0.009	0.014*	0.008
Land Rent-out	-0.048*	0.007	-0.074***	0.018	-0.041	0.030
Female Head	0.025***	0.001	-0.023**	0.011	-0.008	0.009
Age of Head	-0.002*	0.000	-0.003*	0.002	-0.001	0.001
$(\text{Age of Head})^2$	0.000	0.001	0.000**	0.000	0.000	0.000
Highest Education Attained	-0.001	0.006	0.002**	0.000	0.002***	0.001
Loan	-0.000	0.006	0.019**	0.008	0.006	0.008
Min. distance to school	0.000	0.000	0.000	0.000	0.000	0.000
Housing Index	0.007**	0.003	-0.010	0.007	-0.003	0.004
Constant	-0.376***	0.046	0.887	0.114	0.669***	0.083
N	2138		2138		2138	
R^2 and Pseudo-R^2	0.15		0.07		0.16	

Notes: ***99%, **95%, *90% confidence level

Data Appendix

Our three farm outputs – cash crops, food crops, livestock and other crops – are measured as total value of production. Cash crops, as identified by the survey, include cocoa, coffee, pineapple, sheanut/butter, cola nut, cotton, tobacco and sugarcane. On the other hand food crops include any roots, fruits, vegetables and other crops harvested piecemeal. Within this category maize and rice have also been included, even if in other countries these two products are more likely to be considered as cash crops. In the third farm output aggregate we have summed up livestock and *other crops* value of production. As we discussed above, livestock is defined as the sum of in-cash and in-kind incomes from livestock produce, plus sales and rents from livestock and the value of own consumption. *Other crops* instead include kenef, rubber and wood. This residual category of *other crops* presents generally very low values. Therefore even for the few households which have positive levels of *other crops*, it does not alter the interpretation of this aggregate as mostly the livestock output.

With respect to the livestock as an input, we tried to convert different species to a common livestock unit. The concept of Tropical Livestock Units (TLU), for example, provides a convenient method for quantifying different livestock types; however, we do not have all the conversion factors for livestock species present in rural Ghana. Thus we first explored the possibility to value livestock at their sale price, i.e. value of the sales divided by the quantity sold. Then we imputed the median price of the cluster to those observations which lacked either sale values or quantities or both of them. With this method we had to recur to imputation for a large number of observations. Therefore we decided to value livestock at the price the household member thought was the sale price. We also calculated the mean level of input by calculating initial and ending period livestock levels. The final livestock level corresponds to the livestock in hand when the household was surveyed (k(1)). The number of units owned at the beginning of this period (k(0)), was obtained by adding the sales and subtracting the purchases that occurred within the previous twelve months. Then we measure the mean livestock units as the simple average of the initial and ending stocks. Observations with negative average values, as well as with estimated negative initial stocks were dropped.

Operating expenses include both agricultural and non-farm enterprises expenditures. In the survey agricultural costs are divided into crop costs, livestock costs and fishing costs. Crop costs include the yearly expenses for several inputs such as fertilizers, insecticides, herbicides, storage, seed and seedlings, irrigation, bags, petrol/diesel/oil, transport, renting animals and equipment, spare parts and hired labor. Livestock costs include animal feed, veterinary services, paid labor, maintenance of stables, transport of feed, commission on sale, compensation for damage. Here we did not take into account the cost of land-leasing, since we have already used plots' size in acres as the variable which measures land usage.

As far as the control variables concerns, the housing quality index was created through principal component analysis, where the variables used are the source of water, the source of light, the source of energy, the materials used for the walls and the roof, and the bathroom facility type. We retained the first three factors, whose eigenvalues were equal to 3.29, 0.98 and 0.91. In practice our choice was not very far from the Kaiser criterion, which suggests the retention of factors with eigenvalues greater than 1. In this case we would extract at least as much as the equivalent of one original variable. The cumulative variance explained by the first three components is equal to 74%, which is not overwhelming but substantial.

Finally, the issue of outlier observations was tackled in several steps. First, we constructed measures of partial productivity such as the total value of production divided by land usage, number of workers, and operating expenses. Outliers were defined as units with a partial productivity index greater in absolute terms than the median plus three times the standard deviation. Secondly, we ran a regression of the logarithm of the total value of the output over the logarithm of the inputs and some control variables. As we did in the previous outlier check, units with the residual greater in absolute terms than three standard deviations were considered outliers.

References

Aigner, Dennis, C A Knox Lovell and Peter Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, **6**(1): 21-37.

Anríquez, Gustavo and Silvio Daidone (2008), "Linkages between Farm and Non-Farm Sectors at the Household Level in Rural Ghana. A consistent stochastic distance function approach," *FAO ESA Working Papers* Rome.

Anríquez, Gustavo and Kostas Stamoulis (2007), "Rural Development and Poverty Reduction: Is Agriculture Still the Key?," *FAO, ESA Working Papers* Rome**:** 41.

Atkinson, Scott E and Daniel Primont (2002), "Stochastic Estimation of Firm Technology, Inefficiency, and Productivity Growth Using Shadow Cost and Distance Functions," *Journal of Econometrics*, **108**(2): 203-25.

Chavas, Jean Paul, Ragan Petrie and Michael Roth (2005), "Farm Household Production Efficiency: Evidence from the Gambia," *American Journal of Agricultural Economics*, **87**(1): 160-79.

Coelli, Tim (2000), "On the Econometric Estimation of the Distance Function Representation of a Production Technology," *Working Paper* Louvain**:** 24.

Coelli, Tim, Gholamreza Hajargasht and C.A Knox Lovell (2007), "Econometric Estimation of an Input Distance Function in a System of Equations," *Working Paper, Centre for Efficiency and Productivity Analysis* Brisbane**:** 48.

Davis, Benjamin, Thomas Reardon, Kostas Stamoulis and Paul Winters (2002), *Promoting farm/non-farm linkages for rural development. Case Studies from Africa and Latin America*, Rome, FAO.

Davis, Benjamin, Paul Winters, Gero Carletto, Katia Covarrubias, Esteban Quinones, Alberto Zezza, Kostas Stamoulis, Genny Bonomi and Stefania DiGiuseppe (2007), "Rural Income Generating Activities: A Cross Country Comparison," *Background paper for the World Development Report 2008* Rome, FAO.

Färe, Rolf, Shawna Grosskopf and C A Knox Lovell (1994), *Production frontiers*, Cambridge; New York and Melbourne, Cambridge University Press.

Farrell, M.J. (1957), "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, **125**(3): 253-81.

Ghana Statistical Service (1999), "Data User's Guide. Ghana Living Standards Survey Round Four (GLSS 4) 1998/99," Accra.

Ghana Statistical Service (2000), "Poverty Trends in Ghana in the 1990s," Accra, Ghana Statistical Service**:** 70.

Gorman, Ian E (1985), "Conditions for Economies of Scope in the Presence of Fixed Costs," *RAND Journal of Economics*, **16**(3): 431-36.

Grosskopf, Shawna and Kathy Hayes (1993), "Local Public Sector Bureaucrats and Their Input Choices," *Journal of Urban Economics*, **33**(2): 151-66.

Haggblade, Steven, Peter Hazell and James Brown (1989), "Farm-Non Farm Linkages in Rural Sub-Saharan Africa," *World Development*, **17**(8): 1173-1202.

Haggblade, Steven, Peter Hazell and Thomas Reardon, Eds. (2007), Transforming the Rural Nonfarm Economy: Opportunities and Threats in the Developing World Baltimore, The Johns Hopkins University Press

Hirschman, Albert O. (1958), *The Strategy of Economic Development*, New Haven, Connecticut, Yale University Press.

Katz, Eliakim and Oded Stark (1986), "Labor Migration and Risk Aversion in Less Developed Countries," *Journal of Labor Economics*, **4**(1): 134-49.

Kumbhakar, Subal C and Giannis Karagiannis (2004), "On the Equivalence of Two Normalizations in Estimating Shadow Cost Functions," *Economics Letters*, **82**(1): 15-19.
Kumbhakar, Subal C. and C.A. Knox Lovell (2000), *Stochastic Frontier Analysis*, Cambridge, Cambridge University Press.
Kuznets, Simon S. (1957), "Quantitative Aspects of the Economic Growth of Nations II," *Economic Development and Cultural Change*, **4**(Supplement): 3-11.
Maietta, Ornella Wanda (2002), "The Effect of the Normalisation of the Shadow Price Vector on the Cost Function Estimation," *Economics Letters*, **77**(3): 381-85.
Reardon, Thomas, Julio Berdegué, Christopher B. Barret and Kostas Stamoulis (2007), "Household Income Diversification into Rural Nonfarm Activities," *Transforming the Rural Nonfarm Economy.Opportunities and Threats in the Developing World*, S. Haggblade, P. Hazell and T. Reardon, Eds., Baltimore and London, The John Hopkins University Press.
Reardon, Thomas, Kostas Stamoulis, M.E. Cruz, A. Balisacan, Julio Berdegué and B. Banks (1998), "Rural Non-farm Income in Developing Countries," *FAO: the State of Food and Agriculture, 1998*, Rome, FAO.
Singh, Inderjit, Lyn Squire and John Strauss, Eds. (1986), Agricultural Household Models. Extensions, applications, and policy. Baltimore and London, The John Hopkins University Press for the World Bank.
Valdés, Alberto and William Foster (2003), "The Positive Externalities of Chilean Agriculture: The Significance of its Growth and Export Orientation," *ROA Project - FAO* Rome.
Vogel, Stephen J (1994), "Structural Changes in Agriculture: Production Linkages and Agricultural Demand-Led Industrialization," *Oxford Economic Papers*, **46**(1): 136-56.
Wooldridge, Jeffrey M. (2001), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachusetts, The MIT Press.

ESA Working Papers

WORKING PAPERS

The ESA Working Papers are produced by the Agricultural Development Economics Division (ESA) of the Economic and Social Development Department of the United Nations Food and Agriculture Organization (FAO). The series presents ESA's ongoing research. Working papers are circulated to stimulate discussion and comments. They are made available to the public through the Division's website. The analysis and conclusions are those of the authors and do not indicate concurrence by FAO.

ESA

The Agricultural Development Economics Division (ESA) is FAO's focal point for economic research and policy analysis on issues relating to world food security and sustainable development. ESA contributes to the generation of knowledge and evolution of scientific thought on hunger and poverty alleviation through its economic studies publications which include this working paper series as well as periodic and occasional publications.

Agricultural Development Economics Division (ESA)
The Food and Agriculture Organization
Viale delle Terme di Caracalla
00153 Rome
Italy

Contact:
Office of the Director
Telephone: +39 06 57054358
Facsimile: + 39 06 57055522
Website: www.fao.org/es/esa
e-mail: ESA@fao.org