Decomposing the inverse land size-yield relationship

Simone Pieralli

Research Associate, Department of Agricultural Economics, Humboldt-Universität zu Berlin

and

Ph.D. Candidate at Department of Agricultural and Resource Economics at University of Maryland College Park

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting, Minneapolis, MN, July 27-29, 2014.

Decomposing the inverse land size-yield relationship

Simone Pieralli
Department of Agricultural Economics,
Humboldt University in Berlin,
Unter den Linden 6, D-10115 Berlin, Germany
and
Department of Agricultural and Resource Economics,
University of Maryland College Park,
2200 Symons Hall, College Park, MD 20742, USA.

Draft paper: comments are welcome.
For citation please contact the author.

Abstract

Faster agricultural development requires understanding whether the inverse land size-yield relationship exists or not. To verify the presence of this relationship, this study decomposes a yield index into separate components attributable to (1) efficiency, (2) soil quality, (3) land size, (4) variable inputs, (5) capital inputs, and (6) outputs. Non-parametric productivity accounting methods are used to decompose the inverse land size-yield relationship in a multi-output representation of the technology without specific assumptions on returns to scale. A strongly significant inverse (positively convex) land size-yield relationship is present in the Kenyan data, but vanishes in favor of a linear inverse relationship when accounting for the effect of outputs' diversification.
KEYWORDS: inverse land size-yield relationship, productivity decomposition, efficiency, yield, Kenya.
JEL Codes: D20, C14, C43.

1 Introduction

The inverse land size-yield relationship is a long-standing empirical phenomenon since Chayanov observed it for the first time in 1926. Many explanations for the phenomenon have been considered, ranging from incomplete and imperfect markets, to erroneous measurement of land size, or of effective soil quality. Recent contributions have focused on verifying the physical measurement of land size and soil quality. On one hand, Carletto et al. (2011) have observed a strong negative relationship between land size and yield, when using precise satellite data on land size. On the other hand, Barrett et al. (2010) have noticed that soil-quality differences, even when measured through chemical soil-analysis data, only account for one third of the inverse land size-yield relationship.

Whether there exists a systematic relationship between land size and yield, potentially has important policy implications for agricultural development. If an inverse relationship truly exists, small farmers can be promoted not only on egalitarian grounds, but also because they are more productive. On the other hand, if the perceived relationship is an artifact of the methods used, such policies may have less economic support, and may be tailored to intensifying agriculture.

This study examines the inverse land size-yield relationship using more general methods than those used in previous contributions. In particular, the present study develops a productivity accounting method that does not rely on specific assumptions on returns to scale, and accommodates variation of multiple outputs. The method consists in decomposing an index of yield, as a partial productivity measure, into six components: (1) efficiency differences, (2) soil-quality component, (3) land-size component, (4) variable-inputs component, (5) capital-inputs component, and (6) outputs component.

The empirical application of this study is to a sample of Kenyan household farmers. The results evidence a strong (positively) convex inverse relationship between different measures of average land product (including yield) and land size, while excluding the effect of outputs' diversification. By accounting appropriately for the effect in yield terms of outputs' diversification, the convex inverse land size-yield relationship disappears in favor of a simple linear relationship.

The next section discusses the methodology and the following section presents the data. The fourth section presents the empirical results, and the last section concludes.

2 Methodology

The present methodology is grounded on the key observation that yield is a partial productivity measure. Therefore, once yield is converted into index form by comparing it to some base-level yield, simple methods rooted in the theory of index numbers and productivity accounting (Kumar and Russell, 2002) can be used to analyze it exactly as any other partial productivity measure.

Empirical analyses on the inverse land size-yield relationship frequently do not recognize that yield is a partial productivity measure. Lacking that recognition, they cannot accommodate appropriately effects arising from the contemporaneous variation of multiple outputs and multiple inputs. The present contribution adapts methods developed by Färe et al. (1994) and Kumar and Russell (2002) to accommodate variable returns to scale, multiple outputs[1], and a number of production aggregates higher than considered in previous studies.

2.1 Multi-output technology

Let $\mathbf{y} \in \mathbb{R}^S_+$ denote a vector of outputs and $\mathbf{x} \in \mathbb{R}^U_+$ denote variable inputs, and let $l \in \mathbb{R}_+$ denote land area, $\mathbf{c} \in \mathbb{R}^C_+$ denote soil characteristics, and $\mathbf{b} \in \mathbb{R}^B_+$ denote capital inputs. The multi-output technology set T is defined:

$$T = \left\{ (\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) \in \mathbb{R}^{U+1+C+B+S}_+ : (\mathbf{x}, l, \mathbf{c}, \mathbf{b}) \, \text{can be used by households to produce} \, \mathbf{y} \right\}.$$

T satisfies:

A.1: Convexity: If $(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \mathbf{y}_1) \in T$ and $(\mathbf{x}_2, l_2, \mathbf{c}_2, \mathbf{b}_2, \mathbf{y}_2) \in T$, then $\forall \alpha \in [0,1] : \alpha(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \mathbf{y}_1) + (1-\alpha)(\mathbf{x}_2, l_2, \mathbf{c}_2, \mathbf{b}_2, \mathbf{y}_2) \in T$.

A.2: Closeness of the technology set T.

A.3: Boundedness of output set: $Z(\mathbf{x}, l, \mathbf{c}, \mathbf{b}) = \{\mathbf{y} : (\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) \in T\}$, $\forall (\mathbf{x}, l, \mathbf{c}, \mathbf{b}) \in \mathbb{R}^{U+1+C+B}_+$.

A.4: Strong disposability of outputs: if $(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) \in T$ then $0 \leqq \mathbf{y}' \leqq \mathbf{y} \Rightarrow \mathbf{y}' \in T$.

A.5.A: Strong disposability of inputs $(\mathbf{x}, l, \mathbf{b})$: if $(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) \in T$ then $(\mathbf{x}', l', \mathbf{c}, \mathbf{b}') \geqq (\mathbf{x}, l, \mathbf{c}, \mathbf{b}) \Rightarrow (\mathbf{x}', l', \mathbf{c}, \mathbf{b}', \mathbf{y}) \in T$.

A.5.B: Convexity of the set: $V(\mathbf{y}; \mathbf{x}, l, \mathbf{b}) = \{\mathbf{c} : (\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) \in T\}, \forall (\mathbf{y}; \mathbf{x}, l, \mathbf{b})$.

[1] My thanks go to Professor Dr Robert G. Chambers who suggested this method.

Under these assumptions, a Farrell output efficiency score can represent the multi-output technology:

$$E(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) = \max\{e \in \mathbb{R}_+ \colon (\mathbf{x}, l, \mathbf{c}, \mathbf{b}, e\mathbf{y}) \in T\} \tag{1}$$

if $\exists\, e$ s.t. $(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, e\mathbf{y}) \in T$ and 0 otherwise, and where $E : \mathbb{R}_+^U \times \mathbb{R}_+ \times \mathbb{R}_+^C \times \mathbb{R}_+^B \times \mathbb{R}_+^S \to \mathbb{R}_+$. By strong disposability of outputs **A.4**:

$$E(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) \geq 1 \Leftrightarrow (\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}) \in T \tag{2}$$

so that $E(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y})$ is a complete function representation of the technology. In this multi-output technology, one can isolate a scalar output y^M (in this study, maize), and a sub-vector of outputs $\mathbf{y}^N$ to be all the other outputs. As the efficiency measure E is positively homogeneous of degree minus one in outputs $(\mathbf{y}^N, y^M)$, it satisfies:

$$E(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mu\mathbf{y}^N, \mu y^M) = \mu^{-1} E(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}^N, y^M), \quad \mu > 0. \tag{3}$$

By following definition (3) and taking $\mu = (y^M)^{-1}$, one can rewrite the maximal producible output associated with input and output levels $(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}^N, y^M)$ as:

$$E(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \frac{\mathbf{y}^N}{y^M}, 1) = y^M E(\mathbf{x}, l, \mathbf{c}, \mathbf{b}, \mathbf{y}^N, y^M), \quad y^M > 0. \tag{4}$$

2.2 Yield index decomposition

A yield index for farm 1 is defined as the ratio of its yield to a base-level yield (for example, the yield of farm 0):

$$\frac{y_1/l_1}{y_0/l_0}. \tag{5}$$

By using (4), the yield index (5) becomes:

$$\frac{y_1^M/l_1}{y_0^M/l_0} = \frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)/l_1}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)/l_0} \frac{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \mathbf{y}_0^N, y_0^M)}{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \mathbf{y}_1^N, y_1^M)}. \tag{6}$$

The last ratio on the right-hand side is a usual relative efficiency index. It measures relative efficiency of farm 1 with respect to farm 0. The first ratio on the right, instead, is the focus of the attention in this methodological section.

It is a ratio of maximal average producible outputs, given inputs[2]. Leaving out the ratio of land sizes (l_0/l_1), the remaining of this section concentrates on the decomposition of the ratio of maximal producible outputs:

$$\frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}. \tag{7}$$

It is possible to obtain different decompositions of (7). To illustrate, first multiply and divide by $E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_0^N}{y_0^M}, 1) E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1) E(\mathbf{x}_1, l_1, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)$ $E(\mathbf{x}_1, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)$ to obtain:

$$\frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)} = \frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_0^N}{y_0^M}, 1)} \frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}$$

$$\frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_1, l_1, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)} \frac{E(\mathbf{x}_1, l_1, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_1, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)} \frac{E(\mathbf{x}_1, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}. \tag{8}$$

All of these five ratios on the right-hand side are legitimate index numbers. That is, only one argument changes in every ratio, and every ratio measures relative changes in maximal producible output due to the varying argument. In particular, the first of the right-hand side terms represents the distance between two maximal producible outputs given by a change in normalized outputs. The second of the right-hand side terms evaluates instead the distance between two maximal producible outputs given by a change in capital inputs. The third of the right-hand side terms measures the distance between two maximal producible outputs given by a change in soil characteristics. The fourth of the right-hand side terms is the value of the distance between two maximal producible outputs given by a change in land size. Finally, the last of the right-hand side terms quantifies the distance between maximal producible outputs given by a change in variable inputs, keeping normalized outputs, capital inputs, soil characteristics, and land size fixed.

[2]Depending on the issue studied, the analysis is possible at different levels of aggregation allowing for more or less aggregated groups of inputs and outputs to vary separately. Depending on how these aggregates are defined, the decomposition results may change.

The proposed decomposition is not unique. In particular, it is also possible to decompose (7) by multiplying and dividing by

$$E(\mathbf{x}_0, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)E(\mathbf{x}_0, l_0, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_1^N}{y_1^M}, 1).$$

This obtains:

$$\frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)} = \frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)} \frac{E(\mathbf{x}_0, l_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}$$

$$\frac{E(\mathbf{x}_0, l_0, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)} \frac{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_1^N}{y_1^M}, 1)} \frac{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}. \tag{9}$$

Also in this case every right-hand side term represents a proper index. But terms corresponding to changes in same variables in (8) and (9) are not necessarily the same. For example, the soil-quality component need not be the same in the two decompositions:

$$\frac{E(\mathbf{x}_1, l_1, \mathbf{c}_1, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_1, l_1, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)} \neq \frac{E(\mathbf{x}_0, l_0, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}, 1)}. \tag{10}$$

This problem is well known in the productivity literature and it is referred to as path dependency. To illustrate the path dependency issue, one can consider figure 1 where, for exemplifying purposes, only one input x and land size l change. Different paths attribute different measures to changes in x and l. One can determine the paths by changing the variables in different orders. To illustrate, let the comparison be the change between $g(x_1, l_1)$ and $g(x_0, l_0)$ where $g : \mathbb{R}^2_+ \to \mathbb{R}_+$. One can either move from point $g(x_1, l_1)$ to point $g(x_1, l_0)$, and then to point $g(x_0, l_0)$ (first path). Or one can move from point $g(x_1, l_1)$, to point $g(x_0, l_1)$, and then to point $g(x_0, l_0)$ (second path). The problem of the ambiguity of the paths arises because, as in this example in the picture, the portions of the change from $g(x_1, l_1)$ to $g(x_0, l_0)$ attributed to each component are different depending on the path followed. In figure 1 only two different paths are present, but in the case of ratio (7) one hundred and twenty different paths are possible when changing five aggregates. The proposed solution to resolve the ambiguity in the method of decomposition is to follow Fisher in creating his ideal index, and generalize the results by Gini (1937). In other words, one may take the geometric average of the different decompositions to obtain appropriately calculated components.

The proposed decomposition of the yield index into five different index components, together with the efficiency index $EI = \frac{E(\mathbf{x}_0,l_0,\mathbf{c}_0,\mathbf{b}_0,\mathbf{y}_0)}{E(\mathbf{x}_1,l_1,\mathbf{c}_1,\mathbf{b}_1,\mathbf{y}_1)}$, obtains:

$$\frac{y_1^M/l_1}{y_0^M/l_0} = I^{(\mathbf{x}_1,\mathbf{x}_0)}(l_0,\mathbf{c}_0,\mathbf{b}_0,\frac{\mathbf{y}_0^N}{y_0^M};l_1,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M})\bar{L}^{(l_1,l_0)}(\mathbf{x}_0,\mathbf{c}_0,\mathbf{b}_0,\frac{\mathbf{y}_0^N}{y_0^M};\mathbf{x}_1,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M})l_0/l_1$$
$$Q^{(c_1,\mathbf{c}_0)}(\mathbf{x}_0,l_0,\mathbf{b}_0,\frac{\mathbf{y}_0^N}{y_0^M};\mathbf{x}_1,l_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M})K^{(\mathbf{b}_1,\mathbf{b}_0)}(\mathbf{x}_0,l_0,\mathbf{c}_0,\frac{\mathbf{y}_0^N}{y_0^M};\mathbf{x}_1,l_1,\mathbf{c}_1,\frac{\mathbf{y}_1^N}{y_1^M})$$
$$Y^{(\frac{\mathbf{y}_1^N}{y_1^M},\frac{\mathbf{y}_0^N}{y_0^M})}(\mathbf{x}_0,l_0,\mathbf{c}_0,\mathbf{b}_0;\mathbf{x}_1,l_1,\mathbf{c}_1,\mathbf{b}_1)\,EI, \tag{11}$$

where I is a variable-inputs component, $\bar{L}$ is a land-size component, Q is a soil-quality component, K is a capital-inputs component, and Y is an outputs component. Each index can be decomposed in different ways by varying the decomposition paths. Collecting the equal terms and following de Boer (2009) the index for variable inputs $\mathbf{x}$ can be, for instance, restated as follows:

$$I^{(\mathbf{x}_1,\mathbf{x}_0)}(l_0,\mathbf{c}_0,\mathbf{b}_0,\frac{\mathbf{y}_0^N}{y_0^M};l_1,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M}) = \tag{12}$$

$$\left\{\frac{E(\mathbf{x}_1,l_1,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_1,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/5}\left\{\frac{E(\mathbf{x}_1,l_1,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_0^N}{y_0^M},1)}{E(\mathbf{x}_0,l_1,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_0^N}{y_0^M},1)}\right\}^{1/20}\left\{\frac{E(\mathbf{x}_1,l_1,\mathbf{c}_1,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_1,\mathbf{c}_1,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/20}$$

$$\left\{\frac{E(\mathbf{x}_1,l_1,\mathbf{c}_0,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_1,\mathbf{c}_0,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/20}\left\{\frac{E(\mathbf{x}_1,l_0,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_0,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/20}\left\{\frac{E(\mathbf{x}_1,l_1,\mathbf{c}_1,\mathbf{b}_0,\frac{\mathbf{y}_0^N}{y_0^M},1)}{E(\mathbf{x}_0,l_1,\mathbf{c}_1,\mathbf{b}_0,\frac{\mathbf{y}_0^N}{y_0^M},1)}\right\}^{1/30}$$

$$\left\{\frac{E(\mathbf{x}_1,l_1,\mathbf{c}_0,\mathbf{b}_1,\frac{\mathbf{y}_0^N}{y_0^M},1)}{E(\mathbf{x}_0,l_1,\mathbf{c}_0,\mathbf{b}_1,\frac{\mathbf{y}_0^N}{y_0^M},1)}\right\}^{1/30}\left\{\frac{E(\mathbf{x}_1,l_0,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_0^N}{y_0^M},1)}{E(\mathbf{x}_0,l_0,\mathbf{c}_1,\mathbf{b}_1,\frac{\mathbf{y}_0^N}{y_0^M},1)}\right\}^{1/30}\left\{\frac{E(\mathbf{x}_1,l_0,\mathbf{c}_0,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_0,\mathbf{c}_0,\mathbf{b}_1,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/30}$$

$$\left\{\frac{E(\mathbf{x}_1,l_0,\mathbf{c}_1,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_0,\mathbf{c}_1,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/30}\left\{\frac{E(\mathbf{x}_1,l_1,\mathbf{c}_0,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_1,\mathbf{c}_0,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/30}\left\{\frac{E(\mathbf{x}_1,l_0,\mathbf{c}_0,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}{E(\mathbf{x}_0,l_0,\mathbf{c}_0,\mathbf{b}_0,\frac{\mathbf{y}_1^N}{y_1^M},1)}\right\}^{1/20}$$

$$\left\{\frac{E(\mathbf{x}_1, l_0, \mathbf{c}_0, \mathbf{b}_1, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_1, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}\right\}^{1/20} \left\{\frac{E(\mathbf{x}_1, l_0, \mathbf{c}_1, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_1, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}\right\}^{1/20} \left\{\frac{E(\mathbf{x}_1, l_1, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_0, l_1, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}\right\}^{1/20}$$
$$\left\{\frac{E(\mathbf{x}_1, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}{E(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}, 1)}\right\}^{1/5}.$$

This index shows the change in the maximal producible outputs due to the change in agricultural inputs $\mathbf{x}$, while keeping the other inputs and outputs $(l_., \mathbf{c}_., \mathbf{b}_., \frac{\mathbf{y}_.^N}{y_.^M})$ constant. The other four indexes are similar to this, and need not be repeated here.

One can obtain empirical estimates of the six components by applying nonparametric linear programming methods without specific assumptions on returns to scale. In the estimation the choice of the reference farmer $(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \mathbf{y}_0)$ imposes just a different normalization on the productivity estimates. Depending on the normalization, the estimates might appear infeasible in some cases. To minimize this issue, the reference unit in this study is a farmer with very high production potential (highest input levels and average soil characteristics) and lowest realized output:

$$\begin{aligned}(\mathbf{x}_0, l_0, \mathbf{c}_0, \mathbf{b}_0, \mathbf{y}_0) &= (\max\{\mathbf{x}\}_u, \max\{l\}, mean\{\mathbf{c}\}_c, \max\{\mathbf{b}\}_b, \min\{\mathbf{y}\}_s),\\ &\forall u \in 1, \ldots, U,\ c \in 1, \ldots, C,\ b \in 1, \ldots, B,\ s \in 1, \ldots, S. \qquad (13)\end{aligned}$$

2.3 Testing for the inverse land size-yield relationship

The land-size component to production important for testing the inverse land size-yield relationship is an average land-size component, which one can obtain by multiplying the index $\bar{L}$ by the land ratio l_0/l_1:

$$L^{(l_1,l_0)}(\mathbf{x}_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}; \mathbf{x}_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}) = l_0/l_1\, \bar{L}^{(l_1,l_0)}(\mathbf{x}_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}; \mathbf{x}_1, \mathbf{c}_1, \mathbf{b}_1, \frac{\mathbf{y}_1^N}{y_1^M}).$$

L can be considered as the average product of land normalized on the reference level chosen. The inverse land size-yield relationship suggests that an increase in land size causes a decrease in average product of land: $L^{(l_i,l_0)}(\mathbf{x}_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}; \mathbf{x}_i, \mathbf{c}_i, \mathbf{b}_i, \frac{\mathbf{y}_i^N}{y_i^M})$. In terms of the decomposition, testing for the inverse land size-yield relationship is equivalent to testing:

$$\frac{\partial}{\partial l_i}\left(L^{(l_i,l_0)}(\mathbf{x}_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}; \mathbf{x}_i, \mathbf{c}_i, \mathbf{b}_i, \frac{\mathbf{y}_i^N}{y_i^M})\right) < 0. \qquad (14)$$

One may test this hypothesis through robust weighted least squares regression of the average product $L^{(l_i,l_0)}(\mathbf{x}_0, \mathbf{c}_0, \mathbf{b}_0, \frac{\mathbf{y}_0^N}{y_0^M}; \mathbf{x}_i, \mathbf{c}_i, \mathbf{b}_i, \frac{\mathbf{y}_i^N}{y_i^M})$ on land size l_i and on a quadratic term of land size l_i^2:

$$L^{(l_i,l_0)} = \zeta_L + \alpha_L l_i + \beta_L l_i^2 + \eta_{Li}. \tag{15}$$

If a negative relationship is present, this is evidence of an inverse land size-yield relationship.
Even if an inverse relationship exists in (15), it is possible that different land average products of interest to the present study are not inversely related to land size. For example, one can test the hypothesis that a long-term soil-quality adjusted average land product is negatively correlated to land size through the following robust regression:

$$L^{(l_i,l_0)} Q^{(\mathbf{c}_i,\mathbf{c}_0)} = \zeta_{LQ} + \alpha_{LQ} l_i + \beta_{LQ} l_i^2 + \eta_{LQi}.$$

It is of interest to test also whether the average land product, adjusted for short-term productivity-increasing practices (such as the usage of fertilizers, or of hand hoes), is negatively correlated to land size through the following robust regression:

$$L^{(l_i,l_0)} I^{(\mathbf{x}_i,\mathbf{x}_0)} = \zeta_{LI} + \alpha_{LI} l_i + \beta_{LI} l_i^2 + \eta_{LIi}.$$

In addition, one may test if an average land product, adjusted both for short-term productivity-increasing practices and long-term soil-quality, is negatively correlated to land size through the following robust regression:

$$L^{(l_i,l_0)} Q^{(\mathbf{c}_i,\mathbf{c}_0)} I^{(\mathbf{x}_i,\mathbf{x}_0)} = \zeta_{LQI} + \alpha_{LQI} l_i + \beta_{LQI} l_i^2 + \eta_{LQIi}.$$

On the other hand one may consider whether the average product of land, adjusted for capital inputs, is negatively correlated with land size through the following robust regression:

$$L^{(l_i,l_0)} K^{(\mathbf{b}_i,\mathbf{b}_0)} = \zeta_{LK} + \alpha_{LK} l_i + \beta_{LK} l_i^2 + \eta_{LKi}.$$

Consequently, a further robust regression may relate an average land product, adjusted for long-term soil-quality and capital inputs, with land size:

$$L^{(l_i,l_0)} Q^{(\mathbf{c}_i,\mathbf{c}_0)} K^{(\mathbf{b}_i,\mathbf{b}_0)} = \zeta_{LQK} + \alpha_{LQK} l_i + \beta_{LQK} l_i^2 + \eta_{LQKi}.$$

Whether a maximal average producible output index that excludes changes in other outputs is correlated with land size is testable through the following robust regression:

$$L^{(l_i,l_0)}Q^{(\mathbf{c}_i,\mathbf{c}_0)}K^{(\mathbf{b}_i,\mathbf{b}_0)}I^{(\mathbf{x}_i,\mathbf{x}_0)} = \zeta_{LQKI} + \alpha_{LQKI}l_i + \beta_{LQKI}l_i^2 + \eta_{LQKIi}.$$

Finally, one may test whether a yield index measure—that includes appropriately the effect of outputs' diversification, but excludes inefficiency—is correlated with land size by performing the following robust regression:

$$L^{(l_i,l_0)}Q^{(\mathbf{c}_i,\mathbf{c}_0)}K^{(\mathbf{b}_i,\mathbf{b}_0)}I^{(\mathbf{x}_i,\mathbf{x}_0)}Y^{(\frac{\mathbf{y}_i^N}{y_i^M},\frac{\mathbf{y}_0^N}{y_0^M})} = \zeta_{LQKIY} + \alpha_{LQKIY}l_i + \beta_{LQKIY}l_i^2 + \eta_{LQKIYi}.$$

3 Data

The empirical data in this study are a random sample of households from 99 sub-locations in Kenya. Data are from early 2007 and refer to preceding short and long growing seasons. The survey is named "Research on Poverty, Environment and Agricultural Technologies (REPEAT): Panel studies in Africa". Survey data are obtained from the National Graduate Institute for Policy Studies (21st century Center of Excellence Program) in Japan. The present cross-section sample is composed of 590 household farmers of which only 452 have data on soil quality.[3] The decomposition proposed in this study requires strictly positive maize yield: only 443 households satisfy this condition, and these are in the final sample.

Variables used in this analysis are on 9 inputs and 3 outputs. Table 2 shows input and output summary statistics for the households. Main agricultural output is harvested maize which is totally rain-fed. All present farmers produce maize. Other representative outputs are milk and off-farm income. Apart from land area, the measured variable inputs are seeds, fertilizers, cost of temporary hired workers, and hand hoes. The label 'capital inputs' is used somehow arbitrarily to design the human labor capital, and animal capital of the households: household members, adjusted for age and educational level,

[3] More households are included in the original sampling scheme but are left out of the analysis because have incomplete entries. No soil-quality chemical quantitative characteristics in household farmers panel surveys are available in developing countries until the present moment.

and number of dairy cows.[4]

Data on physical characteristics of land in the largest maize plot of each household are available for mid-2003. The present analysis focuses on critical measures of soil structure and soil texture. The critical soil characteristic for soil structure is soil carbon, while the soil characteristic that represents soil texture is soil clay content. These variables indicate the long-term soil quality available to the households. Soil carbon is modifiable only in the long-run, while soil clay is hardly changeable.

Agriculture is mainly manual. Hand hoes are common: the median is 4 hand hoes per household. Median household annual off-farm earnings from all sources are 41100 Kenyan Shillings ($587 in February 2007 dollars). Households at the 25^{th} percentile of the off-farm income distribution earned the equivalent of about $163. The median household head education level is primary. Median application of fertilizers on maize fields is 375 Kgs per family. Median household maize harvest is 765 Kgs per family. Median household size is 4.5 people, who live at home, plus another family member who works off-farm. Half of the households have at most 1 dairy cow, which may also be used as draft power.

4 Results

Figure 2 depicts visually one of the main results of this study:[5] the relationship between land area, and the productivity components of land size $L^{(l_i,l_0)}$

[4]Faithful to the human capital approach pioneered by Jorgenson and Griliches (1967), on-farm labor input is adjusted for differences in quality, both due to education and age. To adjust labor for education, estimates of the impact of education on agricultural productivity in Kenya from the analysis by Husbands et al. (1996) are used. Given an 85% probability that a primary school-completer household head increases household profitability by 40%, an expected return equal among years of primary education (8 years) obtains a 4.25% average increase in productivity per additional year of any level of education. It is possible that different methods of partitioning the increase in productivity would have achieved a different result. But this topic is not the focus of the present contribution. The same is done for households with a primary school completer, who is not the household head (increase by 29.75%). All of the family members living at the household that are above three years old are included. Most kids start completing tasks in the family at this age. To adjust labor for age, number of children who, by United Nations Children Fund classification, are considered those less than 15 years old, are divided in half.

[5]Due to infeasibilities in the linear programs under convexity of the soil-characteristics requirement set, only 401 units have complete results.

and soil quality $Q^{(\mathbf{c}_i,\mathbf{c}_0)}$. The average product of land is negatively correlated to land acreage, signaling a decreasing average product of land when increasing area. This negative correlation is evidence in favor of an inverse land size-yield relationship.

Figure 2 also pictures the relationship between the soil-quality component and land area. The quality component is more variable among families with plots of middle size than among families with either very small or very large plots, which are of better quality. While farmers with very small plots (less than 0.4 acres) seem to be constrained to produce the staple maize in plots with very good quality to obtain a subsistence level of production, farmers with plots larger than 0.4 acres plant maize also in plots of lower quality. A potential explanation is that these farmers with medium extensions need not plant a staple, such as maize, in their best-quality plots to obtain a subsistence level of production. If farmers plant more than 2 acres, it is interesting to notice that the variability in soil-quality component decreases. Farmers with extensions larger than 2 acres apparently seek obtaining (possibly because of specialization) plentiful maize harvests by planting on good-quality soil.

Figure 3 presents a matrix of all six yield components. The graphs outside the diagonal represent each component against the others. The graphs on the diagonal are histograms that show the distributions of each component: efficiency, land size, quality, capital inputs, variable inputs, and outputs.

The first element on the diagonal from the upper-left corner evidences how the efficiency scores have a mass around 1. The non-efficient households have a modal peak around 0.4 in the range between 0 and 1. The second element on the diagonal is the histogram of the average product of land. Most estimates of land average product are concentrated in the lower half of the range, with only very few in the higher half. As can be seen in figure 2, low average land product estimates correspond to farmers with large plots, while very large average land product estimates correspond to households with small plots. Thus, the negative land size-yield relationship seems to be especially due to the few farmers with small plots and with very high average land product.

The third element on the diagonal from the left depicts the distribution of the estimated soil-quality components. Even though soil-quality component estimates are more concentrated around and below 1, there are estimates also higher than 1. Numbers higher than 1 reflect the possibility that the average values of soil characteristics are not the values that allow highest producible output under convexity of the soil-characteristics requirement set.

As can be seen in figure 4, soil-quality component estimates are non-monotonic with respect to increases in the soil characteristics. Both soil characteristics show first an increasing, and then a decreasing portion. In particular, estimated soil-quality components plotted against soil carbon present a steep increase up to around 3% of soil weight, and a smooth decrease thereafter. A decreasing soil-quality component implies congestion of soil carbon after 3%. On the other hand, along the soil-clay distribution, increases and decreases of the soil-quality component are less dramatic and smoother. Nonetheless, percentages of soil clay above 26% appear generally detrimental to soil quality.

Estimated capital-inputs components are depicted as the fourth plot on the diagonal. Because of the assumption of strong disposability on these inputs, estimates of the capital-inputs component distribute with increasing probability up to the maximum value of 1.

Non-monotonic distributions, instead, are shown both for the variable-inputs component, the fifth element on the diagonal, and for the outputs component, the last element on the diagonal. The estimates of the variable-inputs component appear bi-modal, signaling the presence of two distinct groups of households: one group with estimated variable-inputs components less than half the component of the reference farm (mode around 0.2), and one group with estimated components more similar to the reference farm (mode around 0.75). The group with low levels of variable-inputs component hires less labor, and uses lower amount of seeds than the group with higher variable-inputs components.

The estimates of the outputs component show also a bi-modal distribution: some component estimates are concentrated around 1, while others have lower mode around 0.45. Farmers whose estimated output components are at 1 attain a potential level of production at least as high as the level attainable by the reference farm. Farmers who have an outputs component lower than 1 may have decided to diversify more their production. Because the index decomposition is in terms of yield, the farmers who obtain an outputs component around 1 are more specialized in farming than in milk production or in working off-farm. Farmers with an outputs component lower than 1 are instead substituting more maize harvests with other products, at different degrees.

In this setting, it is expected that farmers diversify production. Figure 5 represents the output mixes of the farmers with maize harvest less than 5000

Kgs, with less than 10000 litres of milk produced, and less than 600000 KSh of earned off-farm income. As visible in figure 5, the farmers with low harvests show diversification in output production, especially in milk production. It is sensible to expect synergies in the production of milk and maize. However, as expected, there is less synergy between production of maize and off-farm employment.

In the scatter plots outside the diagonal in figure 3 a negative relationship between average product of land component and variable inputs component is present: high estimates of average land product are related to low levels of variable-inputs component. This negative relationship signals that farmers with small plots (who have high average land product) employ less variable inputs than farmers with large plots. The same farmers with low usage levels of variable inputs are also somewhat more efficient, compared to farmers who have high usage levels of variable inputs.

The estimates of the variable-inputs component do not appear correlated significantly in any direction to the estimates of the soil-quality components. Different levels of variable-inputs components are roughly possible at any quality component level. However, high values of the quality component appear correlated to high capital-inputs components, to low levels of efficiency, and to low or medium levels of the average land product. In other words, high soil quality is in highly capitalized farmers with relatively large plots. As also shown in figure 2, this is evidence once more of a somewhat unequal distribution of soil quality along the land-size distribution.

4.1 Test results on the inverse land size-yield relationship

In figure 6, ten graphs show different aggregates of productivity components and the results of the fitted robust regressions against land size. If there is an inverse land size-yield relationship (as it is suggested in figures 2 and 3), it is critical to understand whether it is statistically significant and how farmers' present input and output choices, and long-term soil quality interact with the inverse land size-yield relationship. The numerical results of the robust regressions are in table 2.

The plot on the top-left corner of figure 6 represents the average product of land $L^{(l_i,l_0)}$ against land size. There is a strong negative relationship between the average product of land $L^{(l_i,l_0)}$ and land area. The results support hypothesis (14) of a negative relationship between yield and land area. This key result for the present study is corroborated by very strongly negatively significant coefficients in table 2. The results support a convex decay of the average product of land at increases in land area. The convexity also reinforces the finding that such an inverse land size-yield relationship may be due to few small farmers with very high average products of land.

The top-right graph of figure 6 pictures a quality-adjusted average product of land, derived as the multiplication of the quality component $Q^{(\mathbf{c}_i,\mathbf{c}_0)}$ and the average product of land $L^{(l_i,l_0)}$, against land area. The graph shows that even a quality-adjusted average product of land is negatively correlated to land area. Indeed, the higher variability in the long-term soil-quality components in figure 2 comes mostly from the farmers in the middle of the size distribution. Even though the precision and the strength of the negative coefficients of the first panel of table 2 are weakened, the results still support a convex decay of the quality-adjusted average product of land, when land area increases.

The left graph on the second row of figure 6 shows how land area is related to an input-adjusted average product of land, derived as the multiplication of the average product of land $L^{(l_i,l_0)}$ and the variable-inputs component $I^{(\mathbf{x}_i,\mathbf{x}_0)}$. This input-adjusted land component looks at the possibility that short-term productivity increasing agricultural practices (such as usage of fertilizers, or of hand hoes) might reduce the intensity of the negative land size-yield relationship. Indeed, the intensity and the precision of the coefficients in the quadratic regression are lower than when considering only the average product of land. Nonetheless, the quantitative results in table 2 maintain a (positively) convex decreasing relationship between an average product of land, adjusted for short-term productivity increasing practices, and land area. Short-term productivity increasing practices have a stronger impact in diluting the inverse land size-yield relationship than long-term soil quality has.

The graph on the right in the second row of figure 6 looks at the interaction of short-term productivity-increasing practices $I^{(\mathbf{x}_i,\mathbf{x}_0)}$ and long-term soil-quality $Q^{(\mathbf{c}_i,\mathbf{c}_0)}$ with average product of land $L^{(l_i,l_0)}$ against land size. From the graphical results it is possible to see that the inverse land size-yield relationship is further diluted by interacting soil quality and variable inputs. The statistical results confirm this intuition in table 2. The coefficients that characterize a (positively) convex decreasing relationship remain significant even if they are weaker than when each productivity component ($I^{(\mathbf{x}_i,\mathbf{x}_0)}$ or

$Q^{(\mathbf{c}_i,\mathbf{c}_0)}$) is considered alone in interaction with $L^{(l_i,l_0)}$.

The graph on the left in the third row of figure 6 portrays an average product of land adjusted for capital inputs $K^{(\mathbf{b}_i,\mathbf{b}_0)}$. The (positively) convex decreasing relationship between land size and yield remains strongly significant. Including, additionally to capital inputs $K^{(\mathbf{b}_i,\mathbf{b}_0)}$, long-term soil quality $Q^{(\mathbf{c}_i,\mathbf{c}_0)}$ in the graph on the right of the third row of figure 6 leaves the inverse land size-yield relationship significant and convex. It is worth noting that the long-term soil-quality component mitigates the inverse relationship further when interacted with capital inputs.

The left graph on the fourth row of figure 6 displays an index of maximal average producible output ($L^{(l_i,l_0)}\ Q^{(\mathbf{c}_i,\mathbf{c}_0)}\ K^{(\mathbf{b}_i,\mathbf{b}_0)}\ I^{(\mathbf{x}_i,\mathbf{x}_0)}$), which excludes the effects of outputs' diversification, against land size. Despite the relationship is weaker than in previous regression coefficients of table 2, there is still a significantly (positively) convex decreasing relationship of $L^{(l_i,l_0)}\ Q^{(\mathbf{c}_i,\mathbf{c}_0)}$ $K^{(\mathbf{b}_i,\mathbf{b}_0)}\ I^{(\mathbf{x}_i,\mathbf{x}_0)}$ with land size. In particular, from the comparison with previous regression coefficients, it is clear that the impact of the variable-inputs component is strong in diminishing the precision of the quadratic term estimate. As expected, the farmers with small plots use low levels of variable inputs.

Finally, the right graph of the fourth row of figure 6 presents a maximal average producible output measure that accounts for outputs' diversification ($L^{(l_i,l_0)}Q^{(\mathbf{c}_i,\mathbf{c}_0)}\ K^{(\mathbf{b}_i,\mathbf{b}_0)}\ I^{(\mathbf{x}_i,\mathbf{x}_0)}\ Y^{(\frac{\mathbf{y}_i^N}{y_i^M},\frac{\mathbf{y}_0^N}{y_0^M})}$) against land size. The highest values of maximal average producible output $L^{(l_i,l_0)}Q^{(\mathbf{c}_i,\mathbf{c}_0)}K^{(\mathbf{b}_i,\mathbf{b}_0)}I^{(\mathbf{x}_i,\mathbf{x}_0)}$, which excludes the outputs component in the left graph of the same row, are decreased further when including the outputs component. The numerical results in table 2 show a totally insignificant quadratic term, and a non strongly significant linear effect. The low significance of the linear effect is jeopardized by the insignificance of the quadratic term. Theory wants that the marginal effect of the lower-order term in this quadratic regression cannot be interpreted if the higher-order term is insignificant. Nonetheless, if one looks at a simple linear relationship, there is still a negatively significant linear correlation. The numerical coefficients are presented in table 2.

The (positively) convex inverse land size-yield relationship disappears when including the productivity effect of outputs' diversification in terms of yield. Among farmers with small plots, the estimated outputs components are lower than among farmers with large plots. As can be seen from figure 7, farmers with small plots, who have high average product of land (last plot from figure 7) but low estimated outputs components, are more diversified into non-maize production than farmers with large plots. The relative productivity, in yield terms, of higher outputs' diversification is lower for these farmers compared to other farmers who dedicate more of their production efforts to the production of maize. Accounting appropriately for the contribution in yield terms of farmers' diversification into other outputs, while controlling for inefficiency, dissipates the convexity of the inverse land size-yield relationship.

In the last row of figure 6 the efficiency estimates (left graph) and the maize yields (right graph) are plotted against land area. A U-shaped relationship of efficiency with land size is apparent. Many farmers with plots less than 1 acre (most with plots below 0.4 acres) and many farmers with plots above 4 acres are completely efficient while an increasing variability of efficiency estimates is visible among farmers with plots between 0.4 and 4 acres. The coefficients in table 2 support this conclusion. As expected, the results in the right graph and the numerical coefficients in table 2 support a (positively) convex decay of yield when increasing land size.

5 Conclusions

The methods presented in this study help understanding the long-debated inverse land size-yield relationship, and some of its possible explanations. The inverse relationship is analyzed by accounting for productivity differences of inputs and outputs with methods that drop assumptions on returns to scale, production efficiency, and parametric assumptions on the technology. Not assuming production efficiency allows decomposing a ratio of maximal producible outputs and not of observed yields, which may include inefficiency. Absence of specific technological functional form assumptions (apart from piecewise linearity) allows not imposing unrealistic properties among inputs and outputs, and specific returns to scale a priori.

The methodology purges out the inefficiency from a yield index, and directly decomposes, under variable returns to scale, the distance among maximal producible outputs into five components: land size, soil-quality characteristics, variable inputs, capital inputs, and outputs. The present study proposes to calculate the six components with nonparametric productivity accounting methods.

The results show, visually and quantitatively, a strongly significant (positively) convex inverse relationship between yield and land size. The yield components estimated show a strongly significant (positively) convex inverse relationship of average product of land with land area, which is evidence in support of an inverse land size-yield relationship. Only by accounting appropriately for the productivity in yield terms of diversification into products different from maize, the convexity of the inverse land size-yield relationship disappears in favor of a linear inverse relationship.

The results imply that in this sample of Kenyan households the long-standing empirical phenomenon of an inverse land size-yield relationship is confirmed. In particular, if the implications of this analysis were to be pushed further, faster agricultural development would come from farmers with production patterns similar to the farmers with smallest extensions of maize. These farmers with very small plots use little levels of agricultural inputs in an efficient manner on good-quality soil to produce not only maize but a portfolio of differentiated products. These conclusions are nonetheless only valid for this sample and for these technology assumptions.

Considering the importance of dynamics in soil fertility, this study is moreover only an approximation of the results obtainable if soil-quality household panel data were available. Once these data were to become available, a generalized version of this study would be possible. A generalized study would allow disentangling completely the dynamic interplay of long-term soil quality, household choices, and the land size-yield relationship.

Figure 1: Graphical representation of the problem of path dependency in a two-dimensional case

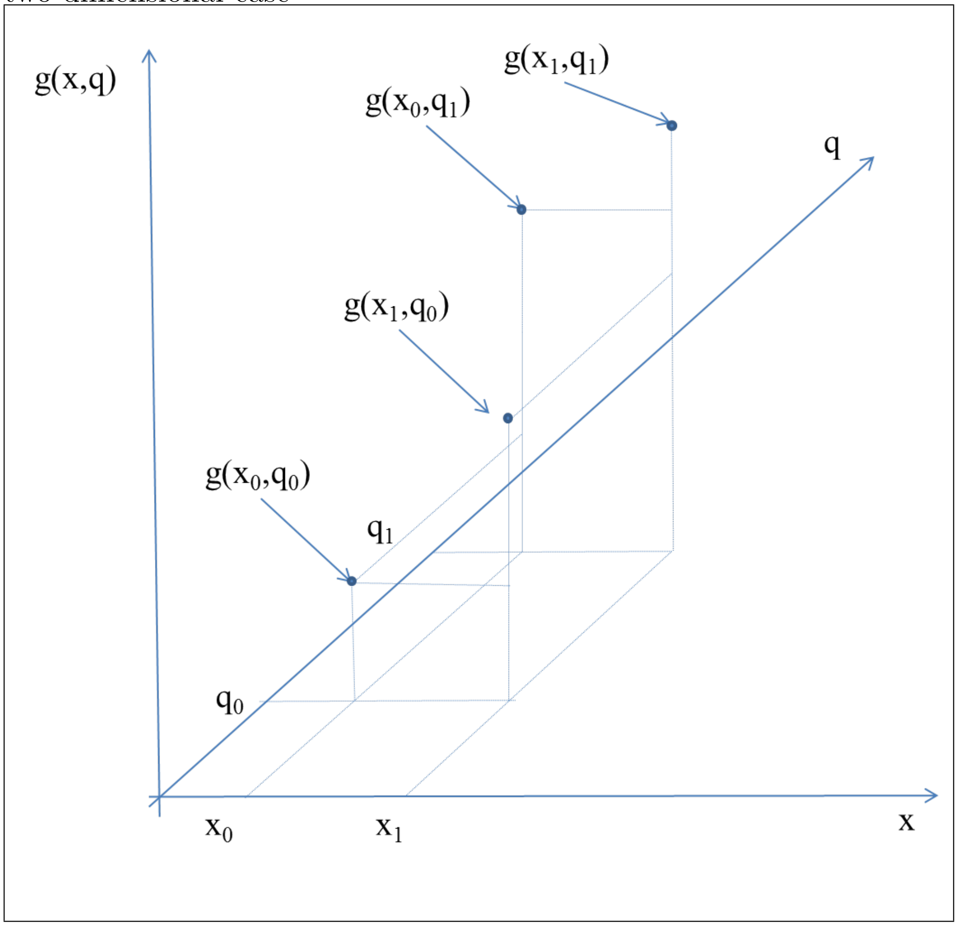


Figure 2: Land size against productivity components of land size and soil quality

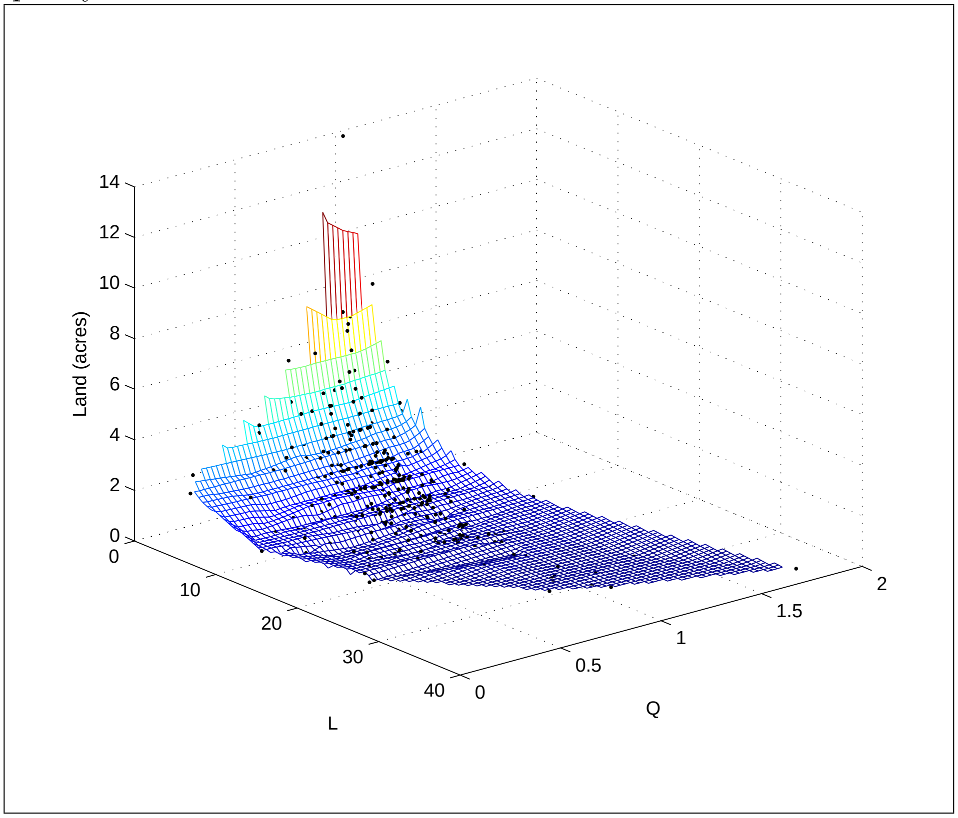


Figure 3: Matrix of plots of yield components under variable returns to scale

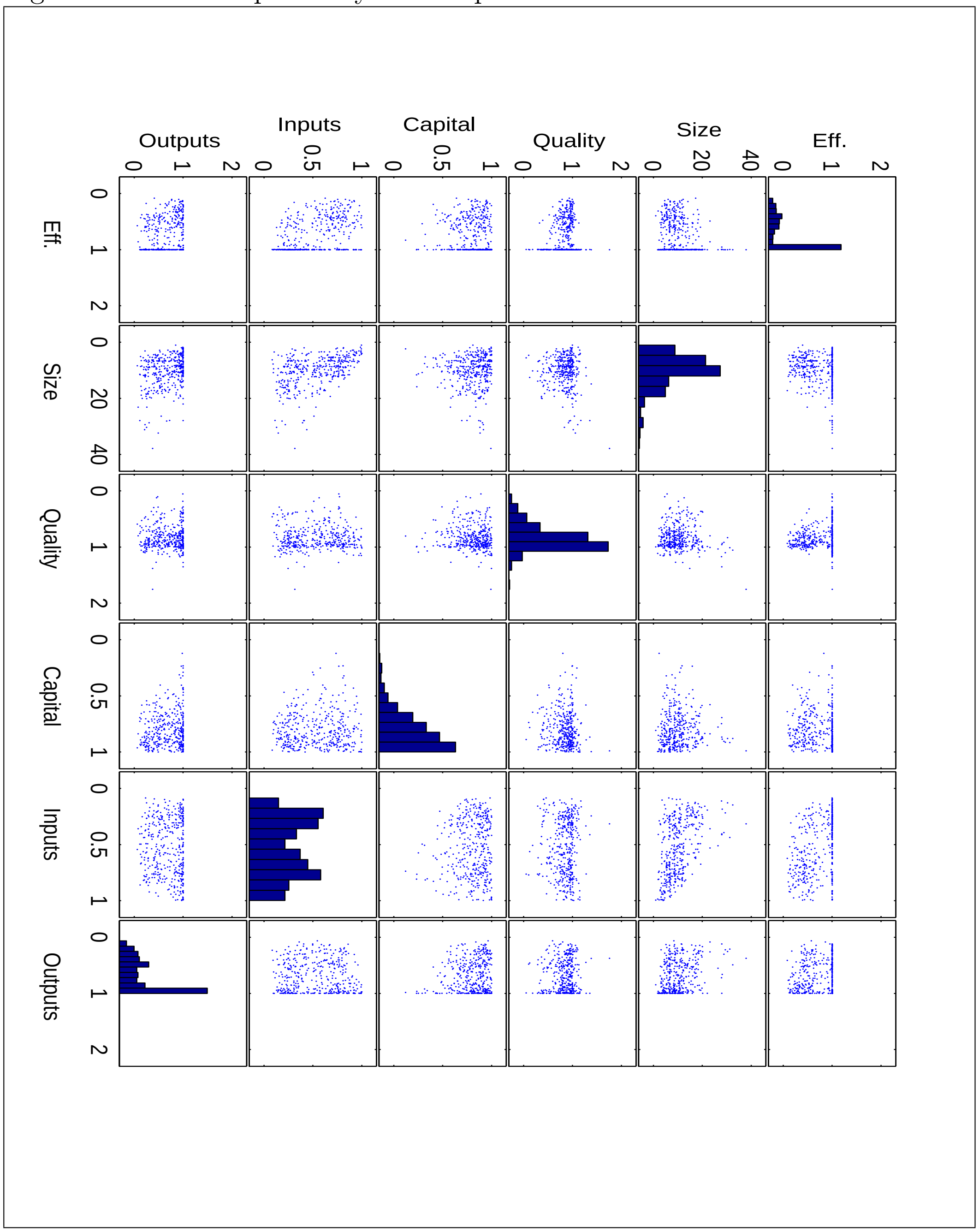


Figure 4: Soil-quality component against soil carbon and soil clay

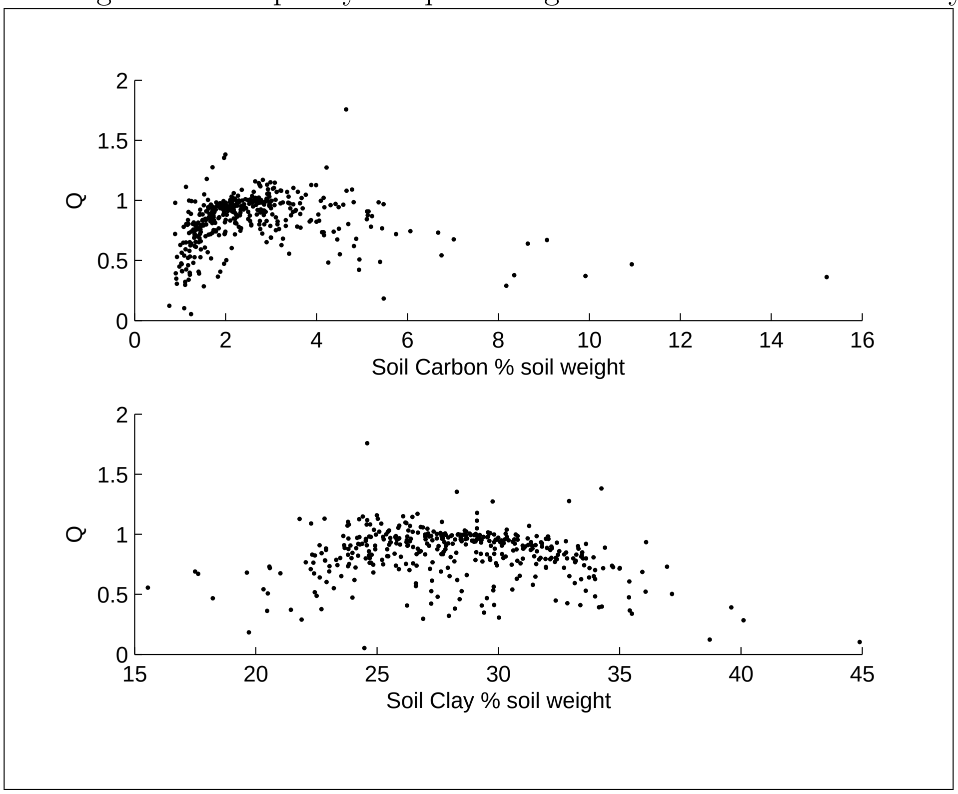


Figure 5: Milk production and off-farm income against maize harvest

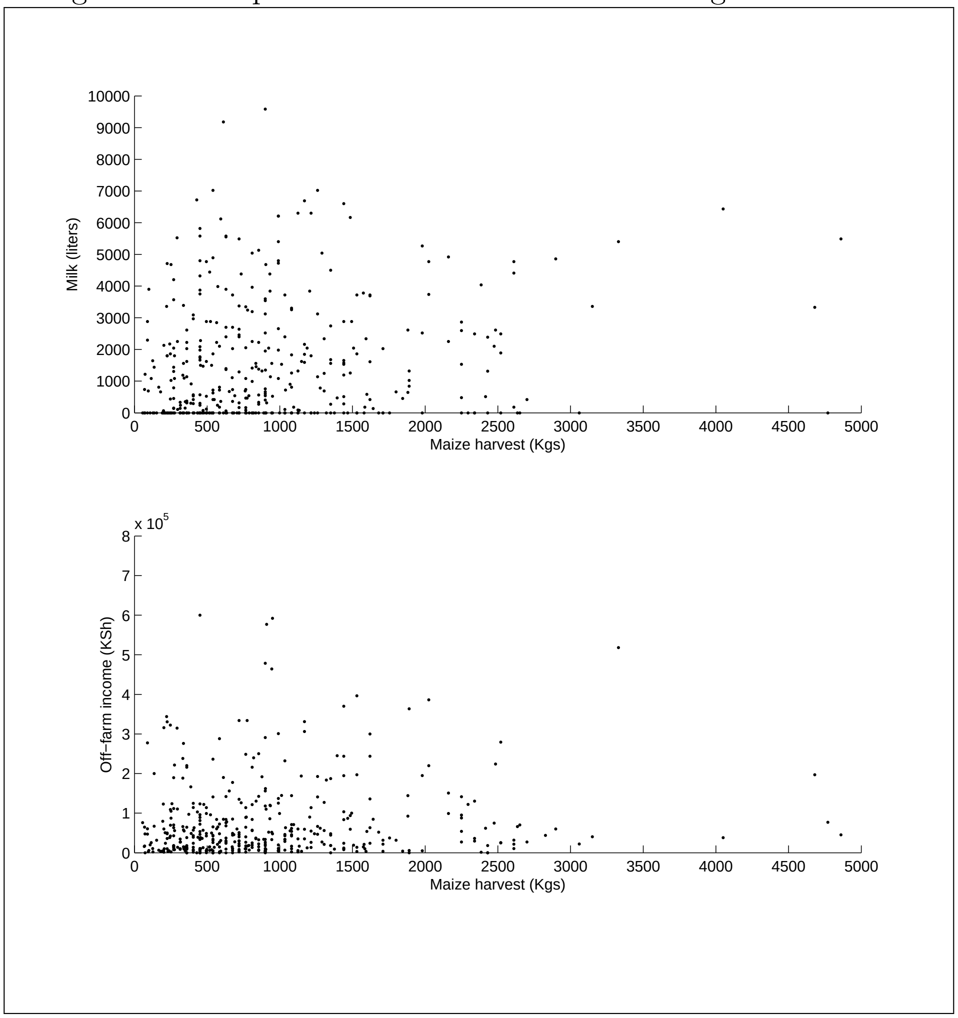
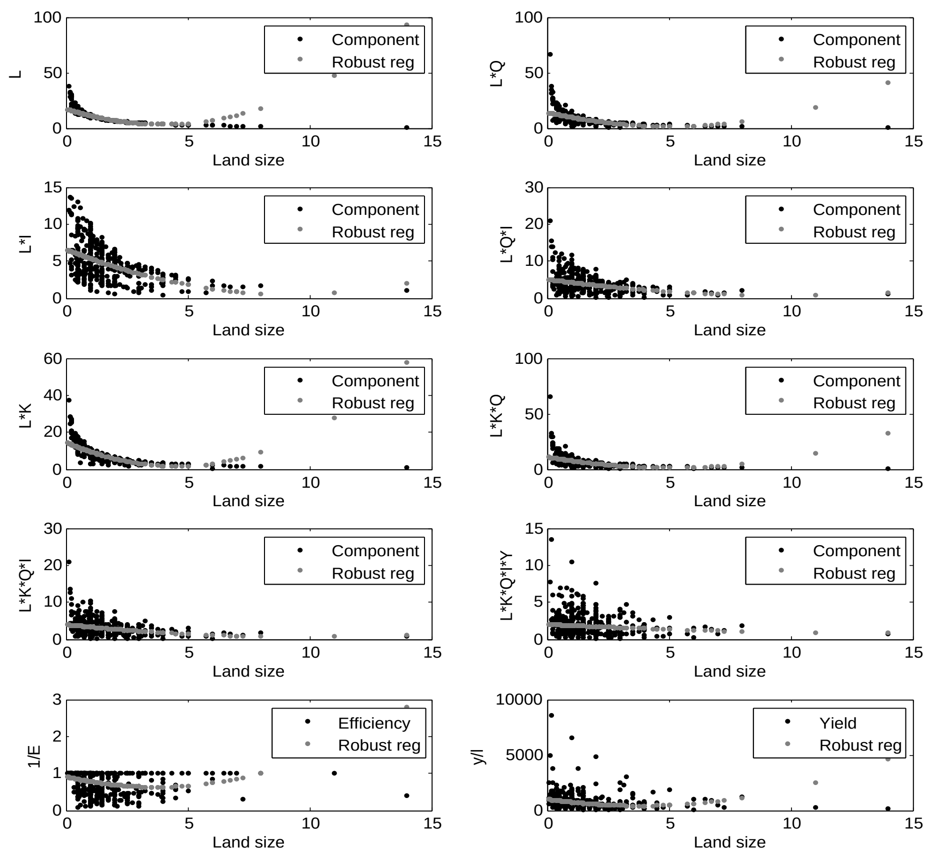


Figure 6: Robust quadratic regressions of different aggregates of yield components under variable returns to scale for a multi-output technology



Note: Components are plotted as black dots. Predictions from robust quadratic regression are plotted with light gray dots.

Figure 7: Output quantities and land-size average product against outputs component

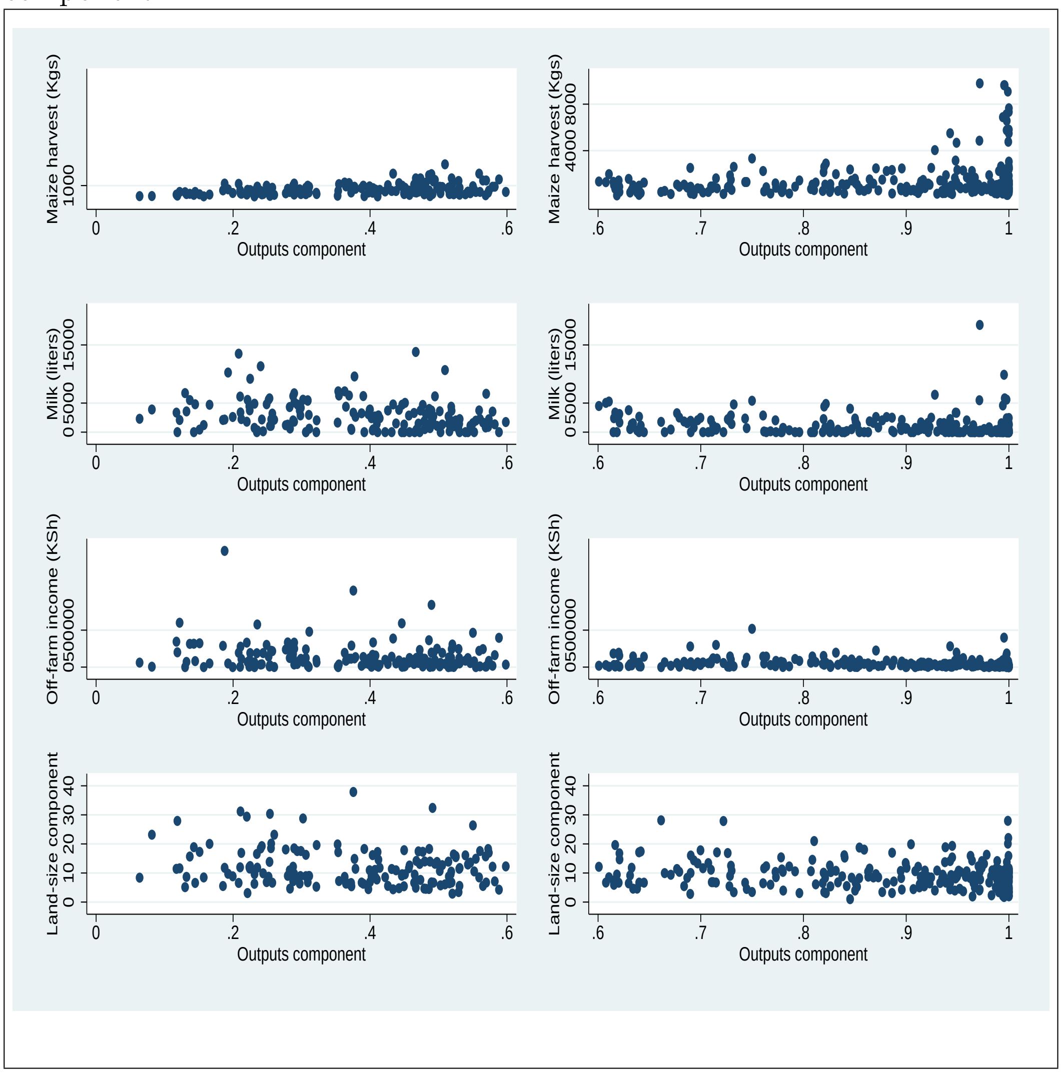


Table 1: Summary statistics of inputs, outputs, and soil-quality physical characteristics

Variable	Mean	Std.Dev.	Min	Max
Inputs				
fertilizers (kgs)	800.1	1233.5	0	9000
quantity of seeds (kgs)	13.4	10.9	1	78
number of hand hoes	3.9	2.1	0	15
hired labor (cost in KSh)	2975.3	4940.4	0	48160
land area (acres)	1.6	1.4	0.1	14
members of the family	5.7	2.8	1	17.5
milking cows	1	0.9	0	5
Outputs				
total harvest maize (kg of dry maize equivalent)	1114.8	1337.9	56.2	9790
milk (liters)	1714.2	2449.1	0	18600
off-farm income (in KSh)	85941	147022.5	0	1572400
Soil-quality physical characteristics				
soil-carbon content (% of soil weight)	2.6	1.5	0.8	15.2
soil-clay content (% of soil weight)	28.3	3.9	15.5	44.9
Observations	443			

Table 2: Robust regression coefficient estimates of different yield components under variable returns to scale for a multi-output technology

	Coefficient	SE	T statistic	P Value
Dependent variable: L				
Constant	17.359	0.080	218.092	0.000
l	-7.067	0.056	-126.498	0.000
l^2	0.889	0.006	139.311	0.000
Dependent variable: L*Q				
Constant	14.187	0.240	59.045	0.000
l	-5.036	0.169	-29.864	0.000
l^2	0.497	0.019	25.788	0.000
Dependent variable: L*I				
Constant	6.580	0.234	28.111	0.000
l	-1.311	0.164	-7.979	0.000
l^2	0.070	0.019	3.739	0.000
Dependent variable: L*Q*I				
Constant	5.032	0.209	24.076	0.000
l	-0.882	0.147	-6.010	0.000
l^2	0.044	0.017	2.632	0.009
Dependent variable: L*K				
Constant	14.664	0.190	77.129	0.000
l	-5.745	0.133	-43.050	0.000
l^2	0.631	0.015	41.370	0.000
Dependent variable: L*Q*K				
Constant	11.085	0.260	42.674	0.000
l	-3.897	0.182	-21.373	0.000
l^2	0.386	0.021	18.515	0.000
Dependent variable: L*Q*K*I				
Constant	3.944	0.175	22.526	0.000
l	-0.680	0.123	-5.535	0.000
l^2	0.035	0.014	2.489	0.013
Dependent variable: L*Q*K*I*Y				
Constant	2.010	0.117	17.247	0.000
l	-0.165	0.082	-2.020	0.044
l^2	0.006	0.009	0.613	0.540
Dependent variable: L*Q*K*I*Y				
Constant	1.955	0.089	21.883	0.000
l	-0.119	0.041	-2.887	0.004
Dependent variable: Efficiency				
Constant	0.914	0.028	32.099	0.000
l	-0.153	0.020	-7.488	0.000
l^2	0.021	0.002	9.033	0.000
Dependent variable: Yield				
Constant	1014.582	38.206	26.556	0.000
l	-314.561	27.450	-11.460	0.000
l^2	40.939	3.058	13.386	0.000

References

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.

Carletto, C., S. Savastano, and A. Zezza (2011, December). Fact or artefact : the impact of measurement errors on the farm size - productivity relationship. Policy Research Working Paper Series 5908, The World Bank.

Chayanov, A. (1926). *The theory of peasant economy.* The University of Wisconsin Press published in 1986.

de Boer, P. (2009). Generalized Fisher index or Siegel Shapley decomposition? *Energy Economics 31*(5), 810 – 814.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *The American Economic Review 84*(1), pp. 66–83.

Fisher, I. (1922). *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability.* Houghton Mifflin.

Gini, C. (1937). Methods of eliminating the influence of several groups of factors. *Econometrica 5*(1), pp. 56–73.

Husbands, K., T. Konyango, and T. Pinckney (1996). Education and agricultural productivity in Kenya. *In "The evaluation of public expenditure in Africa", eds. H.J. Bruton, C. Hill, World Bank Publication, Economic Development Institute, Washington, D.C.*, 111–138.

Jorgenson, D. W. and Z. Griliches (1967). The explanation of productivity change. *The Review of Economic Studies 34*(3), 249–283.

Kumar, S. and R. R. Russell (2002, JUN). Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence. *American Economic Review 92*(3), 527–548. PT: J.