

Farm Size and Productivity  
A “Direct-Inverse-Direct” Relationship

*Sara Savastano*  
*Pasquale L. Scandizzo*



**WORLD BANK GROUP**

Development Economics

Global Indicators Group

June 2017

## Abstract

This paper proposes a new interpretation of the farm size–productivity relationship. Using two rounds of the Ethiopian Rural Household Survey, and drawing on earlier work on five countries in Sub-Saharan Africa, the paper shows that the relationship between farm size and productivity is neither monotonic nor univocal. Most previous studies that tested the inverse farm size–productivity relationship used ordinary least squares estimation, therefore reporting parameter estimates at the conditional mean of productivity. By expanding these important findings to consider the entire distribution of agricultural productivity, the analysis finds sign switches across the distribution, pointing to a “direct-inverse-direct” relationship. Less productive farmers exhibit an inverted U-shape relationship between land productivity and farm size, while more productive farmers show a U-shape relationship that

reverses the relationship. In both cases, the relationship points toward a threshold value of farm size; however, the threshold is a minimum for the less productive farmers and a maximum for the more productive ones. To the left of the threshold, for very small farmers, the relationship between productivity and farm size is positive; for the range of middle farm size, the relationship is negative; and to the right of the threshold, the relationship is direct (positive) again. From a policy perspective, these findings imply that efficiency-enhancing and redistributive land reform should consider farm size in the proper context of the present and potential levels of agricultural productivity. The results and their policy implications underline the relevance of the most recent efforts of the international development community to collect more reliable georeferenced data on farm size and agricultural productivity.

---

This paper is a product of the Global Indicators Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [ssavastano@worldbank.org](mailto:ssavastano@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# Farm Size and Productivity: A “Direct-Inverse-Direct” Relationship

Sara Savastano<sup>1,2</sup> and Pasquale L. Scandizzo<sup>2</sup>  
<sup>1</sup>World Bank, <sup>2</sup>University of Rome Tor Vergata -DEF

**JEL codes:** O12, Q12, Q15

*Keywords:* Agricultural Policies, Agriculture Productivity, Agriculture and Food Security, Ethiopia

Sara Savastano is Senior Economist at the Global Indicators Group, Development Economics and Assistant Professor of Economics at the University of Rome Tor Vergata (currently on leave). Pasquale L. Scandizzo is professor at the University of Rome Tor Vergata. For their comments the authors thank participants of the conference on “Farm Size and Productivity: A Global Look” organized by USDA – ERS & Farm Foundation on February 2-3, 2017, and participants of the 21th ICABR Conference on “Bioeconomy in Transition: New Players and New Tools”, Berkeley May 30 – June 2, 2017. The authors acknowledge the FAO & CEIS-Tor Vergata Letter of Agreement on “Smallholder agriculture in transition: Behavior, constraints and policies in Ethiopia” for the use of the data.

## Introduction

The existence of an inverse farm size–productivity relationship (IR), defined as a smooth tendency of productivity to decline with farm size, has been considered as a well-established finding of agricultural economics, and a stylized fact to guide rural development strategies and land reform to promote efficiency and overall equity. First observed in Russia (Chayanov, 1926), then in the dualistic agriculture labor market in India (Bardhan 1973, Sen 1975, Srinivasan 1972), IR was initially explained by labor market imperfections and supervision constraints, with small-scale farmers having lower opportunity and supervision costs of their labor than operators of larger farms.

From the economic perspective, IR implies that diseconomies of scale characterize agriculture systems for several possible reasons, including the failure of land and labor markets to equalize production efficiency across farm size distribution. From the policy perspective, reforms to redistribute land from large towards small farms appear justified not only on equity but also on efficiency grounds.

In most empirical studies, IR appears to be a smooth tendency of land productivity to decline with farm size and thus is not limited to a different pattern of resource uses between large and small farms. While different reservation wages could account for family versus non-family farms, this would not explain why land productivity appears to decline within small family farms as well.

The IR hypothesis has been tested in a variety of studies using households and plot level data, in cross sectional or country frameworks, over time using panel data to account for farmers' and plots' unobserved heterogeneity, and controlling for land quality. The availability of a more accurate measure of land area by means of GPS has seemed to add further proof for the existence of IR, even after eliminating farmers' self-reported land bias. More recently, the use of geo-referenced measures of soil quality and infrastructure has helped in rejecting the hypothesis that IR might be the fruit of statistical artefacts.

In this paper, we analyze the land productivity–farm size relationship using a quantile regression approach to test whether farm size affects differently farmers with low productivity as compared with farmers of average or higher levels of productivity. In this framework, we aim to understand how farm size affects farmers' productivity achievements across the whole distribution of land productivity.

The large majority of the IR studies have modeled farm productivity as a function of land area and other covariates using an OLS approach. The use of georeferenced variables that correct for measurement errors, objective measures of soil quality, and use of panel data to account for farmer and plot heterogeneity do not change this core framework, which is essentially based on the study of the effect of farm size on the conditional mean of land productivity. However, this approach clearly provides only a partial view of the relationship, as we might be interested in describing the relationship at different points in the conditional distribution of agriculture productivity. By supplementing the estimation of conditional mean functions with techniques for estimating an entire family of conditional quantile functions, we aim to provide a more complete statistical analysis of the stochastic relationships among random variables.

Quantile regression is not a novel approach. The econometric specification was introduced in 1968 by Koenker and Basset. More recent empirical applications concerned education (Koenker and Biliias, 2001) and peer effects (Levin, 2001), analysis of duration of employment spell (Horowitz and Neumann, 1987; Fitzenberger, 1997), ecology (Cade and Noon, 2003; Chamaillé-Jammes and Blumstein, 2012; and Lemke K., 2012), health economics (Johar and Katayama, 2012), and in the context of political economy aspects of trade policy (Billger, 2009; Imai et al., 2012). To our knowledge, application to agriculture policies are found only in Evenson and Mwabu (E-M, 2001) and Scandizzo and Savastano (S-S, 2017). E-M studied the impact of Kenya's extension service, and found that productivity response to acreage, measured through quantile regression, was not significantly different across quantiles, but displayed a concave shape, first rising and then falling with the size of the cultivated area. Testing the IR on national representative surveys from five Sub-Saharan African countries, S-S pointed out that IR may be an artifact of the central tendency indicators

captured by the OLS technique, and depending on conditional expectations. Using quantile regression, S-S found a U-shape and an inverted U-shape relation with sign switches at different points of the distribution of farms' productivity in all the countries.

In this paper, we corroborate the above S-S results, using two rounds of the Ethiopian Rural Household Survey (ERHS),<sup>1</sup> a unique multi-purpose longitudinal household data set representative of the rural population of the country. The contribution of our paper is twofold. First, we test the productivity–farm size relationship, taking into account farmers' heterogeneity; second, we consider the possible interaction between productivity and management quality and efficiency.

By considering the entire distribution of agriculture productivity, we found sign switches across the distribution pointing to the existence of a “Direct-Inverse-Direct” (DID) relationship. Less productive farmers exhibit an inverted U-shape relationship between land productivity and farm size, while the more productive farmers show a U-shape relationship that reverses the relationship. In both cases the relationship points towards the existence of a threshold value of farm size, which is, however a minimum for the less productive farmers and a maximum for the more productive ones. The results show that the inverse relationship holds for middle size farms. To the left of the threshold, for very small farmers, the relationship between productivity and farm size is positive, and farmers can invest in the extensive margins through land expansion as a safety-first strategy to achieve a minimum level of production to guarantee subsistence. For the range of middle farm size, the relationship is negative. Middle farmers tend to invest in new agriculture technologies and income diversification to increase their productivity. They may improve their position by increasing their intensive margins, namely getting more productivity from land already under cultivation. To the right of the threshold the relationship is direct (positive) again.

Our findings indicate that, while IR may be important for certain ranges of farm efficiency and size, it is by no means a ubiquitous characteristic of agriculture. Whether the relationship between productivity and size is positive or negative in fact appears to depend crucially on other factors, including agro-ecological conditions and non-farm opportunities.

The plan of the rest of the paper is as follows. First, we provide some background and motivation through a brief survey and discussion of the literature. Second, we describe the conceptual framework and the estimation strategy. We turn next to a discussion of the model used, the data employed, the estimation techniques and the discussion of the results. The paper concludes with policy recommendations.

## **1. Background and Motivation**

The fact that land productivity is negatively related to the size of the land operated (both owned and rented) appears historically established by a variety of studies starting in the late 1970s (e.g. Berry and Cline, 1979, Kutcher and Scandizzo, 1981, Binswanger et al., 1995).

The literature has emphasized different explanations for this “empirical regularity”. The early studies focused on factor market imperfections in land and other markets such as credit and modern inputs (Scandizzo and Kutcher, 1981; Eswaran and Kotwal, 1985; 1986; Barrett, 1996; Benjamin and Brandt 2002; Berry and Cline 1979; Feder 1985; Binswanger et al. 1995, Ali and Deininger, 2014). More recently the literature has focused on statistical issues and measurement errors in land and agriculture output, omission of soil quality measurements that are inversely correlated with farm or plot size but positively associated with yields (Lamb, 2003; Barrett et al., 2010; Goldstein and Udry, 1999, De Groot and Traorè, 2005). The availability of sophisticated remote sensing and georeferenced data has increased the statistical debate of the IR. Carletto et al. (2013 and 2015) show that the IR persists even after controlling for more objective GPS-based land area measures. Deininger et al. (2012) and Kilic et al. (2017) showed that the inverse relationship could be driven by errors in self-reported survey data on crop production. Desiere and Jolliffe

---

<sup>1</sup> The project was conducted by the Department of Economics at Addis Ababa University, the Centre for the Study of African Economies (CSAE), the University of Oxford and the International Food Policy Research Institute (IFPRI).

(2017) compared self-reported production and crop-cuts estimates, and found that the IR is stronger when based on self-reported production, but disappears when based on crop-cut estimates. Finally, Bevis and Barrett (2017) identified the “edge of the plot” effect, namely that productivity may be greater along the edge of small plots where labor is concentrated. The standard explanation of the IR is the presence of omitted variables and unobserved heterogeneity: omission of soil quality measurements that are inversely correlated with farm or plot size but positively associated with yields (Carter 1984; Bhalla and Roy 1988; Walker and Ryan, 1990; Benjamin 1995; Lamb 2003; Assunção and Braido, 2007; Assunção and Ghatak, 2007; Benjamin, 1995).

The IR finding, however robust across many studies, seems at the same time puzzling, for several reasons. First, it is not limited to small versus large farms, but in most studies, there is a smooth tendency for land productivity to decline with farm size. This result seems to be in contrast to the equalization of factor prices predicted by market equilibrium theory and not simply explained by lower reservation wages for family farms, because land productivity appears to decline within all ranges of family as well as non-family farms. In this respect, Feder’s (1985) alternative explanation appeals to a more general transaction effect, reminiscent of Coase’s theory of the firm. According to this explanation, smaller farms are based on more intense use of family labor, because of its higher efficiency and motivation than hired labor, and the fact that supply of working capital is directly related to farm size. Srinivasan (1972) explains the inverse relationship by yield risk, by defining utility over income, and imposing restrictions on the coefficients of risk aversion and on how risk enters production, under constant returns. Hazell and Scandizzo (1974) provide a rationale for producers to reduce planned production in response to the negative correlation between supply and prices, and Barrett (1996) shows that IR can emerge from price risk if farmers are net buyers of the crop produced, since in this case, risk aversion implies labor overemployment to protect consumption.

As Savastano and Scandizzo (2009) have shown, a relationship between productivity and operated area may arise, because of the risk associated to the investment decision such as, for example, to increase one’s farmland. Under dynamic uncertainty, in fact, the amount of land operated by a farmer will depend on the timing of the exercise of the option to invest in land development. With decreasing returns to scale, this will imply a non-monotonic relationship between revenue per ha and operated land. If land is available on the market in fixed quantities (i.e. supply of plots for rents or sale, or entire farms of discrete size), and/or investment is lumpy, small farms will exhibit lower revenue thresholds for investment, and thus lower revenues per unit of land than larger farms. This implies, in particular, that the relationship between productivity and size may exhibit turning points, as farmers switch from one type of investment to another (e.g. from land improvements to irrigation) as their operating land increases as a result of previous investment decisions.

Most recently, using data from five Sub-Saharan African countries, Scandizzo and Savastano (2017) found a nonlinear relationship between farm size and land productivity, and signs switches across farm size groups. The analysis overcomes the statistical and measurement biases by testing the IR with the LSMS-ISA, nationally representative geo-referenced surveys of Malawi, Niger, Nigeria, Tanzania and Uganda. The estimation approach used also controls for many exogenous common and comparative geo-spatial measures of land quality, infrastructure and access to markets, climate conditions, soil conditions and topography.

Using quantile regression on a cross sectional framework, the results showed that average land productivity (ALP) exhibits an inverted U-shaped relation for less productive farmers, with ALP first rising and then falling after a threshold farm size. ALP shows the opposite pattern of a U-shaped relationship in the upper tail of the productivity distribution, first decreasing and then increasing after reaching a lower threshold. Farms in the lower tail of the ALP distribution thus experienced IR only once they have reached a critical size after which only by moving to a higher ALP class can they avoid the IR relationship from diseconomies of scale. Vice versa, more productive farmers experience IR only if they are below a critical size, which, in general, tends to be larger (and sometime much larger) than the critical size of the lower end

farms. Thus, small and large farm behavior tends to diverge, since farms in the lower deciles of the land productivity distribution experience IR for a smaller range of farm sizes than farmers of the higher deciles. The results are consistent across countries, and in the pooled analysis.

The main conclusions of this analysis, which are corroborated by the results in the present paper, can be summarized as follows:

- (1) ALP will display respectively a negative (the IR) or a positive correlation with farm operating size, depending on whether the threshold effect (the higher incentive for larger farms to hold undeveloped land as an option) prevails or is overwhelmed by the management effect (the tendency of ALP to increase with farm size since larger farms attract better managers).
- (2) Depending on the functional form of their relationship, the two tendencies may equal each other, once a threshold of farm size is reached, after which the net effect on ALP will be reversed.
- (3) Both tendencies and the level of the threshold will depend on the management quality and thus can be expected to vary across farmers, depending on the distribution of management quality and the extent to which the market for managers succeeds in allocating them to larger farms.

These conclusions support the idea that the IR relationship may indeed be present in many farming systems, but we should expect it to be neither ubiquitous nor monotonic. In particular, if farmers face dynamic uncertainty by holding a waiting option for land development and unobservable management quality is positively correlated with farm size, both a reverse and a direct effect of operating size on average productivity may be present at any one time. This in turn implies that the net impact of increasing operating size on land productivity will depend on whether a threshold is crossed where the two effects exactly balance each other.

On the technical estimation side, we further argue that OLS will generally provide an estimate of the relationship based on mean response if factor productivity is distributed normally, with a constant variance. In other words, OLS will allow us to estimate a response coefficient that will quantify the average response of the dependent variable (e.g. land productivity) to farm size increases. If the distribution of the response around the mean, estimated according with OLS, is not satisfactorily described by a single variance, however, quantile regression (Koenker and Bassett, 1978) promises a more robust and appropriate estimate, especially if variance is systematically related to the increase in the response variable (heteroscedasticity). We also conjecture that the relationship between productivity and alternative measures of size (land available, land under cultivation etc.) may be considerably different for farmers who, for various reasons that cannot be captured by the econometric model, have to operate at lower productivity levels, with respect to farmers that operate at higher productivity levels.

## 2. Conceptual Framework

Consider the relationship between land productivity and farm size in the stylized form. For simplicity, we omit the time period:

$$\frac{y_i}{x_i} = \beta_0 + \beta_1 x_i + \gamma Z_i + \epsilon_i \quad (1)$$

where  $y_i$  is some measure of production for the  $i$ th farm,  $x_i$  is a correspondent measure of farm size (e.g. operated area),  $Z_i$  a set of exogenous variables, and  $\epsilon_i$  a random disturbance. It is important to underline the fact that equation (1) is not a production function, but the result of farmers' choices, on the basis, inter

alia, of an underlying technology. If we assume that farmers have adjusted production (either through optimization or through any other behavioral rule) to the circumstances outside their control, including exogenous variables, states of nature etc., the coefficient  $\beta_1$  in (1) should be zero. In other words, all systematic differences in production per acre between farms should be accounted for by differences in the  $Z_i$  variables or in the random term  $\epsilon_i$ . A  $\beta_1$  different of zero, on the other hand, would imply the existence of systematic differences across farmers that are not accounted for in the rest of the equation: these differences could be due to different behavioral rules, different abilities in following the same rules or different levels of information or other omitted variables that are correlated with farm size.

It is also important to notice that a non-zero  $\beta_1$  may be caused by discontinuities in the behavioral function that underlies farmers' adjustment to the exogenous variables. These discontinuities are implied by most of the explanations of the inverse productivity relationship based on anthropological differences between "family" and "non-family" farms or systematic divergence in behavior between "small" and "large" farms (e.g. Feder, 1989; Cornia, 1985). However, if IR is the result of these discontinuities, it should only concern the differences across the two extreme groups of farmers, and not the differences within the groups themselves.

In order to test for the existence of a relationship between land productivity and farm size, we use both OLS regression models and quantile regressions (Koenker and Bassett, 1978). While OLS focuses on modeling the conditional mean of the response variable without accounting for its distribution, the quantile regression model accounts for the full conditional distributional properties of the response variable (or is residual after accounting for the exogenous variables) thereby differing on the assumptions about the error terms of the regression model.

In the case of equation (1), the OLS model is based on the assumption that the error term is normally distributed with zero mean and constant variance:  $\epsilon_i \sim \text{i. i. d. } N(0, \sigma^2)$ .

The mean zero assumption on the error term implies that the model fits the conditional mean, namely  $E[y - \gamma Z|x] = \beta_0 + \beta_1 x_i$  which can be interpreted as the average value of productivity, after accounting for the effect of the exogenous variables  $Z$ , corresponding to a given value of the covariate  $x$  (i.e. farm size). The linear regression model describes how the conditional distribution behaves by utilizing the mean of a distribution to represent its central tendency, a choice that appears appropriate under the assumption of homoscedasticity, namely of constant variance for all values of the covariate  $x$ .

The quantile-regression model (QRM) estimates the potential differential effect of a covariate (farm size) on various quantiles in the conditional distribution. A conditional quantile is a statistic corresponding to the probability level of a given distribution, according to a function (the quantile function) defined as  $q(p) = \{y: \Pr(Y \leq y) = p\}$ . By considering the different quantiles, the QRM estimates how the effect of a covariate varies with the distribution of the response variable and accommodates heteroscedasticity. The QRM corresponding to the LRM in Equation (1) can be expressed as:

$$y_i = \beta_0^{(q)} + \beta_1^{(q)} x_i + \gamma^{(q)} Z_i + \epsilon_i^{(q)} \quad (2)$$

The parameter vector,  $[\beta_0^{(q)} \beta_1^{(q)} \gamma^{(q)}]$  is obtained by minimizing the sum of absolute deviations from an arbitrarily chosen quantile of a farm yield across farmers. In the case of Equation (2) this sum can be expressed as:

$$\text{Minimize } \sum_i |y_i^q - [\beta_0^{(q)} + \beta_1^{(q)} x_i + \sum_j \gamma_j^q Z_{ij}]| \quad (3)$$

where  $y_i^q$  = average productivity for farmer  $i$  at quantile  $q$ , ( $i=1, \dots, n$ );  $x_i$ = farm size  $Z_{ij}^q$  = covariate  $j$  for farmer  $i$  ( $j=1, \dots, K$ ).



The solution to Equation (3) is found by rewriting the expression as a linear programming problem over the entire sample (see Chamberlain, 1994) and solving for the values of the parameters. Both the squared-error and absolute-error loss functions are symmetric, as the sign of the prediction error is not relevant. While OLS can be inefficient if the errors are highly non-normal, quantile regression is more robust to non-normal errors and outliers. QR also provides a richer characterization of the data, gives information on heterogeneity in the effect of land on the productivity, and allows to consider the impact of a covariate on the entire distribution of  $y$ , not merely its conditional mean. In particular, for each quantile, it can be shown whether the effect of farm size on productivity is positive or negative, and how large this effect is compared to other quantiles.

### 3. Data Source and Descriptive Statistics

We use the last two rounds of the Ethiopian Rural Household Survey (ERHS),<sup>2</sup> a unique multi-purpose longitudinal survey conducted since 1989, which covers four of the nine administrative regions in Ethiopia, where the largest proportion of the country's farmers are located, and includes 15 woredas (districts) stratified over the three major agricultural systems found in five agro-ecological zones. Although not nationally representative, the sample is broadly representative for the main farming systems in the country.

Table 1 shows the main characteristics of farmers in both rounds. With 1.4 ha of land per household and 6 family members, small, resource-poor, subsistence farmers dominate the rural sector in Ethiopia. Given the large number of plots cultivated by a farmer (4.7 and 5.1 plots in 2004 and 2009 respectively), fragmentation and the lack of improved technologies are a major challenge for spurring agriculture productivity and to allowing farmers to meet their subsistence needs. While farm size is on average low and has changed very little between the two rounds (1.4 to 1.5 ha), since its distribution is highly skewed, we complement our descriptive analysis with the concept of midpoint farm size introduced by Hoppe et al. (2013). The statistic is defined as the size at which half of all cropland is on farms with more cropland than the midpoint, and half is on farms with less. It is a more informative measure of land consolidation than either a simple median (in which half of all farms are either larger or smaller) or the simple mean (which is average land per farm). We found that the midpoint farm size in Ethiopia is 1.6 ha in 2004, and 2.15 ha in 2009, 1/2 ha more than the value of average farm size, thereby indicating little variation given the presence of a very large number of very small farms and the high degree of fragmentation.

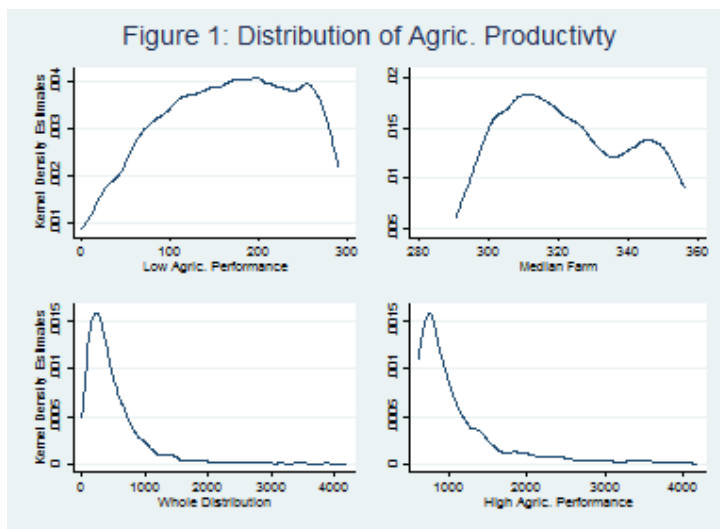
Only one-third of the farmers employ hired labor, and with an average of 6 members per household, they use family members to farm land. Though young (50 years on average), household heads have received little education, 1.6 years of school on average, and so did family members who obtained a maximum of 4.7 years of education.

Farmers rely heavily on agriculture, which accounts for 78% of total income. It is worth noting, however, that the share of income from agriculture has decreased of 9% from 2004 to 2009, while the opposite has occurred for wage and other income sources thereby suggesting a shift from owner-operated farming to wage labor, perhaps also as a consequence of the high food prices of 2008.

Agriculture productivity has increase over time from 446 to 549 US\$/ha with an average of both periods of 497 US\$/ha. We recognize the importance of mean effect, but one of the assumption of the OLS regression is that the error term – and the dependent variable – is normally distributed. To investigate whether this assumption is satisfied, we display in Figure 1 the kernel density estimates for gross agriculture income per hectare, separately by agriculture productivity below and above the median and along the whole distribution.

---

<sup>2</sup> The ERHS represents a collaborative effort of the Department of Economics at Addis Ababa University, the Centre for the Study of African Economies (CSAE), the University of Oxford and the International Food Policy Research Institute (IFPRI).



Source: Authors' calculation based on ERHS

None of the conditional distributions appears to be Gaussian, suggesting correlation across farms.<sup>3</sup> For low productivity performance, the distribution appears as an inverted U-shape, while for high performance it is skewed to the right. For the median productivity farm the distribution appears bimodal. This seems to justify the choice of quantile regression (QR), since, unlike OLS estimation, QR does not require a normally distributed error term.

We use information on households' land ownership and income to compute the Gini coefficient of the farmers in both survey years of the panel. With a Gini total income coefficient of 0.48 in the first and 0.46 in the second period, income inequalities appear to have decreased very little between the two periods. With a land Gini coefficient of 0.43 and 0.44 in each year respectively, inequality in assets is slightly lower than income inequality, a result that differs from most asset and income distributions in the rest of the world and is probably due to a combination of extreme poverty and small size farming. The higher level of the Gini coefficient for income also suggests that inequality may depend on different levels of ability rather than on asset holdings.

Farming conditions are relatively favorable with, on average, less than one plot per household with steep slope, 54 percent of land area with good soil quality, but with one-third of farmers being affected by climate change and reporting to have little rain on fields during the rainy season.

Under these conditions there are few agriculture investment options that farmers can exercise over time. For the sake of simplicity and based on the survey questions available in both rounds, we have identified three main options: land expansion, investment in soil conservation practices, and raising livestock.

Land expansion is the main option for farmers, with 52 percent of farmers having exercised the option to invest in new land, mostly through some kind of rental agreement. We only focus on those who increased their holdings, and we observe that, with an average size of land expanded of 0.75 ha, farmers doubled their land operation in five years. In line with the literature, a well-functioning land market allows land to be transferred to more productive producers and to those who want to stay in agriculture.

For Ethiopia, which faced a severe drought in 1973, investments in soil erosion, also promoted by the government, are an important way to increase agriculture productivity and income. As a second option,

<sup>3</sup> The central limit theorem posits random variables independently drawn from independent distributions converge in distribution to the normal with zero mean and constant variance.

we have computed the share of households who decided to invest in soil conservation practices in the second round of the panel. We observe that in 2009, 18 percent of farmers decided to combat land degradation and invest in one or more soil conservation practices such as physical structures (cross-slope barriers, terraces, and biological practices – mainly agro-forestry). In general, farmers invest in a set of soil conservation practices rather than in a single one; therefore, we use a combined option, instead of focusing on individual options such as agro-forestry, which is exercised only by 5% of the farmers in the sample.

#### 4. Empirical Strategy

##### 4.1. Estimation of the IR with panel data

Our estimates at the household level are based on the functional form originally proposed by Binswanger, Deininger and Feder (1995), in line with more recent approaches (Barrett et al. 2010) but in the context of panel data:

(4)

$$\frac{Y_{it}}{A_{it}} = \beta_0 + \beta_1 A_{it} + \beta_2 A_{it}^2 + \gamma X_{it} + \alpha_i + u_{it}$$

In (4)  $Y_{it}$  represent the unit value of gross revenue from agriculture production of household  $i$  at time  $t = 1, 2$ .  $A_{it}$  is farmers' land endowment in each year, and  $A_{it}^2$  is the square term to test for non-linearities in the purported relationship.  $X_{it}$  is a matrix of household and farm characteristics such as head's age, education of the head, and gender of the head. To account for time-variant differences in land quality, we add some soil quality variables, as well as variables related to topography. We also control for the impact of non-farm activities by introducing the value of non-farm income. We draw on Reardon (1997) who found that in 23 field studies in Africa, the share (on average) of non-farm income in total income was twice as great in upper income tercile households as in lower tercile households.

We use a linear approximation strategy over log-linear and log-log approaches for two main reasons. First, neither of those were found to provide a better fit with the data, and second, because the presence of outliers is not our major concern since the QR approach inherits its robustness property from median regression, and can produce good and reliable estimates even in the presence of extreme outliers.

We investigate the existence of an average land productivity (ALP)- farm size relationship by using different models. First, we run a pooled OLS regression and compare the results with a fixed-effect regression over the two periods of time (Table 2). We complement our analysis with a panel quantile regression model following Powel (2016) to test the robustness of the results, and to extend them to the entire distribution of farm productivity (Table 3).

The fixed-effects (FE) model allows us to control for time-invariant, unobserved idiosyncratic factors which may affect farmers' ALP, such as know-how, ability and expertise in farming activities, and which are included in the  $\alpha_i$  term, whereas  $u_{it}$  represents the time-varying error term that is uncorrelated with the ALP outcome variable.

The time-invariant component may encompass any form of latent individual effects, or farmers' non-ignorable heterogeneity and even unobserved farmers' personality traits or cognitive and non-cognitive abilities (Almlund et al., 2011; Heckman, 2011; Ravallion 2012). Considering that the unobserved characteristics do not change over time, any change in the dependent variable can be attributed to influences other than these fixed characteristics. Farmers' intrinsic characteristics or risky attitude will determine their willingness to exercise different options, and embark on productivity enhancing activities. The whole of these idiosyncratic characteristics that we define as farmers' abilities does not vary over time, once we control for a full set of predictors. According to this definition, the farmer is able to maximize the level of his ALP over time, given a set of initial endowments in terms of factors of production or household characteristics.

For what concerns the sign of the IR, the introduction of the square terms allows to capture different possible relationships, such as a U shape or inverted-U shape relationship between farm size and productivity. Based on the sign and significance of the coefficients  $\beta_1$  and  $\beta_2$  we can compute a threshold level (maximum for inverted U-shape, and minimum for U-Shape) for farm size that implies a sign-switch in the relationship. If a nonlinear relationship exists, the pattern of thresholds will be even more interesting to analyze in the context of a quantile regressions for panel data (Powell, 2016), as we can analyze non-linearities and sign switches along the whole distribution of productivity.

#### **4.2. Robustness checks and testing for changes in agriculture productivity**

We complement our analysis with some robustness checks to test the validity of the panel quantile on the one hand, and for analyzing how the sign of the ALP impacts changes with agriculture productivity.

We use first the panel structure of our data to corroborate our results by using a quantile regression on the second year of the survey with lagged independent variables to avoid endogeneity. While the empirical application of cross-sectional conditional quantile regression models has been widely developed and tested, corresponding methods for panel data have been developed only recently. The large majority of the panel quantile studies has thus dealt with macroeconomic applications, longer time spans and fewer numbers of observations (Koenker, 2004; Lamarche, 2010; Kato and Galvao, 2010; Kato, Galvao and Montes-Rojas, 2012). The main problems associated with fixed effect quantile regression models is that, as it is the case with nonlinear panel data models, the method of differencing out the fixed effects used for the conditional linear-mean model does not carry over to the conditional quantiles. Panel quantile methods also suffer from the standard incidental parameter problem (Neyman and Scott, 1948). This results in an inconsistent estimate of the coefficient of interest (e.g., the land area) in an individual fixed effects quantile regression model when the number of individuals grows large (tends to infinity) and the number of time periods is fixed (small) (Kato and Galvao, 2011, Flores et al., 2014).

Given this background, we use the same functional form and estimate the relationship of interest with a quantile regression model for 2009, the second year of the survey (Table 4). Agriculture ALP is thus regressed on its lagged value, on farm size and its square, on lagged non-farm income and on a set of exogenous controls. As farm size (and not land operated) and past off-farm incomes are the results of past decisions of farmers, no endogeneity arises. As an additional test, we also run the cross-section quantile regression on ALP changes between 2004 and 2009 (Table 5).

### **5. Estimation Results**

Table 2 presents results for a simple OLS pooled regression with year dummy (Column 1), and for fixed-effect regressions (column 2) with gross agriculture income per hectare (our ALP) as the dependent variable, using the two rounds of data, 2004 and 2009.

By measuring ALP as gross revenue per ha, we interpret the functions that we estimate as a normalized revenue function, in the spirit of McFadden, i.e. as a gauge function that characterizes farm production along a, possibly bounded, optimization path. This implies that the revenue per ha observed already incorporates the same period farm choices relative to input levels and is a function only of exogenous variables, including prices and farm size.

We find a nonlinear significant U-shape relationship between farm size and productivity. This suggests the presence non-constant returns to scale in agriculture, first decreasing up to the threshold (3.44 ha) and then increasing.

However, the analysis of panel quantile regression suggests a different picture. Even though not significant, the relationship between ALP and farm size for low performers (low agriculture productivity) is an inverted U-shaped form, and after median productivity is reached, switches to a U-shaped form. This in turn suggests that returns to scale switch across the distribution of both productivity and farm size.

Given the short time span, we choose to corroborate our results using the panel structure of our data on the second year cross-section sample. Tables 3 and 4 present results of quantile regression analysis on the second round of the survey where the dependent variables are, respectively, ALP (Table 3) and ALP growth (table 4).

Our major finding suggests that land productivity and farm size are systematically related, but the sign and the intensity of this relationship depend on one or more latent variables, directly related to total factor productivity. For farmers in the lowest deciles of the ALP distribution, where this variable has lower values, revenue per ha tends first to increase with farm size, and once reached a maximum (3 ha in average), to decrease. These farms, therefore, may be using land expansion (or expansive margins), perhaps as a safety-first strategy, to increase their welfare by ensuring subsistence. They would benefit first from an increase in farm size, but their deficit of management capacity would then become a priority, so that they would lose from further expanding their land size without achieving independent and substantial gains in ALP or, more effectively, in total factor productivity (TFP). For farmers in the highest deciles of the land productivity distribution, on the other hand, where management ability and related TFP are more developed, the situation would be reversed. These farmers would benefit from expanding their farm size beyond the minimum threshold, while they would be constrained and show negative results by expanding within the threshold itself.

We also find that productivity growth is driven by similar factors, with a pattern of behavior that reinforces our conclusions beyond the cross-section results. Not only the farms that are in the lowest productivity deciles, in fact, but also the farms that are less dynamic in terms of productivity growth, display a lack of management ability that translates into diseconomies of scale. Farms in the highest deciles of average productivity growth, on the other hand, display the opposite pattern of response to farm size increase, with increases in productivity first negatively and then positively correlated with the scale of operations.

Figure 2 summarizes the results of our analysis and analyzes the farm-size productivity relationship among less and more performant farmers. Put it in a general perspective, beyond the value of the threshold, it appears that the relationship between farm size and productivity follows two distinct patterns for less and more productive farmers, with latent variables pushing farm size to converge to a certain threshold. Given a level of productivity, the relationship is direct for very small farms, and for large ones (below and above the threshold respectively), and it is inversely related to agriculture productivity for a whole range of median (or middle farms). This pattern suggests that the ALP farm size relationship can be characterized as DID (Direct-Inverse-Direct). This result goes beyond the overall farm size. Figure 3 shows the distribution of land among low and high performance (below and above the median land productivity respectively). With a maximum farm size in the sample of about 6 ha, Figure 3 points towards the existence of a variety of farms size within the two classes of productivity.

One interesting result is the significance and sign of the variable non-farm income, a proxy for non-farm employment and agriculture diversification. The coefficient is positive and significant in the pooled OLS regression, and the panel fixed effect (Table 2). This points towards the fact that the income generated by off-farm activities is likely to be used as a safety net activity to increase rural income and to reduce income inequalities. Farmers that invest in off-farm activities or work off-farm can use the income generated to invest in more productive on-farm activities. However, when looking at the entire distribution (Table 4), the coefficient is positive only for the less productive farms but not significant for the more productive, though still positive. This suggests, as Figure 2 shows, that agriculture diversification and rural non-farm activities represent a valuable option for the “middle farm” farmers lying on the negative side of the farm-size productivity relationship. This is counterintuitive compared to the conventional wisdom of the negative relationship between non-farm and total income or landholdings. Our interpretation adds to the existing literature that revisits this hypothesis (Reardon et al., 2000; Reardon 1997; Lanjouw and Lanjouw, 1995; Reardon and Taylor, 1996; and Ellis, 1998).

Figure 2: IR and Threshold value of Land

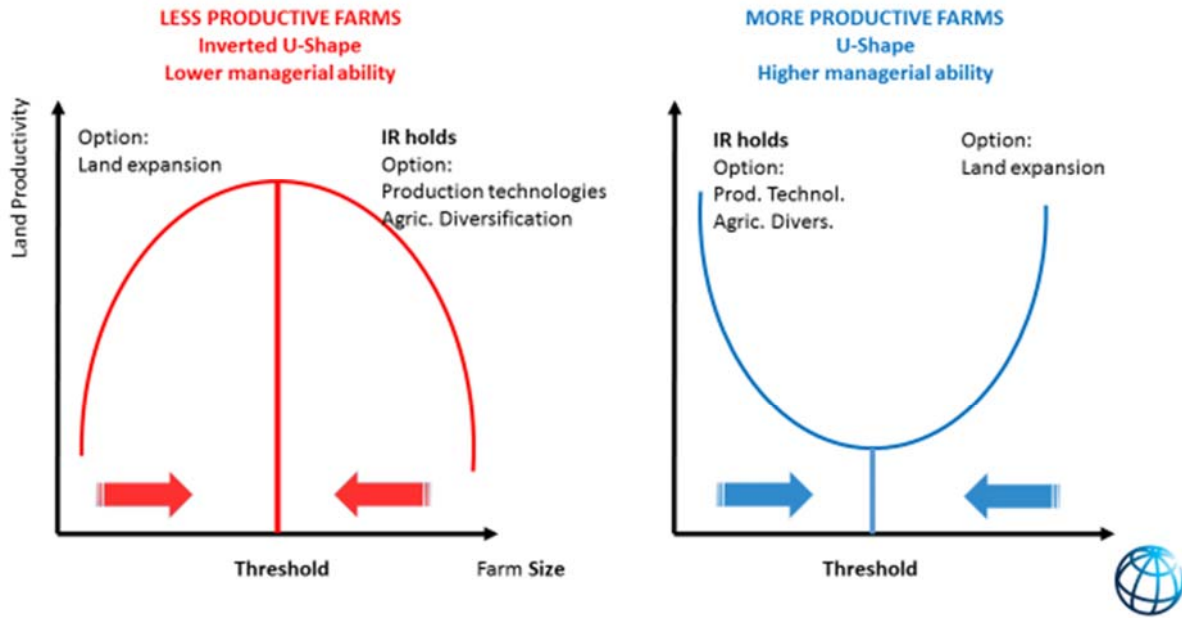
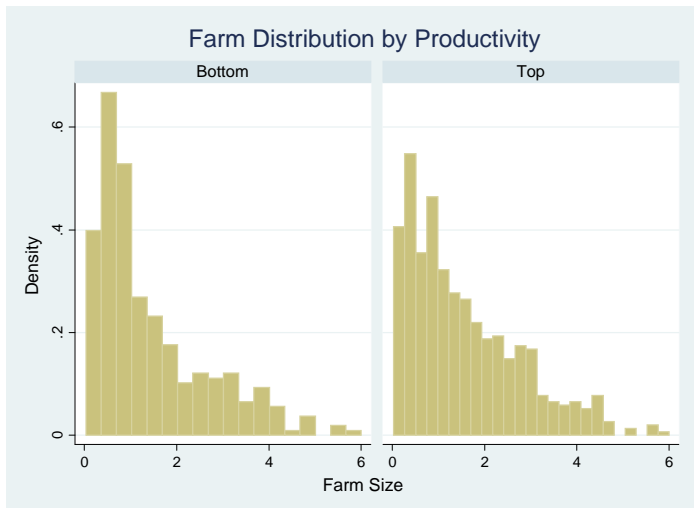


Figure 3: Distribution of farm size by bottom and top performance.



Source: Authors' calculation based on ERHS

These results, which seem reasonable and are confirmed along a vast set of countries, suggest that the ubiquitous IR relationship reported by the literature may have been an artifact of the presumption that

the relationships observed were linear or, at least, a linear approximation to a monotonic relationship. It appears not only that the relationship is pronouncedly non-monotonic, but also that it is drastically different across farms, depending on their position in the distribution of productivity. The policy implications of these findings are also significant. For farms in the lowest TFP ranks, this result in fact implies that some land expansion may be beneficial, but lack of management ability soon creates diseconomies of scale so that better management and technology are needed more than land increases. Moreover, this implication holds also for farms that are less dynamic in terms of their productivity growth. On the contrary, for farms that already have a patrimony of technological know-how and good management, and thus are in the highest deciles of the distribution of productivity or productivity increase, land expansion, and the associated economies of scale, may be the only policy that can make their productivity grow.

## Conclusions

The inverse relationship (IR) between average land productivity (ALP) and land size has been the object of a voluminous literature, raising both objections and explanations. In this study, after a brief review of some of the main arguments, we have presented evidence from two rounds of data from Ethiopia that offers a new and more comprehensive explanation for the relationship. The survey data used are from detailed household interviews and contain accurate georeferenced information on farmers' location, distance from the markets, distance from the main road, and land quality. In order to estimate the ALP–farm size relationship and to test the IR hypothesis, we have used a specification entirely relying on exogenous variables and estimation procedures according to the quantile regression model.

Our results suggest that farm size displays a robust impact on ALP, but this impact is nonlinear and switches shape across farms according with their performance. For low performers the impact is first positive and then negative, while the opposite is true for high performers. As already noted, although in a different context by Evenson and Mwangi (2001), this may be due to individual management factors, so that in the two areas of the ALP distribution, different complementary and substitute relations exist between land sizes and unobserved human capital variables, such as farmers' abilities, skills and experience. In turn, these results are consistent with a revised version of Savastano and Scandizzo's (2009) option model, where management quality is supposed to be positively correlated with farm size.

Interpreted across the range of farm sizes, our results also point to the possible existence of an S-shaped relationship between ALP and acreage at least for a range of performances, with ALP first increasing then decreasing, then increasing again. This DID relationship implies a turning point for the lower quantiles of the ALP residual distribution at which a positive (DR) relationship becomes negative (IR) and one for the upper quantiles where IR becomes positive again (DR). Both turning points are for small to medium farm sizes, but the ones of the lower quantiles tend to be smaller than those for the upper quantiles. Thus, while there is some significant negative relationship between productivity and operating size for low performers over a relevant range of farm sizes, higher performers tend to display IR only over a range from small to medium farm sizes. The interpretation of these results is thus that low productivity farms may exploit to some extent economies of scale for a limited range of farm size increases, but need to become more productive by exercising the right options to invest and not to incur into diseconomies once a maximum farm size is reached. Farmers in the higher quantiles of the productivity distribution, on the other hand, can exploit the economies of scale related to farm size only after they cross a minimum threshold of farm size. Furthermore, a similar relationship holds for productivity growth, in the sense that more productive and more dynamic farms are those that can exploit more fully farm expansion as a source of scale economies and that the IR relationship holds only if they are below a minimum economic level of farm size.

One could question whether the DID relationship could hold in a new external context characterized by new trends such as emergent or mega farm operators, and large-scale land acquisitions (Collier, 2009; Collier and Dercon, 2014; Sitko and Jayne, 2014). We argue that even though our nationally

representative surveys focus on small and medium size farms, the results of our land productivity distribution analysis could be tested also on a larger sample of farms, and identify different levels of middle farms based on the countries' own agro-ecological and farming characteristics. With some exceptions (Ali et al., 2016; Muyanga and Jayne, 2016), the current format and survey design do not allow to make any type of quantitative productivity analysis.

These results have clear policy implications, since they indicate that land reform and redistribution policies may be very effective for less efficient producers only if they are below a critical farm size, after which management and technology are better instruments to improve their lot. Vice versa, for producers that are already at a reasonable level of efficiency and dynamism, the opposite is true: land policies, be they in the form of redistribution or more secure tenure, would be more effective than extension and technological innovation.

In sum, our results confirm that IR may be a ubiquitous relationship, as found in much of the literature, but indicate that its form, shape and importance may significantly differ across the spectrum of farm productivity performance. At the low end of the ALP distribution, IR appears to prevail, once a minimum threshold of farm size is reached, while at the higher hand, IR only appears to be mainly a characteristic of farmers with operating sizes not exceeding medium size thresholds. The literature on transaction costs and the role of the firm suggests that these differences will require a deeper analysis of some of the critical factors determining the performance of the farm as a "productivity agent" and of the role played by management and capabilities in shaping farmers' choices.



## References

- Ali, D. A and K. Deininger. 2014. Is there a farm-size productivity relationship in African agriculture? Evidence from Rwanda. *World Bank Policy Research Paper 6770*.
- Ali, Daniel Ayalew & Deininger, Klaus W. & Harris, Charles Anthony Philip, 2016. "Large farm establishment, smallholder productivity, labor market participation, and resilience : evidence from Ethiopia, "Policy Research Working Paper Series 7576, The World Bank.
- Almlund, Mathilde & Duckworth, Angela Lee & Heckman, James & Kautz, Tim, 2011. "Personality Psychology and Economics,"*Handbook of the Economics of Education*, Elsevier.
- Assunção, J. J. and L. H. B. Braido. 2007. Testing Household-Specific Explanations for the Inverse Productivity Relationship. *American Journal of Agricultural Economics*, 89 (4): 980-90.
- Assunção, J., Ghatak, M., 2003. On the Inverse Relationship between Farm Size and Productivity. *Economics Letters*, Volume 80, No. 2, pp.189-194, August 2003
- Bardhan, P. 1973. Size, productivity and returns to scale: an analysis of farm-level data in Indian agriculture. *Journal of Political Economy*, 81 (6): 1370-86.
- Barrett, C. B. 1996. On Price Risk and the Inverse Farm Size-Productivity Relationship. *Journal of Development Economics*, 51 (2): 193-215.
- Barrett, C. B., M. F. Bellemare and J. Y. Hou. 2010. Reconsidering Conventional Explanations of the Inverse Productivity-Size Relationship. *World Development*, 38 (1): 88-97.
- Benjamin, D. 1995. Can Unobserved Land Quality Explain the Inverse Productivity Relationship? *Journal of Development Economics*, 46 (1): 51-84.
- Benjamin, D., Brandt, L., 2002. Property Rights, Labor Markets, and Efficiency in a Transition Economy: The Case Of Rural China. *Canadian Journal of Economics*, 35(4), 689–716.
- Berry, R.A., Cline, W.R., 1979. *Agrarian Structure and Productivity in Developing Countries*. Johns Hopkins Univ. Press, Baltimore, MD.
- Bevis Leah EM and Christopher B Barrett, 2016, "Close to the Edge: Do Behavioral Explanations Account for the Inverse Productivity Relationship?"; November 2016 working paper.
- Bhalla, S. S. and P. Roy. 1988. *Misspecification in Farm Productivity Analysis: The Role of Land Quality*. *Oxford Economic Papers*, 40 55-73.
- Billger, Sherrilyn M. & Goel, Rajeev K., 2009. "Do existing corruption levels matter in controlling corruption?: Cross-country quantile regression estimates," *Journal of Development Economics*, Elsevier, vol. 90(2), pages 299-305, November.
- Binswanger, H.P., K. Deininger, and G. Feder. 1995. Power distortions revolt and reform in agricultural land relations. In: Behrman, J. and Srinivasan, T.N. (eds). *Handbook of development economics*, Volume III. Amsterdam, the Netherlands: Elsevier Science B.V.
- Cade, B. and B. Noon, 2003. "A Gentle Introduction to Quantile Regression for Ecologists." *Frontiers in Ecology and the Environment*, 1, 412-420.
- Carletto G., Savastano S., and A. Zezza, 2013, "Fact or Artefact: The Impact of Measurement Errors on the Farm-Productivity Relationship", *Journal of Development Economics*, Volume 103, July 2013, Pages 254–261, and World Bank Policy Research Working Paper 5908, December 2011
- Carletto, C.; Gourlay, S., and P. Winters, 2015. "Editor's choice From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis", *Journal of African Economies*, Centre for the Study of African Economies (CSAE), vol. 24(5), pages 593-628.

- Carlos A. Flores, C. A., Flores-Lagunes, A., and D. Kapetanakis, 2014. "Lessons From Quantile Panel Estimation of the Environmental Kuznets Curve," *Econometric Reviews*, Taylor & Francis Journals, vol. 33(8), pages 815-853, November.
- Carter, M. R. and K. Wiebe. 1990. Access to Capital and its Impact on Agrarian Structure and Productivity in Kenya. *American Journal of Agricultural Economics*, 72: 1146-1150.
- Carter, M., 1984. Identification of the Inverse Relationship between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production. *Oxford Economic Papers*, New Series, 36, 1, pp. 131-145.
- Chamaillé-Jammes, S. and Blumstein, D. 2012. "A case for quantile regression in behavioral ecology: getting more out of flight initiation distance data." *Blumstein Behavioral Ecology and Sociobiology*. 66 – 6: 985-992
- Chamberlain, G. , 1994, "Quantile Regression, Censoring, and the Structure of Wages," in C. A. Sims (ed.), *Advances in Econometrics*, Sixth World Congress, Volume 1. Cambridge University Press. Cambridge
- Chayanov, A.V. 1926. *The Theory of Peasant Economy*, In D. Thorner, B. Kerblay, and R.E.F. Smith, eds. Irwin: Homewood.
- Collier, P. and S. Dercon. 2014. African Agriculture in 50 Years: Smallholders in a Rapidly Changing World? *World Development*, 63: 92-101.
- Collier, Paul. 2009. Africa's Organic Peasantry: Beyond Romanticism. *Harvard International Review* Summer: 62-65.
- Cornia, G. A., 1985. Farm Size, Land Yields and the Agricultural Production Function: An Analysis for Fifteen Developing Countries. *World Development*, 13 (4): 513–34.
- De Groote, H., Traoré O., 2005. The Cost of Accuracy in Crop Area Estimation. *Agricultural Systems* 84:21-38
- Deininger, K., and G. Carletto, Savastano S., Muwongue, J., (2012), "Can diaries help in improving agricultural production statistics? Evidence from Uganda" *Journal of Development Economics*, Volume 98, Issue 1, Pages 42-50 (May 2012)
- Desiere S. and D. Jolliffe, 2017 (forthcoming)," Land productivity and plot size: Is measurement error driving the inverse relationship?", Manuscript submitted for publication in Policy Research Working Paper. Washington DC: World Bank Group
- Ellis, F. (1998). *The Determinants of Rural Livelihood Diversification in Developing Countries*, Agricultural Economics Society Annual Conference, University of Reading
- Eswaran, M., Kotwal, A., 1985. A Theory of Contractual Structure in Agriculture, *American Economic Review* 75: 352-67.
- Eswaran, M., Kotwal, A., 1986. Access to Capital and Agrarian Production Organization, *Economic Journal* 96: 482-98.
- Evenson, R.E. and Mwabu, G., 2001, "The Effects of Agricultural Extension on Farm Yields in Kenya", *African Economic Review* Volume 13, Issue 1.
- Feder, G. 1985. The Relation between Farm Size and Farm Productivity: The Role of Family Labor, Supervision and Credit Constraints. *Journal of Development Economics*, 18 (2-3): 297-313.
- Fitzenberger, B. (1997). A Guide to Censored Quantile Regressions. In: *Handbook of Statistics*, Volume 15: Robust Inference (Eds. G.S. Maddala & C.R. Rao), 405-437. Amsterdam: North-Holland.
- Goldstein, M., Udry C., 1999. *Agricultural Innovation and Risk Management in Ghana*. Unpublished, Final report to IFPRI.
- Hazell, P. e Scandizzo, P.L., 1974 "Market Intervention Policies when Production is Risky", *American Journal of Agricultural Economics*, 57 (4), pp. 641-649, 1975 Johns Hopkins University Press.
- Heckman, J., and Kautz, T.D., 2011, "Hard Evidence on Soft Skills" paper presented at the World Bank, December 15, 2011, Washington DC.
- Horowitz, J.L. & G.R. Neumann, 1987. "Semiparametric estimation of employment duration models", *Econometric Reviews* 6, 5—40
- Imai, S., Katayama, H. and K. Kala, 2013. "A quantile-based test of protection for sale model", *Journal of International Economics*, Elsevier, vol. 91(1), pages 40-52.

- Johar, M. And H. Katayama, 2012. "Quantile regression analysis of body mass and wages". *Health Economics*, Volume 21, Issue 5, May , 597–611
- Kato, K., and A. Galvao (2011): "Smoothed Quantile Regression for Panel Data," Working Paper. [811,815] KATO, Kato, K., Galvao K., A. and G. Montes-Rojas (2012): "Asymptotics for Panel Quantile Regression Models With Individual Effects," *Journal of Econometrics*, 170, 76–91. [811,815, 817]
- Kilic, T., Gourlay, S. and D. Lobell, 2017, Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for Inverse Scale - Productivity Relationship in Uganda, unpublished working paper
- Koenker, R. & Y. Biliias, 2001. "Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments", *Empirical Economics* 26, 199—220.
- Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89. [811,815,820]
- Koenker, R., and J. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50. [810, 815-817]
- Kutcher, G. P., and P. L. Scandizzo, 1981. *The Agricultural Economy of Northeast Brazil*. Baltimore, John Hopkins University Press for the World Bank
- Lamarche, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408. [811,820]
- Lamb, R. L. 2003. Inverse Productivity: Land Quality, Labor Markets, and Measurement Error. *Journal of Development Economics*, 71 (1): 71-95.
- Lanjouw J.O. and Lanjouw P. 1995. Rural Nonfarm Employment: Survey. Yale University and the World Bank. World Development Report. Washington D.C.
- Lemke K. 2010 "Modeling Healthcare Expenditures Using PROC QUANTREG", NESUG 2010, 1-7
- Levin, J.,2001. "For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement", *Empirical Economics* 26, 221—246
- MacDonald, James M.; Korb, Penni and Robert A. Hoppe, 2013, Farm Size and the Organization of U.S. Crop Farming, A report summary from the Economic Research Service, August
- Muyanga, M., and T.S. Jayne; 2016, Is small still beautiful? The farm size-productivity relationship revisited in Kenya, Paper presented at the World Bank Land and Poverty Conference, March 15, 2016, Washington, DC.
- Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.
- Ravallion, M., 2012, Poor, or Just Feeling Poor? On Using Subjective Data in Measuring Poverty, Policy Research Working Paper 5968, The World Bank Washington DC.
- Reardon, T. (1997). Using Evidence of Household Income Diversification to Inform Study of the Rural Nonfarm Labor Market in Africa, *World Development*, 25(5)
- Reardon, T., J.E. Taylor, K. Stamoulis, P. Lanjouw, and A. Balisacan, 2000. "Effects of Nonfarm Employment on Rural Income Inequality in Developing Countries: An Investment Perspective," *Journal of Agricultural Economics*, Volume 51, Number 2 - May 2000 - Pages 266-288
- Reardon. T. and Taylor, J. E. (1996). Agroclimatic Shock, Income Inequality, and Poverty: Evidence from Burkina F&o, *World Development*, 24(4), 901-914.
- Savastano S., and P.L. Scandizzo, 2009, "Optimal Farm Size in an Uncertain Land Market: the Case of Kyrgyz Republic", *Agricultural Economics*, 2009, vol. 40, issue s1, pages 745-758
- Scandizzo, P.L., Savastano, S., 2017. Forthcoming, "Revisiting the farm size productivity relationship: New evidence from sub-Saharan African countries", In *Agriculture and Rural Development in a Globalizing World*, edited by Prabhu Pingali and Gershon Feder, Chapter 3 Earthscan Food and Agriculture Series. London: Routledge
- Sen, K. Amartya, 1975. *Employment, technology and development*. Oxford, UK: Clarendon Press.

Sitko, Nicholas J. & Jayne, T.S., 2014. "Structural transformation or elite land capture? The growth of "emergent" farmers in Zambia," *Food Policy*, Elsevier, vol. 48(C), pages 194-202.

Srinivasan, T. N. 1972. Farm size and productivity: Implications of choice under uncertainty. *Sankhya - The Indian Journal of Statistics*, 34 (2): 409-20.

Walker, T.S., Ryan, J.G., 1990. *Village and Household Economies in India's Semi-arid Tropics*. , Johns Hopkins Univ. Press, Baltimore.

**Table 1: Household characteristics**

	2004	2009	Total
<b>Farmers' characteristics</b>			
Head's age	50.49***	52.53***	51.51
Dummy for female HH head	0.26*	0.29*	0.27
Head's education (years)	1.4***	1.8***	1.6
Max years of education of HH members	3.8**	5.7**	4.7
HH size	5.93	5.91	5.92
<b>Income, poverty and Inequality</b>			
Income per capita (Const. 2009 US\$)	105.3	152	129
Consumption per capita (Const. 2009 US\$)	93.5	173.5	134
Gini coefficient of land	0.43	0.44	0.44
Gini coefficient of income	0.48	0.46	0.48
Gini agriculture income pc	0.51	0.52	0.53
Gross agriculture income/ha	446.81***	549.92***	497.8
<i>Share of income from:</i>			
agriculture (crop+liv)	0.82***	0.73***	0.78
from crop	0.61***	0.55***	0.58
from livestock	0.21***	0.18***	0.2
from non-farm income	0.18**	0.27**	0.22
<b>Agriculture and Farming conditions</b>			
Farm size	1.40**	1.52**	1.46
Midpoint farm size	1.6**	2.15**	2.0
Number of plots	4.7	5.1	4.9
Value of inputs /ha	501.82***	991.93***	746.43
Nb. of plots with steep slope	0.16	0.14	0.15
Dummy too little rain on HH fields during rainy season	0.27***	0.32***	0.29
Share of land with good soil quality	53*	56*	54
<b>Agriculture options in round 2</b>			
Dummy Option land expansion		52***	76
Amount of land increased (ha)		0.75	0.75
Dummy option soil conservation		18	18

Source: Authors' computation based on ERHS 2004-2009

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

**Table 2: Panel Analysis ERHS 2004-2009: Dependent Variable: Gross Agriculture Income per ha (US\$/ha)**

Gross Ag. Income/ha in 2009 (US\$/ha)	Pooled OLS	aster	HH Fixed Effect	aster
Farm size	-260.89	***	-271.30	***
Farm size sq.	37.90	***	38.95	***
Age of HH head	0.99		0.81	
Dummy for female HH head	-56.32	**	-55.67	**
Years of education of HH members	39.97	***	37.06	***
Nb. of plots with steep slope	-51.53	***	-44.21	***
Share of land with good soil quality	88.19	***	76.50	***
Producer price Index for Cereals	32.98	***	0.93	***
Non-farm income	0.05	***	0.05	***
Dummy year 2009	-2,650.48	***		
Constant	-2,980.06	***	460.32	***
<b>Threshold Land</b>	<b>3.44</b>		<b>3.48</b>	
Observations	2,126		2,126	
R-squared	0.129			
Number of HHs			1,063	

Source: Authors' computation based on ERHS 2004-2009

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table 3: Panel Quantile Analysis ERHS 2004-2009: Dependent Variable: Gross Agriculture Income per ha (US\$/ha)

Gross Ag. Income/ha in 2009 (US\$/ha)	Low Performers								Median		High Performers							
	Q 0.1	aster	Q 0.2	aster	Q 0.3	aster	Q 0.4	aster	Q 0.5	aster	Q 0.6	aster	Q 0.7	aster	Q 0.8	aster	Q 0.9	aster
<b>Farm size</b>	3.57		0.18		3.57		-6.38		-12.05		-45.24		<b>-171.20</b>	<b>***</b>	<b>-269.49</b>	<b>***</b>	<b>-570.86</b>	<b>***</b>
<b>Farm size sq.</b>	-2.21		-0.07		-1.10		-0.86		0.88		4.09		<b>20.34</b>	<b>**</b>	<b>33.86</b>	<b>*</b>	<b>84.74</b>	<b>**</b>
Age of HH head	1.30		0.01		-0.06		-0.04		-0.03		-0.26		1.11		0.31		2.27	
Dummy for female HH head	2.29		-1.33		22.56		21.15		-4.50		-23.15		-60.90		-51.87		-81.42	
Years of education of HH members	3.12		0.66		22.66	<b>**</b>	18.73	<b>**</b>	8.24		18.42	<b>**</b>	38.89	<b>***</b>	50.64	<b>**</b>	72.38	<b>*</b>
Nb. of plots with steep slope	-0.32		8.20		15.10		-2.11		-2.97		-16.65	<b>*</b>	-34.62		-48.80		-67.99	<b>***</b>
Share of land with good soil quality	5.61		1.49		27.40		27.62		20.43		53.86		111.16	<b>**</b>	112.80		142.06	
Producer price Index for Cereals	1.10		0.01		0.19		0.33		0.23		0.58		1.42		1.84		2.11	
Non-farm income	-2.97		-0.01		0.02		0.03		0.01		0.07		0.04		0.05		0.06	
Constant																		
<b>Threshold Land</b>													<b>4.21</b>		<b>3.98</b>		<b>3.37</b>	
Observations	2,126		2,126		2,126		2,126		2,126		2,126		2,126		2,126		2,126	
R-squared																		
Number of HHs	1,063		1,063		1,063		1,063		1,063		1,063		1,063		1,063		1,063	

Source: Authors' computation based on ERHS 2004-2009

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 4: Cross Section Quantile Regression in 2009: Dependent Variable Gross Agriculture Income/ha (US\$/ha)**

Gross Agriculture Income/ha in 2009	Low Performers								Median		High Performers							
	Q 0.1	aster	Q 0.2	aster	Q 0.3	aster	Q 0.4	aster	Q 0.5	aster	Q 0.6	aster	Q 0.7	aster	Q 0.8	aster	Q 0.9	aster
Gross Ag. Income/ha lagged	0.04		0.08	***	0.08	***	0.08	*	0.13	**	0.22	**	0.24	***	0.28	***	0.24	**
Land (ha)	<b>58.04</b>	<b>**</b>	<b>58.72</b>	<b>**</b>	<b>47.97</b>	<b>*</b>	<b>15.17</b>		<b>-30.09</b>		<b>-76.20</b>		<b>-165.12</b>	<b>***</b>	<b>-268.70</b>	<b>***</b>	<b>-566.20</b>	<b>***</b>
Land sq	<b>-9.08</b>	<b>**</b>	-8.13		<b>-9.11</b>	<b>*</b>	<b>-4.68</b>		<b>0.92</b>		<b>6.91</b>		<b>20.96</b>	<b>*</b>	<b>36.12</b>	<b>**</b>	<b>88.16</b>	<b>***</b>
Head's Age	-0.92		-1.59	***	-1.38		-1.69	*	-1.47		-1.07		-1.20		-1.87		1.27	
Dummy Female Head	-46.73	***	-62.69	***	-81.17	***	-87.34	***	-62.32	*	-54.82	*	-69.03	**	-94.14	*	-125.79	*
Head's Education	10.25	**	17.06	***	23.40	***	26.74	***	30.81	***	36.36	***	34.90	***	48.21	***	60.88	***
Nb. of Plots with Steep Slope	-6.23		-18.88	***	-31.92	***	-40.70	***	-40.16	**	-39.20	***	-46.03	***	-52.78	***	-87.60	***
Share of land with good soil quality	26.38		23.71		31.10		31.51		32.78		65.74	*	60.29		69.56		24.06	
Producer Price Index for Cereals	30.45	***	25.77	***	24.59	***	25.85	***	29.62	***	32.61	***	36.72	***	35.27	***	50.27	***
Non-Farm Income Lagged	0.03	**	0.03	**	0.03	*	0.02		0.03		0.03		0.03		0.04		0.03	
Constant	-5,683.58	***	-4,717.56	***	-4,427.78	***	-4,548.81	***	-5,181.53	***	-5,689.44	***	-6,276.02	***	-5,761.33	***	-8,179.54	***
<b>Threshold Land</b>	<b>3.20</b>				<b>2.63</b>								<b>3.94</b>		<b>3.72</b>		<b>3.21</b>	
Observations	1,063		1,063		1,063		1,063		1,063		1,063		1,063		1,063		1,063	

Source: Authors' computation based on ERHS 2004-2009

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1



Table 5: Cross Section Quantile Regression in 2009: Dependent Variable Difference Gross Agriculture Income/ha (US\$/ha) (2009-2004)

Difference Gross Agriculture Income/ha (2009-2004)	Low Performers								Median		High Performers							
	q10	q10	q20	q20	q30	q30	q40	q40	q50	q50	q60	q60	q70	q70	q80	q80	q90	q90
Land (ha)	152.73	*	123.17	***	90.85	***	62.19	**	22.00		-18.35		-104.73	*	-233.09	***	-520.58	***
Land sq	-21.96		-17.83	**	-15.19	**	-10.73	*	-4.93		1.62		14.59		36.54	***	81.40	***
Head's Age	-2.22		-1.88		-2.28	***	-1.97	***	-1.78	**	-2.73	**	-1.48		-0.12		-0.05	
Dummy Female Head	-56.68		-72.53		-44.65		-58.48	**	-47.63		-36.89		-61.26		-91.40		-120.65	
Head's Education	0.16		-7.03		-3.22		0.76		7.31		10.15		18.36	**	13.19		39.81	**
Nb. of Plots with Steep Slope	12.75		22.67		4.86		2.46		-0.72		1.42		-11.27		-28.19		-59.08	
Share of land with good soil quality	-153.41		-87.92	*	-53.20	*	-20.52		11.21		-1.04		-19.27		-40.82		-20.88	
Producer Price Index for Cereals	53.63	***	40.52	***	41.87	***	35.58	***	32.33	***	25.82	***	28.67	***	36.58	***	49.05	***
Non-Farm Income Lagged	-0.05		0.03		0.02		0.03		0.04	**	0.02		0.05		0.06		0.08	*
Constant	10,454.05	***	7,769.19	***	7,903.02	***	6,649.06	***	5,978.87	***	4,568.63	***	4,987.71	***	6,258.85	***	8,122.64	***
<b>Threshold Land</b>			<b>3.45</b>		<b>2.99</b>		<b>2.90</b>								<b>3.19</b>		<b>3.20</b>	
Observations	1,063		1,063		1,063		1,063		1,063		1,063		1,063		1,063		1,063	

Source: Authors' computation based on ERHS 2004-2009

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$