



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

“Sources of measured agricultural yield difference”

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**

Preliminary Draft for discussion only. Please do not cite without the permission of the author.

Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

Copyright 2012 by Simone Pieralli. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Sources of measured agricultural yield difference

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.

Preliminary Draft for discussion only. Please do not cite without the permission of the author.

June 4, 2012

Abstract

We decompose yield difference relative to a reference level into components attributable to (1) efficiency difference, and movements along the frontier due to (2) land quality, to (3) land size, and to (4) other inputs. The production frontier is built using nonparametric methods requiring no specification of the functional form of the technology. We analyze the contributions to yield relative to a reference unit in terms of the quadripartite decomposition finding that results depend on the choice of the unit of reference. If the reference unit is chosen to be the mean, land size contributions are found to be negatively correlated to yield with usual finite moments regression methods. Also nonparametric correlation confirms the negative sign of the relationship. If the reference unit is chosen to be the median instead, land size contributions are found to be negatively correlated to yield with usual finite moments regression methods. But nonparametric correlation is not statistically significant because many farmers have no contribution to production difference from their different land sizes. Integrated squared density difference tests show in both cases efficiency has a major role in shaping the distribution.

Key words: inverse land size-productivity relationship, productivity decomposition, efficiency, yield, Kenya

JEL classification: D20, C14, C43

Introduction

The introduction of new methodologies and new technologies has led to a sustained interest in the inverse farm size-productivity relationship. Since Chayanov (1926), the inverse rela-

tionship between land size and yield, as a crude measure of productivity, has been the topic of an extensive debate. Unlike older studies, recent empirical literature has revisited the long-standing relationship, focusing especially on the introduction of new data, available thanks to technologies applied innovatively to this old problem.

While recent studies have considerably improved our understanding of the problem, they have also revived the controversy by neglecting the importance of very critical agricultural physical factors such as land quality, even after including newly available data. After Chayanov (1926) who noticed it for the first time, empirical economists emphasized the importance of other factors, such as incomplete and imperfect markets, measurement error and omitted soil quality as the culprit of this relationship in developing countries settings. The latest contributions find little role for omitted soil quality (Barrett, Bellemare, and Hou 2010), and no role for measurement error (Carletto, Savastano, and Zezza 2011), while confirming a strong negative relationship between land size and yield.

Much of the existing empirical literature is summarized in the recent contributions (e.g. Barrett, Bellemare, and Hou (2010) and Carletto, Savastano, and Zezza (2011)). This literature has focused on explaining the relationship with new data but with available methods. One of the first explanations of this relationship in the past was the presence of imperfect labor markets. These imperfections caused, following this explanation, an over-usage of labor in small-holder fields making them appear more productive. Data restrictions have instead caused to formulate the omitted soil quality explanation and the size measurement error explanation. The first indicates soil quality as an omitted variable negatively correlated with land size. By virtue of regression methods, this could provoke the inverse

relationship. The size measurement error explanation instead considers that the inverse relationship could be caused by measurement error attenuation bias (Lamb 2003). These explanations have sometimes caused the relationship to disappear but not unanimously.

Very recently the focus has been on introducing and using newly available data for explaining this old relationship with available methods. The availability of new satellite measurements for plot sizes have allowed Carletto, Savastano, and Zezza (2011) to show a stronger inverse relationship when taking into account the measurement error of plot size among Ugandan households with regression methods. The availability of new quantitative land characteristics measurements have allowed Barrett, Bellemare, and Hou (2010) to show the insignificance of land quality in explaining the inverse relationship with usual regression methods.

Modern data do not explain anything of this relationship when used with usual regression methods, with common functional form assumptions. The goal of this study is to separate spurious empirical relationships from truly significant ones. This inverse relationship is an important topic in development economics. Its truth or falsity has policy implications.

This issue is very important presently because the international agenda is mostly focused on smallholder African agriculture productivity. Smaller farms could be considered the most productive and efficient production units for a better development if the inverse relationship were confirmed using also more assumptions-free methods. If instead the relationship is proved just a result of applied statistical methods, other policies such as land consolidation or formation of aggregate groups of farmers should be investigated more closely.

This paper addresses this question directly. In particular, we decompose an index of yield, a crude measure of productivity difference, into components attributable to (1) efficiency difference, and movements along the frontier due to (2) land quality, to (3) land size, and to (4) other inputs. The first component reflects movements toward (or away from) the frontier as farmers adopt best practice technologies and reduce (or exacerbate) technical inefficiency. The second component reflects movements along the frontier due to land quality, keeping land size, and other inputs fixed. The third component reflects movements along the frontier due to land size, keeping land quality, and other inputs fixed. Finally, the fourth component measures movements along the frontier due to all other inputs, keeping land quality, and land size unchanged. This decomposition sheds light on which of these components is more important in explaining the difference in yield index.

The production frontier is constructed using nonparametric methods requiring no specification of the functional form for the technology and without specific assumptions on returns to scale or on market efficiency. We calculate the above four components of yield difference for a sample of Kenyan households.

These methods, already used by Färe et al. (1994) and Kumar and Russell (2002) to analyze changes in macroeconomic context, are here generalized. Moreover, these methods are here applied to Kenyan households to shed light on a long-standing issue in development and agricultural economics by innovating the methodology applied to already available data. This is done in the hope of obtaining more general results. Any procedure that produces estimates or approximations to the technology frontier (econometric estimation

or Data Envelopment Analysis, DEA, approximation) could be used to obtain empirical versions of each of the theoretical measures developed in this study.

Studies based on standard linear regression methods focusing on the first and second moments of distributions have not provided until now satisfactory explanations of the inverse land size-productivity relationship. In the present case, for example, a crude standard regression of logarithm of yield on logarithm of land size provides an estimate of elasticity of -0.236 significant at 1% level. This means that yield decreases by around a fourth of each percentage increase in land size. Even when relaxing the parametric assumptions the results still show significant negative correlation around the mean of the logarithm of size. This can be seen in the nonparametric regression plot shown in figure 1. But we also see from the nonparametric regression that there are parts of the distribution that are not well described by this simple regression analysis. It is important to understand what is hidden inside the data around the mean. Moreover, we want to relax restrictive assumptions on form of production functions usually embedded in linear regression methods. For these reasons in this study we decide to adopt the DEA methodology.

Although the methods used in the analysis here are quite simple, it provides somewhat fundamentally different results than usually obtained with regression methods: (1) Results are shown to be relative to the reference unit considered. (2) If measured around the median (or the mean), while with usual regression methods there is a negative significant relationship between size and yield, in the present study there is substantial evidence of no important negative contributions of land size to difference in yield when considered with the proposed methods. (3) A lot of the difference in yield is due to efficiency differences.

A caveat on the results shown in this study is granted now. The measures of productivity difference developed here are measures developed for one cross-section of data. This means that there is no time dimension in the results; this is so because there are no land quality panel households data available in the context of developing world countries. Once these data were to become available a generalized version of this study would be in order. This would allow a less arbitrary and more natural choice of reference unit. For the moment we leave this for future research. Moreover these results are done only for one output so no consideration can be given to strategic behavior of the farmers. The methodology is easily generalizable to multiple outputs case. This could help in comparing better the results of this study to previous studies which might be, in this respect, more comprehensive than this.

We should also say that the analysis, because of the index number theory methods used, is not intended to provide causal explanations of the facts observed. It only is a generalized growth-accounting exercise applied to shed light on an important problem in a different field of analysis. The methodology is discussed next. Then data are presented. Finally, empirical results and conclusions are shown.

Methodology

If the inverse land size-productivity relationship reflects physical reality, land could potentially be more productive if large-scale operations were broken into smaller units. Hence the inverse relationship is often offered as an economic argument for land redistribution programs. The inverse land size-productivity relationship is often analyzed (e.g. Assunção and Braidó (2007) and Bardhan (1973)) assuming a Cobb-Douglas production function

with constant returns to scale. But constant returns to scale implies that a proportional increase in all inputs leads to a corresponding proportional increase in all outputs. This is not necessarily true a priori. The use of a production function implies all agents operate in a technically efficient manner. But there are possibly many cases in which incentives are such that agents produce inefficient bundles. In addition, the use of a Cobb-Douglas functional form implies a unitary elasticity of substitution that can mask legitimate changes in the degree of input substitutability as allocative inefficiency.

The inverse land size-productivity problem is often studied by regressing yield (or the natural logarithm of yield) on land size (or on the natural logarithm of land size) while conditioning on other characteristics (among which input factors and, seldom, land quality characteristics). In particular, conditioning linearly on land quality characteristics implies that, in the evaluation of the performance of the farmer, substitution possibilities are not considered, even among land quality characteristics, and the inverse relationship is calculated as if these characteristics were given.

My research addresses the possibility that relaxing too restrictive assumptions and accounting quantitatively for land quality characteristics and land size could change the results obtained from more conventional regression methods on the inverse relationship. The typical measure of productivity used in the empirical literature on land size and productivity is yield. Yield is easily recognized as a partial productivity measure. Therefore, once yield is converted into index form by comparing it to some base-level yield it can be analyzed exactly as other partial productivity measures have been analyzed (Kumar and Russell 2002).

A simple method rooted in the theory of index numbers and productivity accounting can be used to isolate the contribution of different factors to differences in measured productivity.

DEA is the methodology used in this article because it allows to characterize the technology with minimal parametric assumptions (i.e. only piecewise linearity).

Let $y \in \mathbb{R}_+$ and $\mathbf{x} \in \mathbb{R}_+^U$ denote output and inputs respectively and let $l \in \mathbb{R}_+$ and $q \in \mathbb{R}_+$ denote land area devoted to production and land quality respectively. The following is developed in the case of one output to follow the empirical literature on the inverse yield-size relationship but could be extended to a multi-output case. The technology set T_t , where t represents time, is defined:

(1)

$$T_t = \{(\mathbf{x}_t, l_t, q_t, y_t) \in \mathbb{R}_+^{U+1+1+1} : (\mathbf{x}_t, l_t, q_t) \text{ can be used by households to produce } y_t \text{ at time } t\}$$

T_t is assumed to satisfy:

A.1: $(\mathbf{x}_t, l_t, q_t, y_t) \notin T_t$ if $\mathbf{x}_t = \mathbf{0}, l_t = 0, q_t = 0, y_t > 0$.

A.2: If $(\mathbf{x}_{1t}, l_{1t}, q_{1t}, y_{1t}) \in T_t$ and $(\mathbf{x}_{2t}, l_{2t}, q_{2t}, y_{2t}) \in T_t$, then $\forall \alpha \in [0, 1] : (\mathbf{x}_t, l_t, q_t, y_t) = \alpha(\mathbf{x}_{1t}, l_{1t}, q_{1t}, y_{1t}) + (1 - \alpha)(\mathbf{x}_{2t}, l_{2t}, q_{2t}, y_{2t}) \in T_t$.

A.3: T_t is assumed closed $\forall (\mathbf{x}_t, l_t, q_t, y_t) \in \mathbb{R}_+^{U+1+1+1}$.

A.4: T_t is bounded $\forall (\mathbf{x}_t, l_t, q_t) \in \mathbb{R}_+^{U+1+1}$.

A.5: Outputs are strongly disposable: if $y_t \in \mathbb{R}_+ \in T_t \subseteq \mathbb{R}_+^{U+1+1+1}$ then $0 \leq y'_t \leq y_t \Rightarrow y'_t \in T_t$.

A.6: Inputs (\mathbf{x}_t, l_t, q_t) are strongly disposable: if $(\mathbf{x}_t, l_t, q_t) \in \mathbb{R}_+^{U+1+1} \in T_t \subseteq \mathbb{R}_+^{U+1+1+1}$

then $(\mathbf{x}'_t, l'_t, q'_t) \succeq (\mathbf{x}_t, l_t, q_t) \Rightarrow (\mathbf{x}'_t, l'_t, q'_t, y_t) \in T_t$

In the single output case, the Farrell output efficiency score is defined:

$$(2) \quad E(\mathbf{x}_t, l_t, q_t, y_t) = \max \{e_t \in \mathbb{R}_+ : (\mathbf{x}_t, l_t, q_t, e_t y_t) \in T_t\}$$

if $\exists e_t$ s.t. $(\mathbf{x}_t, l_t, q_t, e_t y_t) \in T_t$ and $+\infty$ otherwise. By **A.5**

$$(3) \quad E(\mathbf{x}_t, l_t, q_t, y_t) \geq 1 \Leftrightarrow (\mathbf{x}_t, l_t, q_t, e_t y_t) \in T_t$$

so that $E(\mathbf{x}_t, l_t, q_t, y_t)$ is a complete function representation of the technology. It is also positively homogeneous of degree minus one in y , that is,

$$(4) \quad E(\mathbf{x}_t, l_t, q_t, \mu y_t) = \mu^{-1} E(\mathbf{x}_t, l_t, q_t, y_t) \quad \mu > 0.$$

The method of decomposition of the factors affecting yield difference allows for non constant returns to scale. In doing so, it adapts and generalizes what has been done in productivity studies, for example, by Henderson and Russell (2005) and by Kumar and Russell (2002). But especially it allows for a more general framework in which to study the inverse farm size-relationship. This is developed for one period in time only because we have data on land quality for only one period. But it could be easily generalized to include a technological change component.

We recognize a yield index as a ratio of partial productivity measures. A yield index for one unit (in the following, unit 1) can be defined relative to a base unit (in the following,

the base unit will be unit 0) as:

$$(5) \quad \frac{y_1/l_1}{y_0/l_0} = \frac{f(\mathbf{x}_1, l_1, q_1)/l_1 E(\mathbf{x}_0, l_0, q_0, y_0)}{f(\mathbf{x}_0, l_0, q_0)/l_0 E(\mathbf{x}_1, l_1, q_1, y_1)}$$

Using the fact that the Farrell output efficiency is positively linearly homogeneous of degree minus 1 in its output argument, we can rewrite this expression as:

$$(6) \quad \frac{y_1/l_1}{y_0/l_0} = \frac{f(\mathbf{x}_1, l_1, q_1) E(\mathbf{x}_0, l_0, q_0, y_0/l_0)}{f(\mathbf{x}_0, l_0, q_0) E(\mathbf{x}_1, l_1, q_1, y_1/l_1)}$$

The second right hand term can be considered a usual relative efficiency index measured with inefficiency measures. The rest of the treatment here will concentrate on the first right hand term

$$(7) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)}$$

which can be recognized as a ratio of efficient points on the production function, without necessity of assuming specific returns to scale, nor functional forms a priori.

It is possible to obtain different decompositions of (7). To illustrate, first multiply and divide by $f(\mathbf{x}_1, l_1, q_0)f(\mathbf{x}_1, l_0, q_0)$ to obtain

$$(8) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_1, l_1, q_0)} \frac{f(\mathbf{x}_1, l_1, q_0)}{f(\mathbf{x}_1, l_0, q_0)} \frac{f(\mathbf{x}_1, l_0, q_0)}{f(\mathbf{x}_0, l_0, q_0)}$$

Each of these three terms on the right-hand side:

$$\frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_1, l_1, q_0)},$$

$$\frac{f(\mathbf{x}_1, l_1, q_0)}{f(\mathbf{x}_1, l_0, q_0)},$$

and

$$\frac{f(\mathbf{x}_1, l_0, q_0)}{f(\mathbf{x}_0, l_0, q_0)}$$

are legitimate index numbers. That is, only one argument changes in every ratio and every ratio measures relative changes due to that argument. In particular the first of the right hand terms represents the vertical distance between the two frontier points given by a change in soil quality. The second of the right hand terms represents instead a distance between two frontier points given by a change in land size. The last of the right hand terms represents instead a change in the frontier points given by a change in the inputs other than land quality and land size.

But it is also possible to decompose (7) by multiplying and dividing by $f(\mathbf{x}_0, l_1, q_1)f(\mathbf{x}_0, l_0, q_1)$.

This obtains:

$$(9) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_1, q_1)} \frac{f(\mathbf{x}_0, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_1)} \frac{f(\mathbf{x}_0, l_0, q_1)}{f(\mathbf{x}_0, l_0, q_0)}$$

Also in this case every term represents a proper index. In this case the first term is associated with a change in inputs other than land quality and land size, the second term is associated with a change in land size and the last term is instead associated with a change in land quality. We can see that the corresponding terms of the decompositions are not the same.

For example the land size component is not the same in the two decompositions:

$$(10) \quad \frac{f(\mathbf{x}_1, l_1, q_0)}{f(\mathbf{x}_1, l_0, q_0)} \neq \frac{f(\mathbf{x}_0, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_1)}$$

More generally, it is possible to show that

$$(11) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)}$$

can be decomposed in the following equivalent but different decompositions, in addition to the previous two:

$$(12) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_1, l_0, q_1)} \frac{f(\mathbf{x}_1, l_0, q_1)}{f(\mathbf{x}_1, l_0, q_0)} \frac{f(\mathbf{x}_1, l_0, q_0)}{f(\mathbf{x}_0, l_0, q_0)}$$

$$(13) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_1, q_1)} \frac{f(\mathbf{x}_0, l_1, q_1)}{f(\mathbf{x}_0, l_1, q_0)} \frac{f(\mathbf{x}_0, l_1, q_0)}{f(\mathbf{x}_0, l_0, q_0)}$$

$$(14) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_1, l_1, q_0)} \frac{f(\mathbf{x}_1, l_1, q_0)}{f(\mathbf{x}_0, l_1, q_0)} \frac{f(\mathbf{x}_0, l_1, q_0)}{f(\mathbf{x}_0, l_0, q_0)}$$

$$(15) \quad \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_1, l_0, q_1)} \frac{f(\mathbf{x}_1, l_0, q_1)}{f(\mathbf{x}_0, l_0, q_1)} \frac{f(\mathbf{x}_0, l_0, q_1)}{f(\mathbf{x}_0, l_0, q_0)}$$

Our proposed solution to resolve the ambiguity in the method of decomposition is to pursue the path followed by Fisher in creating his ideal index and by many others since. That is we take the geometric average of the different decompositions to obtain:

$$\frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = \left(\frac{f(\mathbf{x}_1, l_0, q_0)}{f(\mathbf{x}_0, l_0, q_0)} \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_1, q_1)} \frac{f(\mathbf{x}_1, l_0, q_0)}{f(\mathbf{x}_0, l_0, q_0)} \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_1, q_1)} \frac{f(\mathbf{x}_1, l_1, q_0)}{f(\mathbf{x}_0, l_1, q_0)} \frac{f(\mathbf{x}_1, l_0, q_1)}{f(\mathbf{x}_0, l_0, q_1)} \right)^{1/6}$$

$$\left(\frac{f(\mathbf{x}_1, l_1, q_0) f(\mathbf{x}_0, l_1, q_1) f(\mathbf{x}_1, l_1, q_1) f(\mathbf{x}_0, l_1, q_0) f(\mathbf{x}_0, l_1, q_0) f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_1, l_0, q_0) f(\mathbf{x}_0, l_0, q_1) f(\mathbf{x}_1, l_0, q_1) f(\mathbf{x}_0, l_0, q_0) f(\mathbf{x}_0, l_0, q_0) f(\mathbf{x}_1, l_0, q_1)} \right)^{1/6}$$

$$(16) \left(\frac{f(\mathbf{x}_1, l_1, q_1) f(\mathbf{x}_0, l_0, q_1) f(\mathbf{x}_1, l_0, q_1) f(\mathbf{x}_0, l_1, q_1) f(\mathbf{x}_1, l_1, q_1) f(\mathbf{x}_0, l_0, q_1)}{f(\mathbf{x}_1, l_1, q_0) f(\mathbf{x}_0, l_0, q_0) f(\mathbf{x}_1, l_0, q_0) f(\mathbf{x}_0, l_1, q_0) f(\mathbf{x}_1, l_1, q_0) f(\mathbf{x}_0, l_0, q_0)} \right)^{1/6}$$

The first term is a term that considers effects of changes in the inputs (\mathbf{x}_i) keeping quality of land and land size fixed (INPUTS). The second term measures effects of changes in the frontier due to a change in land size (SIZE), and the last term measures the changes in the frontier due to a change in soil quality (QUAL). For later purposes, let us express the decomposition in compact form as follows:

$$(17) \frac{f(\mathbf{x}_1, l_1, q_1)}{f(\mathbf{x}_0, l_0, q_0)} = INPUTS * SIZE * QUAL$$

In this case yield index would be decomposed, following (16) and (6), into an efficiency component, a component due to land size, a component due to soil quality, and a component relative to the other inputs.

Because results are relative to the specific unit of reference (\mathbf{x}_0, l_0, q_0) and they change substantially we show in the results eight different possible scenarios. These different scenarios are useful to shed light on the possible importance of variables of interest such as size. The variables we vary in the scenarios are land quality, land size, and yield.

For each of these three variables we choose one unit with high and one with low value, resulting in eight possible scenarios. One scenario is taking as a reference unit a household with a low land size, low land quality, and low yield. Another scenario is taking as a reference unit a household with big land size, low land quality, and low yield and so on

varying land quality and yield. We also conduct the same calculations by taking the mean of all inputs and outputs as a reference unit. But we recognize that the distributions might be skewed.

In search of an ideal unit of reference, we then calculate the measures taking as a reference the median value of inputs and outputs. For the median values reference scenario, we elaborate the results in more detail to study where the inverse yield-size relationship comes from and which variables are actually most important in the decomposition of productivity differences. The importance of this methodology is in its generality. It can accommodate decompositions of productivity in components related to each different input, if so desired.

To test statistically for the significance of the contribution of different components to productivity difference we look at the linear regression of each component on the observed yield. This is to see if there is any significant relationship to emphasize. But usually applied regression methods are only looking at the behavior around the mean of the distribution. On the other hand, the nonparametric productivity measurements used in this study allow to characterize the position of each point, and not only of the average, with respect to the production frontier, and with respect to the reference unit. This is much more general than focusing only on the first moment characterization proper of usual linear regression methods. To exploit the potential of such richer characterization, nonparametric tests of equality of distributions are used to investigate the importance of relevant contributions to productivity difference (Li, Maasoumi, and Racine 2009). We prefer a nonparametric test of the integrated squared density difference to test for ‘any difference’ among distributions (Li, Maasoumi, and Racine 2009). This test allows to see which of the components isolated

has a decisive impact on shaping the observed yield distribution. Same test is repeated to show if there are any differences among different returns to scale assumptions.

In particular we can rewrite the decomposed productivity difference, using the decomposition in (17) as:

$$(18) \quad \frac{y_1/l_1}{y_0/l_0} = INPUTS * SIZE * QUAL * \frac{E(\mathbf{x}_0, l_0, q_0, y_0/l_0)}{E(\mathbf{x}_1, l_1, q_1, y_1/l_1)}$$

From this decomposition we can, following Kumar and Russell (2002) and adapting their intuition to our context, define different sets of counterfactual distributions. In particular we can rewrite

$$(19) \quad y_1/l_1 = y_0/l_0 * INPUTS * \frac{E(\mathbf{x}_0, l_0, q_0, y_0/l_0)}{E(\mathbf{x}_1, l_1, q_1, y_1/l_1)} * SIZE * QUAL$$

This can be considered as an alternative decomposition. If we multiply each component on the right-hand side we obtain exactly the observed yield distribution on the left-hand side. To isolate the significance of the contributions of inefficiency, land size, and land quality, we can start from a counterfactual distribution that would equal observed yield if there were no differences in land size, land quality, and inefficiency. In particular this can be written as:

$$(20) \quad y_1^I/l_1 = y_0/l_0 * INPUTS$$

We then successively introduce differences in inefficiency to have:

$$(21) \quad y_1^E/l_1 = y_0/l_0 * INPUTS * \frac{E(\mathbf{x}_0, l_0, q_0, y_0/l_0)}{E(\mathbf{x}_1, l_1, q_1, y_1/l_1)}$$

This is the counterfactual distribution of yields if we were to ignore differences in land quality and land size. Then we introduce differences in land size to have

$$(22) \quad y_1^L/l_1 = y_0/l_0 * INPUTS * \frac{E(\mathbf{x}_0, l_0, q_0, y_0/l_0)}{E(\mathbf{x}_1, l_1, q_1, y_1/l_1)} * SIZE$$

This is a counterfactual distribution of yields that does not take into account differences in land quality. Finally, we can introduce the land quality differences to obtain the previous decomposition (19).

But the last step, as the previous ones, could also be done in reverse order. In other words we could introduce the adjustment for land quality, first, to obtain:

$$(23) \quad y_1^Q/l_1 = y_0/l_0 * INPUTS * \frac{E(\mathbf{x}_0, l_0, q_0, y_0/l_0)}{E(\mathbf{x}_1, l_1, q_1, y_1/l_1)} * QUAL$$

And then introduce the land size component to obtain the original decomposition (19).

For later reference, we can also introduce the efficiency component at last. In other words, we can define

$$(24) \quad y_1^{QL}/l_1 = y_0/l_0 * INPUTS * QUAL * SIZE$$

Of course to arrive to this decomposition other counterfactual distributions can be obtained.

In particular we can introduce only the land quality component after (20):

$$(25) \quad y_1^{IQ}/l_1 = y_0/l_0 * INPUTS * QUAL$$

Or we can introduce only the land size component after (20) as follows:

$$(26) \quad y_1^{IL}/l_1 = y_0/l_0 * INPUTS * SIZE$$

At each subsequent step a test (Li, Maasoumi, and Racine 2009) will be done for equality of counterfactual and observed yield distributions to see when the two distributions cannot be statistically distinguished. In this way we can test which component contributes to shape the observed yield distribution.

For example, if the distribution y_1^L/l_1 is not found statistically different than y_1/l_1 , this would mean that land quality (in this case the last excluded factor) would not have a predominant effect in shaping the observed yield distribution. Moreover this would signal a significant impact of land size if, for instance, the previous test of equality of y_1^E/l_1 and y_1/l_1 were to be rejected in preceding comparisons.

Obviously the order of introduction of the subsequent differences is arbitrary. The underlying story behind does not seem to change from changing the order of introduction of different components.

Data

The data are drawn from a sample of households in 99 sub-locations in Kenya in early 2007 and they are relative to the long and short seasons 2005-2006. The survey is part of a panel named “Research on Poverty, Environment and Agricultural Technologies (REPEAT): Panel studies in Africa”. Survey data were obtained from the National Graduate Institute for Policy Studies (21st century Center of Excellence Program) in Japan. The cross-section sample analyzed in this study initially includes 718 households, of which only 579 units are available for calculation. Of these, data on land quality are available for 452 families.

Measured output, representative for agriculture, is harvested dry maize. Faithful to the development economics literature on the topic of yield productivity we choose to take into account only the case of dry maize production. Selectivity of farmers in maize production might be considered as an issue but all households available for estimation produce maize. So this seems less of a concern considering also that the sample has maintained its random sample properties even after elimination of some units due to errors in sampling.

The measured inputs used in maize production directly are seeds, land area, organic and inorganic fertilizers, family worked hours, cost of temporary hired workers, hours worked by permanent and shared workers, and milking cows. Other variables measuring inputs available for household production are number of hand hoes, ploughs, sickles, spray pumps. Table 1 shows input and output summary statistics of the households. Data on physical characteristics of land for the largest maize plot for each household are available for mid-2003. The analysis focuses on two measures that are stable over time: soil carbon and soil clay content. These two variables are aggregated into a ordinal land quality measure, following the methodology developed in Pieralli (2011). In creating the land quality indicator, we will vary the percentiles of reference of the inputs and outputs to see how the results change. Table 1 shows summary statistics of these soil properties but more soil properties could be aggregated into a land quality indicator, if so desired.

Results

In development economics, it is an empirical regularity to encounter a negative relationship between yield and land size, either of the farm or of the plot farmed. Measurement errors seem to reinforce this negative relationship (Carletto, Savastano, and Zezza 2011).

Indeed, even in the present case, common regression methods and nonparametric regression methods show a negative and strongly significant relationship. This can be seen from figure 1. The figure represents the nonparametric regression of logarithm of yield on the logarithm of land size. The nonparametric regression (middle) line is contoured by the 95% confidence intervals to show significance. This figure shows a significant negative relationship between logarithm of land size and logarithm of yield of dry maize per acre. This conceptually means that the unconditional elasticity of yield with respect to land size is negative. This kind of evidence is usually brought forward to signal at first glance a significant negative relationship between yield and size per acre (Barrett, Bellemare, and Hou 2010). The nonparametric regression is significant, at least, around the average. We estimated the elasticity also with parametric methods. The parametric estimate of the elasticity at the average size is -0.236 and it is significant at less than 1% level. This means that per acre production decreases on average almost by one fourth of the percentage increase in acreage. Results are robust also when including inputs and remain qualitatively the same when including also a land quality measure. This is usually taken to signal the presence of a negative relationship between size and yield and the apparent insignificance of the land quality variable.

The problem is that this estimated relationship assumes a specific functional form and studies the relationship (at least usually in parametric cases) around the mean value. Moreover, usually, production efficiency and constant returns to scale in production are assumed. These assumptions are very stringent and possibly the cause of how the estimate results.

In this paper we relax these assumptions to see if the relationship persists. We consider a flexible nonparametric productivity accounting method, separating explicitly the efficiency component, and the influence of land size, land quality, and other inputs. In this way we do not assume a specific functional form, nor efficiency of production, or constance of returns to scale.

The productivity accounting method described in the preceding methodology section produces measures that are relative to the unit of reference considered. In the following we show how results change by changing the unit of reference. We do this by means of graphing the four percentage components of productivity against observed yield. As said in the methodology section we consider eight different cases to show how estimates change for a high and a low value of three characteristics: land quality, land size, and yield. We choose the units using the level of land quality calculated under variable returns to scale. Because the ranks can change, especially between constant returns and the other assumptions, we focus on studying the variable returns to scale as the most general assumption. The exercise can be replicated under different returns to scale and for different characteristics to see how measures change.

To place the units of reference in context of the present sample, we can show the different values on the cumulative distribution functions of yield, land size, and land quality with empirical cumulative distribution functions of the single variables and with joint bivariate histograms. While we tried hard to match this simple theoretical idea with finding the right units of reference for the analysis, we had to accommodate to approximately high and approximately low values to match these ideas with real units. In particular in figure 2 we

can see the empirical cumulative distribution function of land size. We plotted on the graph lines in correspondence of 0.55 acres, 2.25 acres, 2.65 acres, and 4 acres. These lines are in correspondence of values from four units we have chosen as reference units for the analysis and that can help to see also where the other four units used for reference are placed. In figure 3 we show the empirical cumulative distribution function of yield. Corresponding to the previous four values are, respectively, the lines at 981 Kg acre⁻¹, at 240 Kg acre⁻¹, at 135 Kg acre⁻¹, and at 787.5 Kg acre⁻¹. Each of these households has an associated land quality. In particular, in figure 4 we report lines corresponding to previous values at 0.7199, at 0.99, at 0.42, and at 0.35. These four cases allow seeing the eight possibilities we designed for measurement. In particular, the first unit among the four will be the reference unit representative for little size, high yield, and relatively low land quality. The second unit will be one with a relatively big land size, a low yield, and a very high land quality. The third unit instead is an example of a unit with big land size, low yield, and low land quality. Finally the fourth unit is a unit with very big size, high yield, and very low land quality.

The other four units have respectively 0.5 acres of land size, yield of 270 Kg acre⁻¹, and 0.7453 of land quality index, 0.5 acres of land size, yield of 720 Kg acre⁻¹, and 0.95 of land quality index, 2.5 acres of land size, yield of 972 Kg acre⁻¹, and 0.98 of land quality index, and finally 0.6 acres of land size, yield of 250 Kg acre⁻¹, and 0.96 of land quality index. It is possible to visualize the position of these reference units approximately on the joint histogram of land quality and land area in figure 5, land area and yield in figure 6, and land quality and yield in figure 7.

In each of the eight cases we repeated the calculations of the quadripartite decomposition, for understanding what is the relation between yield and the four contributions. Figures from 8 to 15 plot the percentage contributions measures (dots) against the observed yield. Graphs also report a usual regression line (solid) for which the legend says if the relationship is significant or not at the 95% confidence level and a dashed line representing a smoothed Gaussian kernel. The kernel shows a smoothed local regression line.

Consider first the case of reference unit with small land size, low yield, and low land quality. This is shown in figure 8. We expect positive percentages of land size, land quality, and other inputs in contributing to the yield difference between other units and the reference unit. The regression lines show these significant relationships. It is not significant the contribution of efficiency to yield difference in this case. At first sight, these regression lines seem to suggest a completely opposite relationship between land size contribution and observed yield than usually seen in empirical applications. This result changes if we take as a reference unit a household with same characteristics (low yield and low land quality) but with a big land size as we do in figure 9. In this case we see that land size contribution is not correlated with observed yield almost at all, while contributions of land quality and efficiency are positively correlated to observed yield. If we do the same analysis passing from a little land size to a big land size but for a household with high yield as in figures 10 and 11, we can see the same trends in the changes of relationship between land size contributions and land quality. Land size contributions to yield difference are moderately positively significantly correlated to yield in the case of little land size but are negatively significantly correlated to yield in the case of big land size reference unit.

We then consider the cases when the household units of reference have high land quality. In particular, in figures 12 and 13 we consider the cases when the household reference unit has low yield and high land quality, passing from little land size in figure 12 to big land size in figure 13. We observe here the same relationship in the change of land size. In particular, in these cases, because the land quality of reference is high, most other households have negative contributions of land quality to yield difference. Moreover, these contributions are negatively correlated to observed yield if we follow the regression line plotted. But if we look empirically at the dots representing the different contributions we can see that the most negative contributions are for smaller yields. This would mean that actually the households more affected by a difference in land quality compared to a high land quality are the households with lower yields. This would open another branch of research that is not strictly the focus of this study but for sure of critical importance to assess vulnerability of households.

If we then consider the graph of the contributions of land size to yield difference, we can see that increasing size of the land makes insignificant the positively sloped significant regression line. So we go from evidence against most literature on the topic (positive correlation of percentage contributions of land size to yield) to a negative significant or insignificant relationship. In other words, going from a small to a big land size reference unit any relationship between land size contributions and yield, if significant, becomes negative or disappears.

We finally consider the case of a reference unit with high yield, high land quality and we move from a small land size in figure 14 to a big land size in figure 15. The same kind

of positive relationship when considering a small land size unit of reference in figure 14 is inverted in a negative relationship when considering a big land size unit of reference in figure 15.

In some cases, finding a negative relationship of land size contributions with yield would seem to reassure the empirical studies on the topic. But these estimates are relative to a specific reference unit and change substantially. Moreover, the regression line seems particularly not informative of the variation among land size percentage contributions to yield difference.

Considering that these estimates could be misjudged depending on the reference unit used, we also produce the same graphs taking into account as a reference unit the average unit with average values of inputs and outputs. This case should, in principle, be a more meaningful balanced case than the extreme cases considered until now. Graphs to illustrate the average case are reproduced in figures 16, 17, and 18 for constant, non-increasing, and variable returns to scale respectively. These figures suggest that, independently of returns to scale assumptions, there is a significant negative linear relationship, on average, between contributions of land size to yield difference and observed yield. This negative slope is essentially what has led many to argue for the inverse land size-yield relationship. This result is consistent across returns to scale. Moreover, the figures show an insignificant relationship between contributions of land quality to yield difference and observed yield in the south-west quadrant.

Because in usual empirical cases, as in this one, the median is a much more informative statistic given the skewness of some of the distributions of the variables, we repeat the

calculations taking into account as a reference unit the median values of inputs and outputs. We can see in tables from 2 to 13 summary statistics of the components of the quadripartite decomposition of yield. These calculations are done in correspondence of land quality measures calculated for different percentiles of reference levels of inputs and outputs as done in Pieralli (2011) and as adapted to the present case of a single product.

To facilitate the interpretation we report the reciprocal of the efficiency index. In this way we can see that the units were, on average, 45 to 60% as inefficient as the unit of reference. This is true under all returns to scale even though variable returns have slightly lower averages. Land quality contribution is between -10% and -1% for the constant and between -15% and -1% for the non-increasing returns to scale on average. Land quality, under variable returns, has instead a much higher negative contribution to yield difference on average from around -20% to -6%. Land size, on average, has a relatively small effect under constant returns to scale and it is increasing in importance with increasing the percentiles of reference of the land quality measure from around -2% to -3%. Same trend with similar figures is evidenced under non-increasing (-2% to -3%) and variable returns to scale (-2% to -5%). Other inputs instead account for a negative mean contribution of around 50%. In looking at these statistics we have to say that when the considered land quality measure is calculated at lower percentiles of reference of inputs and outputs the convergence of units presents more problems. This is why we concentrate the analysis of the results using the land quality measure originating from the highest percentile of reference of other inputs and outputs, i.e. the measures summarized in the last line of each of the tables. The results shown for the highest percentile of reference level of inputs and outputs (on which the

treatment is concentrated here) are summarized from the results of 443 units because only 443 units have a strictly positive yield in this sample. The results at the highest percentile of reference level for constant and non-increasing returns to scale are from these 443 units. The results at the highest percentile of reference level for variable returns to scale are instead summarized from 403 families. This is because the land quality component of the productivity accounting measures proposed seems very sensitive to jumps among counterfactual measures. The stability of the results across returns to scale assumptions reassures of the non arbitrariness of these results. Moreover we also repeated these calculations and the tests without the 9 units with zero yield and results are qualitatively the same, if not stronger.

As before, three graphs are used to illustrate the results of the calculations for the unit of reference with median values. Figures 19, 20, and 21 present the results in the same format as previously for constant, non-increasing, and variable returns to scale respectively. Across returns to scale assumptions, there is a significantly negative linear relationship between contributions of land size to yield difference and observed yield. This shows that for smaller yields differences in land sizes matter most for productivity differences. But results vary slightly for the land quality component across different returns to scale assumptions. While in the constant returns to scale case the relationship between land quality contributions to yield difference and observed yield is significantly negative, in both non-increasing and variable returns to scale the significance of this relationship disappears.

This is the reading that we could have if we wanted to stop at a characterization of the average behavior of the measures. We could emphasize that in the beginning part of

north-east graph of figure 21, under most general returns to scale assumption, there are many contributions that are positive and then followed by negative contributions at higher levels of yields. But this would leave out a lot of the variation around the observations. In particular, we can see that observations, especially around the beginning of the distribution, are very spread, both on the negative and on the positive side, signaling the inadequacy of first moment parametric comparisons (Kumar and Russell 2002; Li, Maasoumi, and Racine 2009).

For example, a simple Spearman correlation coefficient between land contributions and observed yields is significantly negatively correlated for constant (-0.08) and non-increasing returns to scale (-0.079) only at 10% level. But in the case of variables returns to scale the test is not significant (-0.058 with p-value of 0.2443)¹.

The spread of the observations on the graphs, together with these tests, show that the characterization of the results by only looking at a first moment parametrically might be misleading. So we check more in detail what is hidden around the average in the estimates of land quality and land size contributions. To check which of the three components of the production function has a major role in a relative sense, we calculated the percentage average difference rates due to each of the three components of the decomposition of the production function: land quality, land size, and other inputs. While the percentage average difference rates due to other inputs is predominant in all returns to scale assumptions, we want to concentrate on land quality and land size. The percentages due to land quality and land size differ depending on the returns to scale assumptions. In particular, figure 22 plots mean percentage contribution rates in the last twenty percentiles of reference of inputs and

outputs when obtaining land quality measure. In the constant and non-increasing returns to scale case (the upper and middle graphs respectively) land size contributes on average more to the percentage yield difference. In the variable returns to scale instead (the lower graph) land quality contributes almost the double than land size at each given percentile. This shows that depending on the assumptions the importance of contribution rates can be different. It also shows that land size and land quality contributions can be relatively very important, and in different proportion for different assumptions.

We start a more in depth explanation of the results with median reference unit from the analysis of land quality contributions. We notice (figure 23) that land quality contributions at low levels of size (plotted on the horizontal axis) are more important and more variable for the variable returns to scale (the lower graph) than for constant (the upper graph), and non-increasing returns (the middle graph). This suggests that bigger sizes are less influenced by quality for production. We can see in figure 24 the land quality measure q plotted against size. This graph shows a non well defined relationship. But, especially for variable returns (the lower graph), some units with smaller size have a more variable land quality measure. This could play a role in showing a bigger contribution of land quality to yield difference.

For small farmers of very small sizes the land size percentage contribution is negative across all returns to scale assumptions (see figure 25). This figure shows also that when increasing size the percentage to contribution is increasing systematically at least up to a certain size. This level of size up to which the increase is systematic is around 0.8 acres. Even though this could seem an artifact of the methods presented here, this increase is not

systematic along the whole distribution and it does not reflect in the portion higher than the median in the same way. Many farmers have land size contributions less negative on the left of the median and less positive on the right of the median level, respectively.

The negative contribution of land size for smaller farmers can be seen directly from the kernel smoothing distributions in figure 26. In particular we can notice, in aggregate, a shift of probability mass, even if not statistically significant, between the solid lines (before land size adjustment) and the dashed lines (yield distribution after land size adjustment). This shift is under the three returns to scale assumptions of the same direction: shifting mass to the left. This means more farmers have lower yield after the land size adjustment. These are the smaller farmers up to 0.8 acres but for the purpose of this study we want to see how this changes along the yield distribution if we disaggregate measures of land size contributions.

We divide non negative and negative land size contributions to see how land size contributions behave differently across the yield distribution. We replicate the same comparisons of kernel smoothing distributions between land size unadjusted and observed yields in figure 27. We can see that the presence of negative land size contributions to yield difference moves farmers towards lower yields (that are below the median yield level of 540 Kg acre^{-1} plotted as a vertical line). Figure 28 represents, in the same way, instead the non negative contributions of land size to yield difference. The non negative contributions to yield difference of land size move instead farmers towards higher yields (that are higher than the median level of yield).

To see if it is true that there is a differential impact of land size for higher yielding or lower yielding farmers, we divide precisely between the ones that are below the median or greater than or equal to the median level of 540 Kg acre⁻¹. Among lower yield farmers in figure 29 we see much less clear evidence of shifts of probability mass to the left signaling that not all lower yielding farmers are negatively affected by land size. In the same way when we analyze the higher yield farmers (as the median or higher) in figure 30 we see no particular evidence of shifts of probability mass to the right. A shift to the right would be expected if we were to think that higher yield farmers would be positively affected by land size. This counterintuitive result seems to be caused by the fact that many farmers that are both below and above the median level have a zero measured contribution of land size to yield difference. This is probably where this measurement differs from usual regression methods. While these farmers have differing land sizes, with the present methods, after taking into account land quality and efficiency explicitly, the contribution to yield difference of this difference in land sizes is null. To show this we isolate the farmers with zero contribution of land size to yield and we show their distribution of yields in figure 31. They are more on the lower side of the median. We show what is their distribution of land sizes in figure 32. More importantly we show how the sizes of these household farmers are distributed along the yield distribution in figure 33. In this figure we see that there is a negative relationship between land size and yield among the farmers that, in our measures, have no contribution of size to yield difference. This relationship is strongly significantly negative with nonparametric Spearman correlation tests (-0.3041 for constant returns, -0.3387 for non-increasing returns, and -0.3131 for variable returns all signif-

icant at less than 1% level). This is so because the efficiency index for these farmers is decreasing at the same time. This is shown by a strongly negative relationship between land size and the efficiency index. Nonparametric Spearman correlation tests (-0.4766 for constant returns, -0.5467 for non-increasing returns, and -0.6265 for variable returns) are all significant at less than 1% level. This negative relationship means that increasing land size increases the relative inefficiency of these families with respect to the median unit. There is not such a relationship at the level of the total sample.

This means that with usual methods their land measures are negatively correlated with yield and they are contributing to characterize the negative empirical regularity. But our methods instead predict that these sizes do not change the counterfactual production measures if you separate contributions of land quality, efficiency, and other inputs. No changes in the counterfactual production points are evidenced for these farmers if we take solely the effect of changes in land size into account as in our productivity accounting method. This fact means that these negative contributions to changes in the production measures are mistakenly thought to be caused by land size while instead are probably the outcome of inefficiency. This is evidence of the insignificance of the negative empirical relationship between land size contributions to yield difference and observed yield. But these are only descriptive methods. To discover if land size has actually a statistically significant effect we have to consider any difference on the whole distribution caused by the components of our quadripartite decomposition.

Exploiting the nonparametric nature of the productivity accounting measures used we explore the behavior of the distribution as a whole. In other words, we want to see in

our method which component relevant to our research shapes the observed yield distribution more significantly. We do this in a more general way than correlation tests and in a more statistical way than visual inspection of density distributions by studying any deviations of two distributions, focusing on an integrated squared density difference test by Li, Maasoumi, and Racine (2009). This smoothing test is shown to have advantages on the non-smoothing tests of difference of two distributions, such as the Kolmogorov-Smirnov test (Li, Maasoumi, and Racine 2009).

As shown in the methodological section we ask ourselves which component actually brings the distribution of yields from the unit of reference (in this case the median value) to the observed yields distribution. This is equivalent to a shift from the counterfactual distribution y_1^I/l_1 to the observed yields distribution as can be seen in figure 34. When the null hypothesis of the test by Li, Maasoumi, and Racine (2009) is not rejected anymore by successively testing counterfactual distributions against observed yields, we would then have found the component that plays the major role in shaping observed yield distribution. We can also study the importance of the adjustment by studying the probability value. We can then qualify the nature of the change brought by this component on the counterfactual yield distribution.

We show informally which component is most important comparing kernel smoothing density estimates of observed yields (dashed line) and of different counterfactual distributions of the different yield components. In particular, we present the counterfactual distributions y^L , y^Q , and y^{QL} as defined in the methodological section. We remind here that the counterfactual distributions y^L and y^Q are the observed yield distributions without the

final component of land quality and land size, respectively. So the difference between the counterfactual distributions and the actual yield distribution is the contribution of those characteristics to yield. Analogously, the difference between y^{QL} and the observed yield distribution is the effect of efficiency. Figure 35 shows the difference between y^Q and observed yields with empirical cumulative distributions. Figure 36 shows the difference between y^L and observed yields with empirical cumulative distributions. In figure 37 finally we show instead the effect of efficiency adjustment. This effect seems to be really strong. The efficiency effect seems to be the responsible of a big portion of the shift between y^I and observed yields distribution. Considering the particular, almost bimodal shape of y^I we could ask what are the characteristics of these families, and why this happens, but this is not the focus here and is left for future research.

Exploiting new developments in the statistical nonparametric theory we can then test formally the most important contributor to shaping the observed yield distribution by means of a test by Li, Maasoumi, and Racine (2009). From table 14 we can notice that introducing the other inputs component to obtain y^I does not make the distribution statistically equal to the observed yield distribution. This is true across different returns to scale. We see instead that the efficiency component makes the distributions statistically equal. This is particularly true for constant returns to scale and for non-increasing returns to scale where the p-values reach levels well above 0.25. It is not the same in the case of variable returns to scale where the p-value is only 0.089. This means that still efficiency does not make the counterfactual distribution statistically equal to observed yields, at least for a 10% level test. This also means that for the variable returns to scale case there is a stronger impact of the non

introduced components of land quality and land size, compared to the case of constant and non-increasing returns. This is also confirmed by the mean percentage contribution rates shown in figure 22. In particular, introducing land size component makes the p-value of the test jump to 0.947 in the case of variable returns. This high increase in p-value means that the effect of land size in the variable returns is relatively big compared to the other returns assumptions.

This also means, differentially that there is not much left for land quality to change the distribution if you include already land size. However if land quality were introduced instead of land size, the p-value would grow similarly up to 0.901. This means that, after efficiency, probably land size has a bigger impact than land quality in making the distributions equal. Different are instead the land size and land quality impacts in the constant and non-increasing returns case. The p-value becomes very high when including land size, but on the other hand it moves very little when introducing land quality after efficiency. The distributions move in a different way for different returns to scale. In particular, under variable returns the adjustment for land quality has a bigger impact than under the other assumptions. This means that land quality interacts differently especially for farmers who are on the increasing returns side of the production technology. For them in particular, land quality seems to be very important.

This analysis shows a strong decisive impact of efficiency in shaping the actual yield distribution, confirmed when studying qualitatively the kernel density estimates. This analysis also shows a differential impact of land quality in the variable returns to scale. Moreover, we show that land size assumes an importance in making the distributions equal only un-

der variable returns to scale and only when introduced directly after y_1^E/l_1 . If instead we introduce, after inputs, only land size or only land quality to create respectively y_1^{Ll}/l_1 and y_1^{Ql}/l_1 , there is no significant change. No change even when we include both in order to create y_1^{QL}/l_1 , as can be seen from table 14. This confirms once again the qualitative evidence of the importance of efficiency component shifting the distribution from the counterfactual y_1^I/l_1 to the distribution of observed yields as shown in figure 34.

We also test for difference among returns to scale of the productivity components. As we can see in table 15, there seems to be no particular difference among returns to scale apart for when introducing land size in y_1^L/l_1 and y_1^{Ll}/l_1 . This happens when testing equality of constant returns estimates with non-increasing (p-value of around 13%) and variable returns estimates (p-value less than 5%). The difference between returns to scale assumptions suggests that there are different ways land size interacts with farms on the upper and lower parts of the size distribution. This is where main differences among the two assumptions on scale play a role in a significant (variable returns case) or not so significant way (non-increasing returns case). The same qualitative results can be seen from studying the scenario of a mean value reference unit as can be seen in table 16.

Conclusions

The methods presented in this study allow knowing more on whether the long debated inverse land size-productivity relationship is true or false. Land quality is not taken usually into account quantitatively in the literature. When it is taken into account it is considered with very restrictive statistical assumptions. The hypothesis is that this, together with the other assumptions, among which production efficiency, cause the empirical regularity of

the inverse land size-yield relationship. Ascertaining if this relationship is true is done in this study by taking into account land quality, land size, and efficiency explicitly, in productivity terms.

In particular, we decompose a yield index measure into four parts. We purge out the inefficiency and decompose the efficient production function difference into three components. Components are relative to land size, land quality, and other inputs. Many studies, with few exceptions, found land size empirically negatively correlated to measured yield. In this study many assumptions usually done are taken away. First, no efficiency assumption is done on household dry maize production. Secondly, no specific functional form of the technology is assumed. Thirdly, no returns to scale assumption is done a priori.

The fact of not assuming efficiency allows us to study the decomposition of the efficient points on the production function and not of the observed yields. This allows us to purge what is included in yield measurement but caused by inefficiency. The second assumption of no specific technological functional form comes together with the first and allows not imposing specific properties among inputs and outputs a priori. The third assumption of returns to scale is shown to bear some consequences when analyzing the statistical significance of results but these are not central features of this study.

We replicate usual regression methods and find a significant negative relationship between land size and observed yield. We decompose yield difference into efficiency, land quality, land size, and other inputs components relative to specific units for different returns to scale. Results are done for eight different reference units. We choose reference units with low and high values of respectively land size, land quality, and yield. Regression

of percentage contributions of the quadripartite decomposition of observed yields shows different results depending on the reference unit.

Keeping the other characteristics the same and moving from a small land sized reference unit to a bigger farm transforms the relationship between land size contributions and yield from significantly positive to null or negatively sloped. Because we understand the relativity of these estimates, we repeat the calculations with mean values as reference unit. In this case the relationship between land size and yield is negative and significant. But we realize that, particularly in our case, mean statistics are less informative than medians. We repeat the calculations against the median values taken as a reference unit. But also in this case the simple regression relationship between land size contributions and yield is negatively significantly sloped lending the side to what has usually been suggested as the regular negative yield-size relationship.

This is true when estimating the relationship with linear regressions. With nonparametric measures of correlation the negative relationship is not confirmed. More importantly, the view of the graph of contributions and yield suggests that a simple measure based on the average does not render a proper characterization of the variation of the contributions.

To understand how this works we study the distribution of the yields and land size contributions more closely isolating the families who have positive, zero, and negative contributions. When taking into account land quality, efficiency, size, and other inputs separately, we find that some farmers have zero contributions from land size to productivity. These farmers show evidence of an inverse significant relationship between yield and their land sizes, and at the same time a strong negative relationship of land size and the efficiency

index. This is why many have registered the empirical relationship as a regularity when studying, in aggregate, parametric average behavior measures.

But these are just descriptive methods of the insignificance of the negative relationship between land size contributions to yield difference and observed yields. We follow the intent of exploring statistically more than the first and second moment of these distributions. We study any deviations among yield and several counterfactual distributions with the integrated squared density difference test by Li, Maasoumi, and Racine (2009). In this way we want to see statistically which component is the one that shapes the observed yields distribution. From the results we see that there is a critical role of efficiency.

Land size under variable returns to scale, when applied after efficiency in the counterfactual distributions, makes the counterfactual yield distribution equal to observed yields for any statistically relevant level. But we acknowledge that this is not the case if land size component is applied before efficiency. Efficiency still plays the major role in shaping the distribution of yields. This means that neither land size nor land quality explain the shape of the yield distribution, even though they make up a comparable and, in some cases significant, percentage of yield difference rates. The same results are derived for the mean reference case.

The productivity accounting measures developed in this study show that with usual regression methods the yield-size negative relationship is present even when taking into account efficiency and land quality. With more general nonparametric measures the correlation is not present because a part of farmers with negative yield-size relationship is shown to have no contribution to yield differences when measured against the median. There is

no critical role for land size in shaping the yield distribution also when we test the importance of contributions to yield difference statistically. Returns to scale assumptions do not interact critically with the importance land size and land quality have in shaping the yield distributions.

The last point we want to emphasize once more is the relativity of these measures. The findings on the shape of the distributions are robust to changes between the mean and the median reference choice. But the numeric values change when changing unit of reference. This means that a definitive answer to whether an inverse farm size-yield relationship is present or not could only come, once a less arbitrary reference unit choice would be available. But this is part of this contribution. We try to show that results from the methods proposed, and not solely, depend on the choice of the unit of reference.

Notes

¹This test, if done in the case of not including the 9 zero yield units, shows a significantly negative relationship only under constant returns, while it is insignificant under non-increasing and variable returns to scale. The Spearman correlation coefficient between land contributions and observed yields, in the case of taking the mean as a reference point, is instead significantly negatively correlated for constant (-0.1913), non-increasing (-0.1859), and variable returns to scale (-0.1807) at 1% level.

References

- Assunção, J., and L.H.B. Braido. 2007. "Testing Household-Specific Explanations for the Inverse Productivity Relationship." *American Journal of Agricultural Economics* 89:980–990.
- Bardhan, P.K. 1973. "Size, productivity, and returns to scale: An analysis of farm-level data in Indian agriculture." *Journal of Political Economy* 81:1370–86.
- Barrett, C.B., M.F. Bellemare, and J.Y. Hou. 2010. "Reconsidering Conventional Explanations of the Inverse Productivity-Size Relationship." *World Development* 38:88 – 97.
- Carletto, C., S. Savastano, and A. Zezza. 2011. "Fact or artefact : the impact of measurement errors on the farm size - productivity relationship." Policy Research Working Paper Series No. 5908, The World Bank, Dec.
- Chayanov, A. 1926. *The theory of peasant economy*. The University of Wisconsin Press published in 1986.
- 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:pp. 66–83.
- Henderson, D.J., and R.R. Russell. 2005. "Human capital and convergence: A production-frontier approach." *International Economic Review* 46:1167–1205.
- Kumar, S., and R.R. Russell. 2002. "Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence." *American Economic Review* 92:527–548, PT: J.

- Lamb, R.L. 2003. "Inverse productivity: land quality, labor markets, and measurement error." *Journal of Development Economics* 71:71 – 95.
- Li, Q., E. Maasoumi, and J.S. Racine. 2009. "A nonparametric test for equality of distributions with mixed categorical and continuous data." *Journal of Econometrics* 148:186–200.
- Pieralli, S. 2011. "Land quality index in a separable DEA framework. An application to Kenyan household farmers." *EAERE Conference paper available at <http://www.webmeets.com/EAERE/2011/m/viewpaper.asp?pid=190>*, pp. 1–31.

Figures

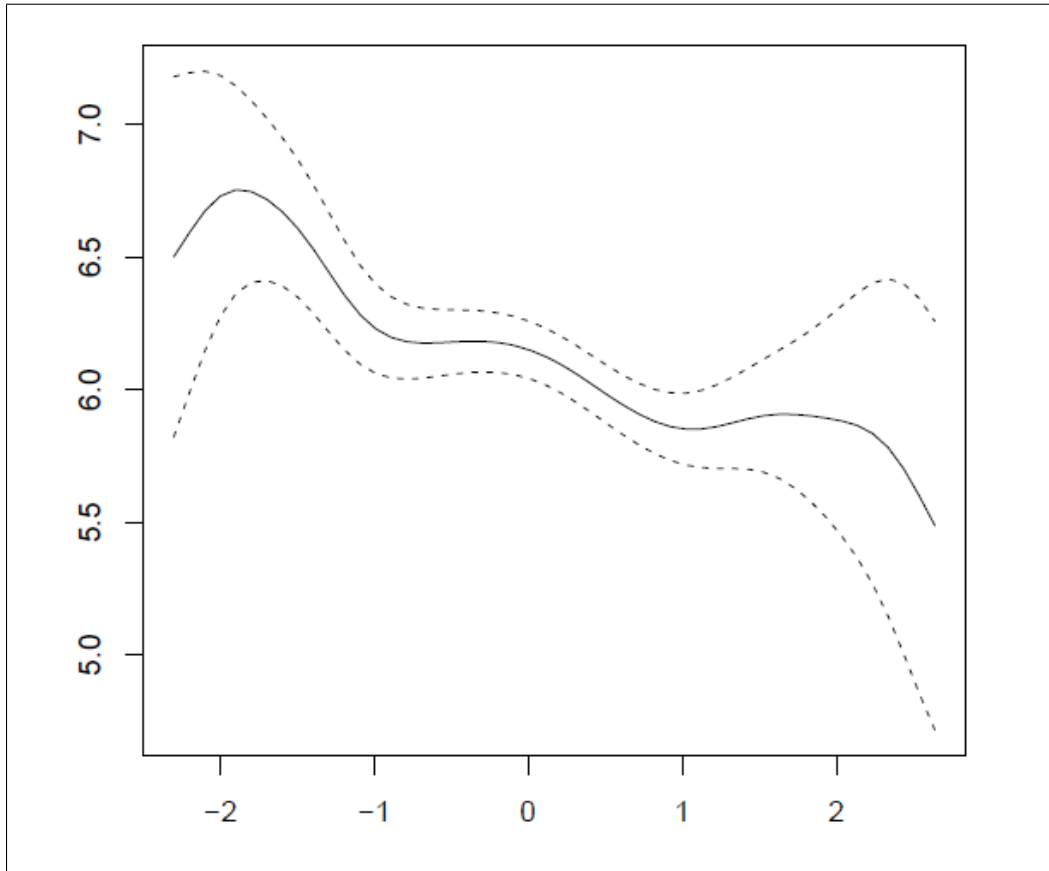


Figure 1. Negative empirical relationship between natural logarithm of yield of dry maize (on the vertical axis) and natural logarithm of land size (on the horizontal axis)

Note: There are only 443 observations considered in this graph because only 443 observations out of the 452 have a strictly positive yield.

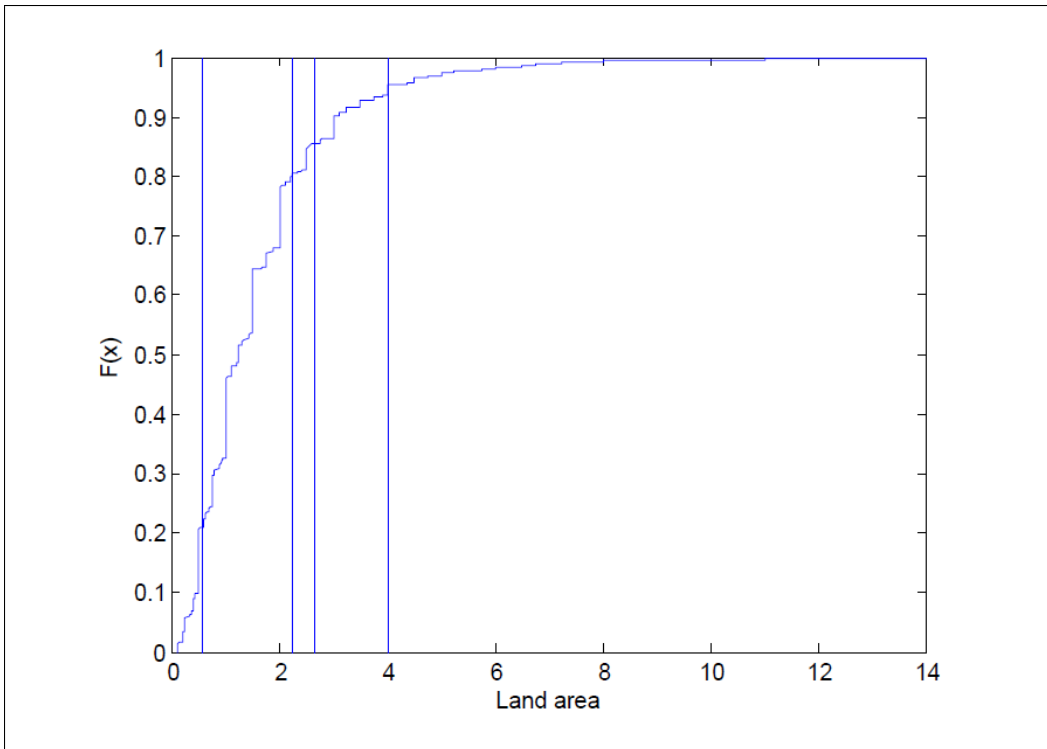


Figure 2. Empirical cumulative distribution of land size

Note: The lines plotted are in correspondence of 0.55 acres, 2.25 acres, 2.65 acres, and 4 acres.

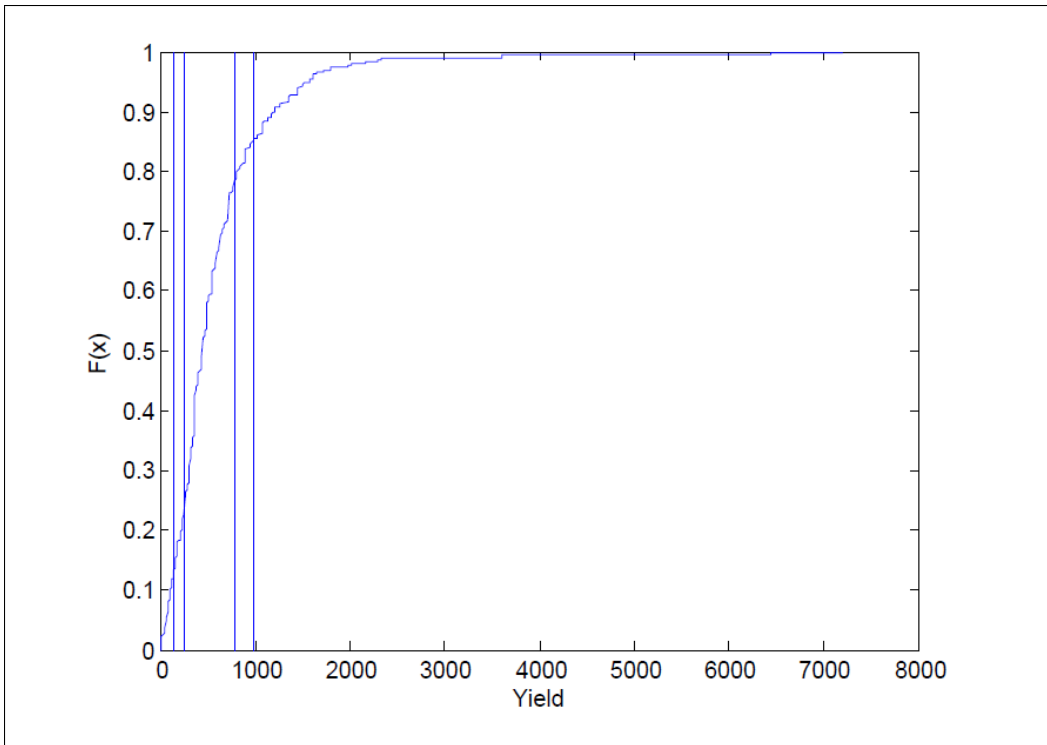


Figure 3. Empirical cumulative distribution of yield

Note: The lines plotted are in correspondence of values of yield of 135 Kg acre⁻¹, 240 Kg acre⁻¹, 787.5 Kg acre⁻¹, and 981 Kg acre⁻¹.

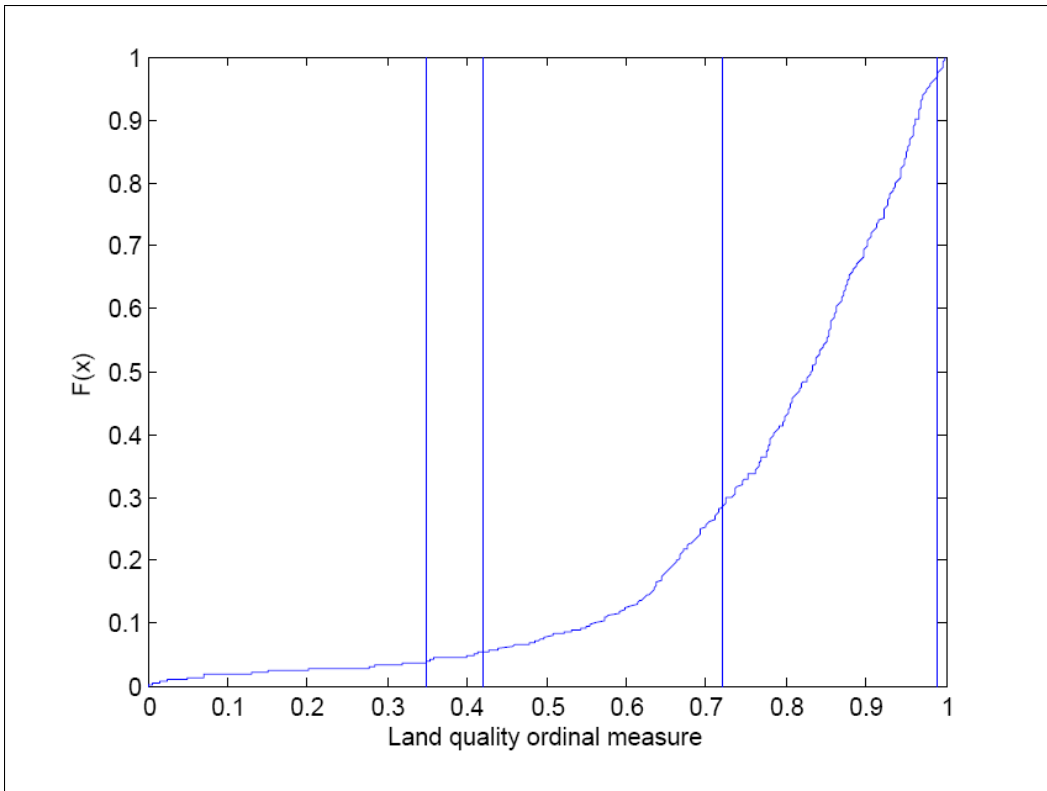


Figure 4. Empirical cumulative distribution of land quality under variable returns to scale

Note: The lines plotted are in correspondence of values of land quality index of 0.35, 0.42, 0.7199, and 0.99.

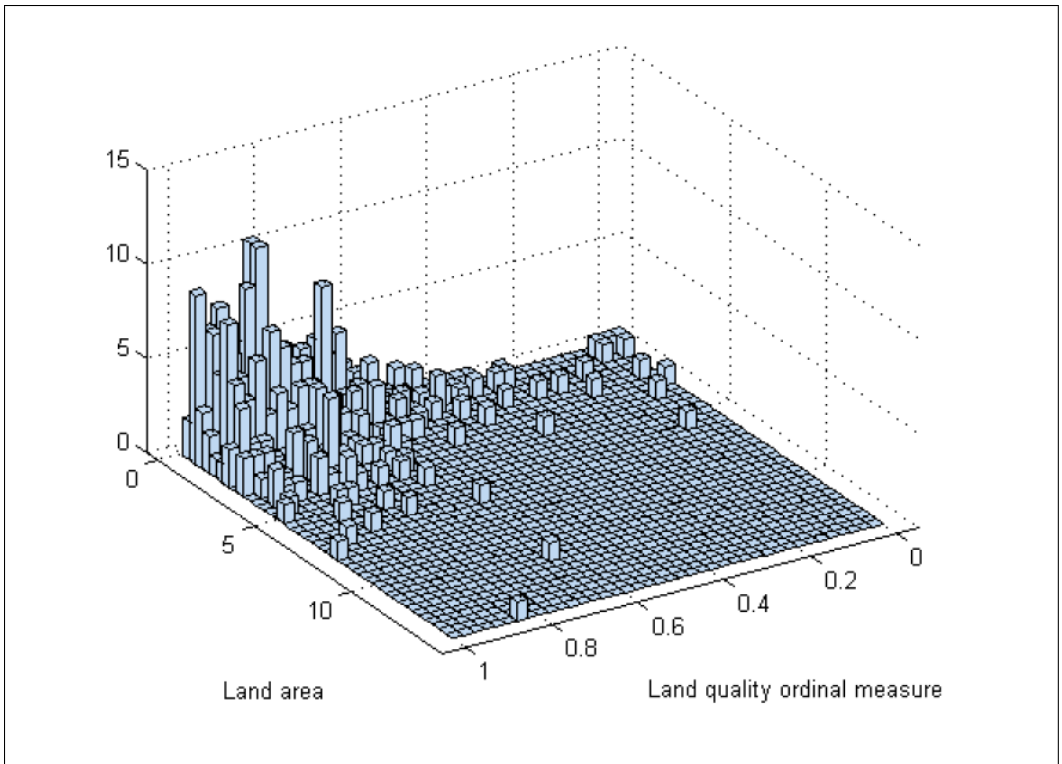


Figure 5. Empirical joint histogram of land area and land quality index under variable returns to scale

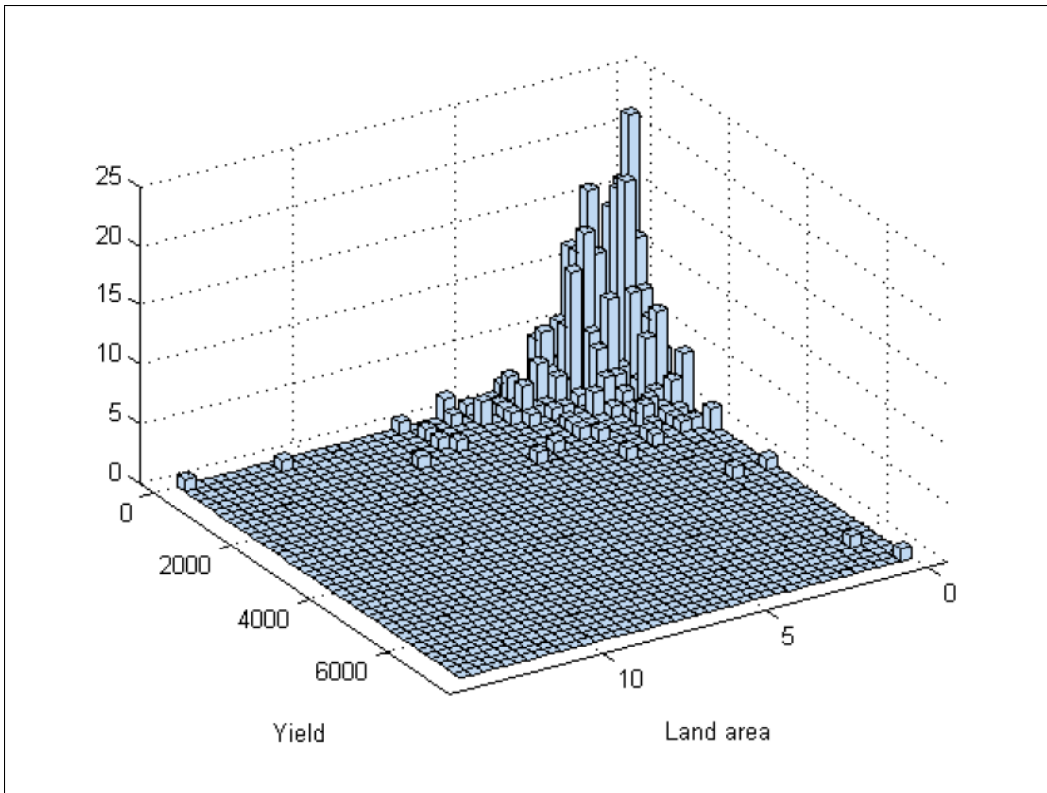


Figure 6. Empirical joint histogram of land area and observed yield

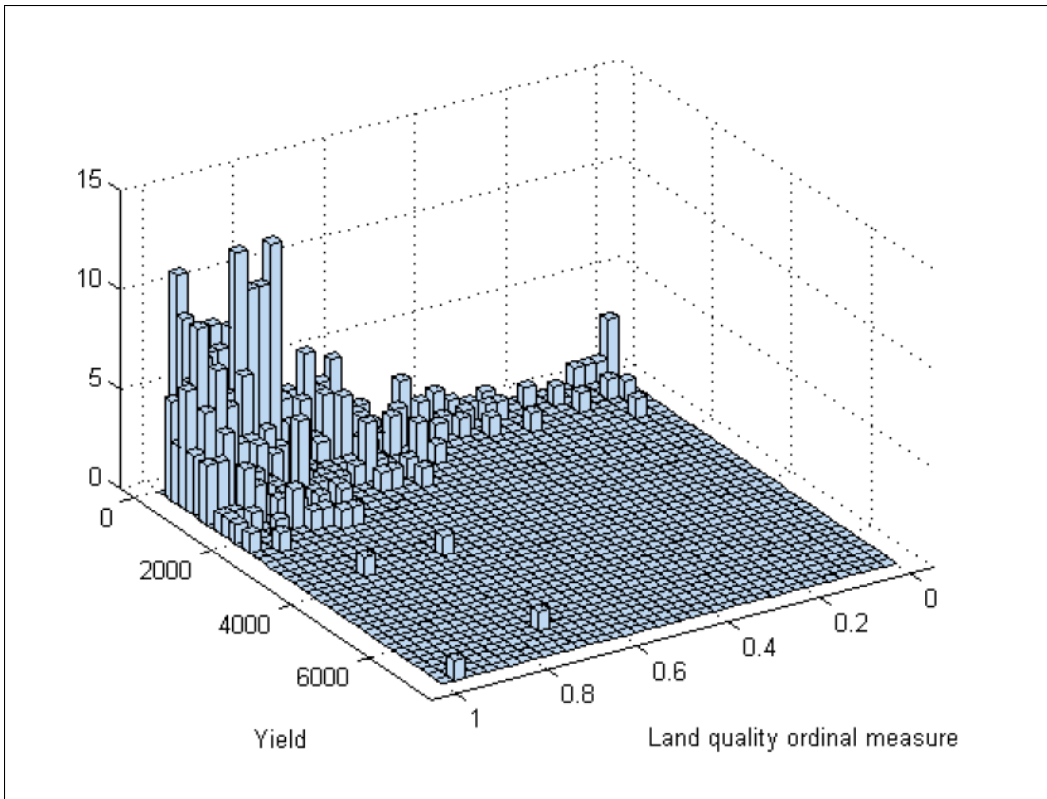


Figure 7. Empirical joint histogram of land quality under variable returns to scale and observed yield

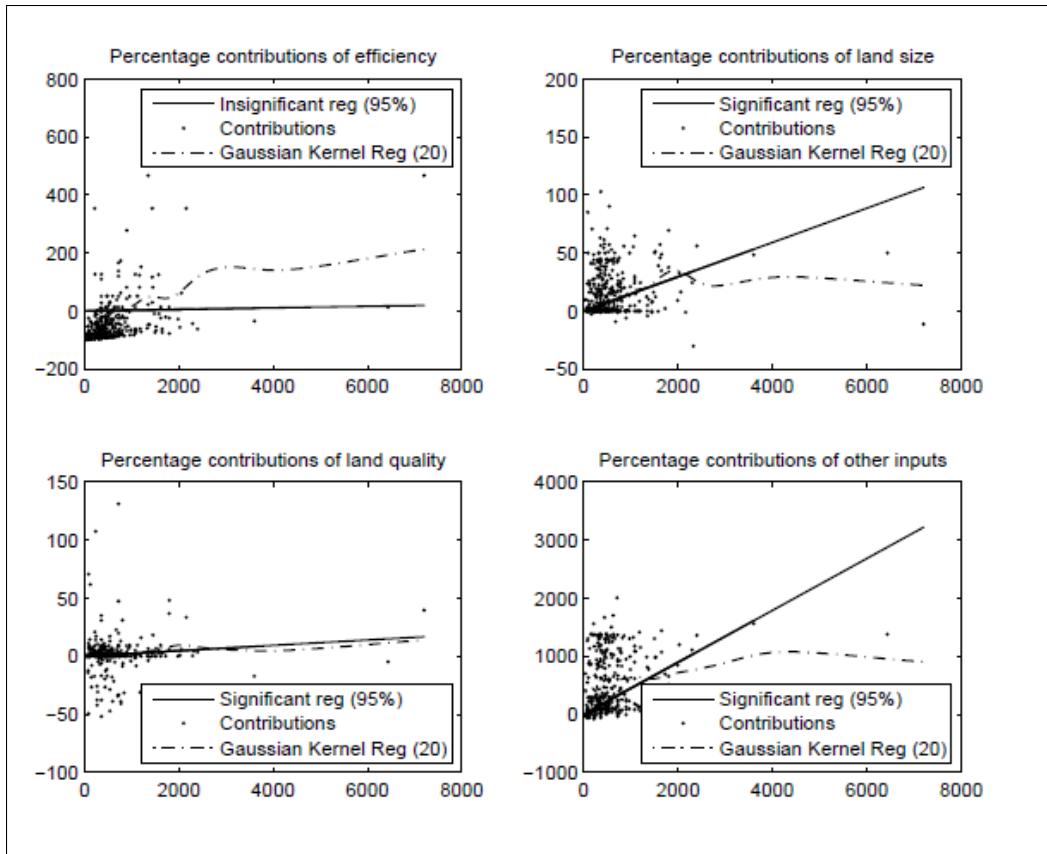


Figure 8. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with little land size, low yield, and low land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

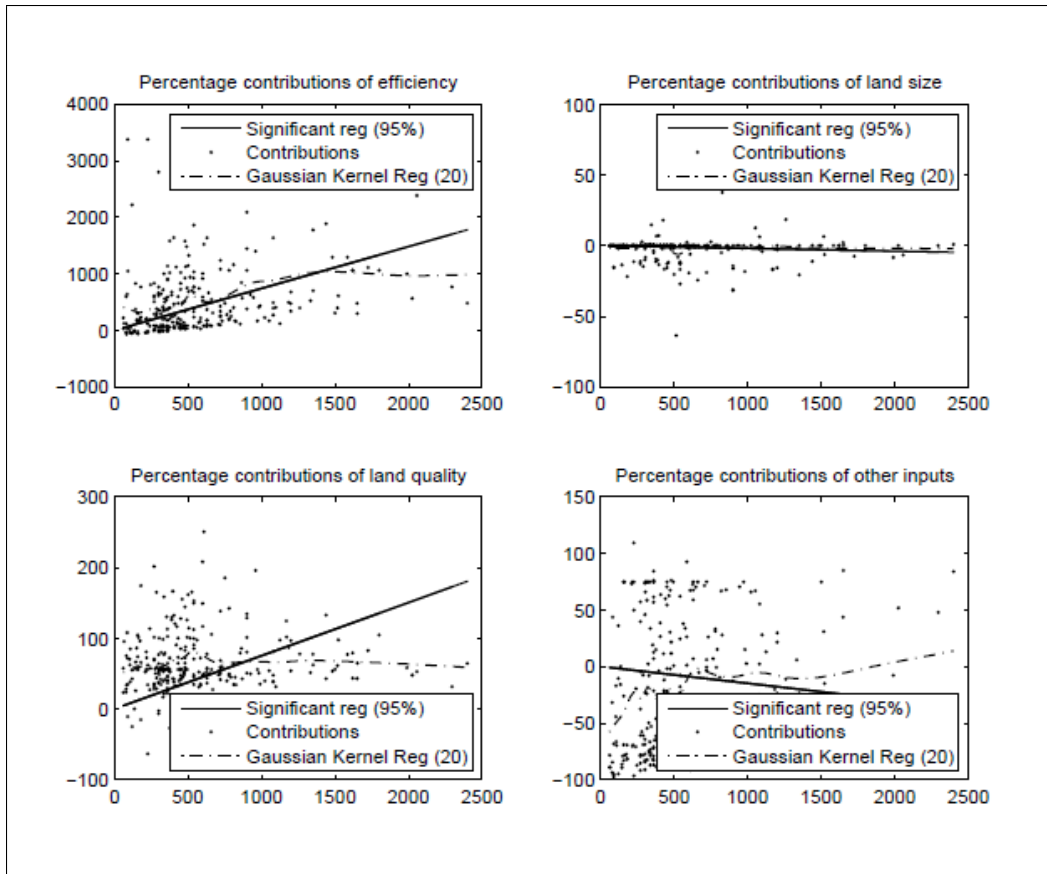


Figure 9. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with big land size, low yield, and low land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

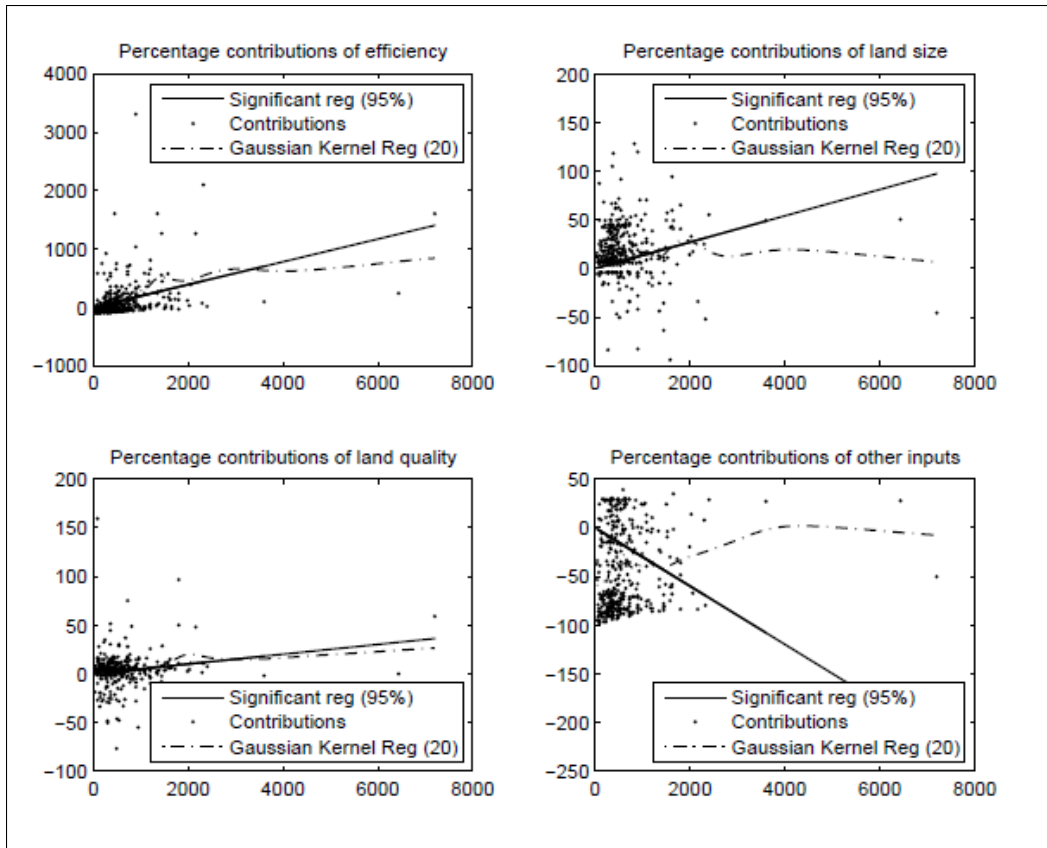


Figure 10. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with little land size, high yield, and low land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

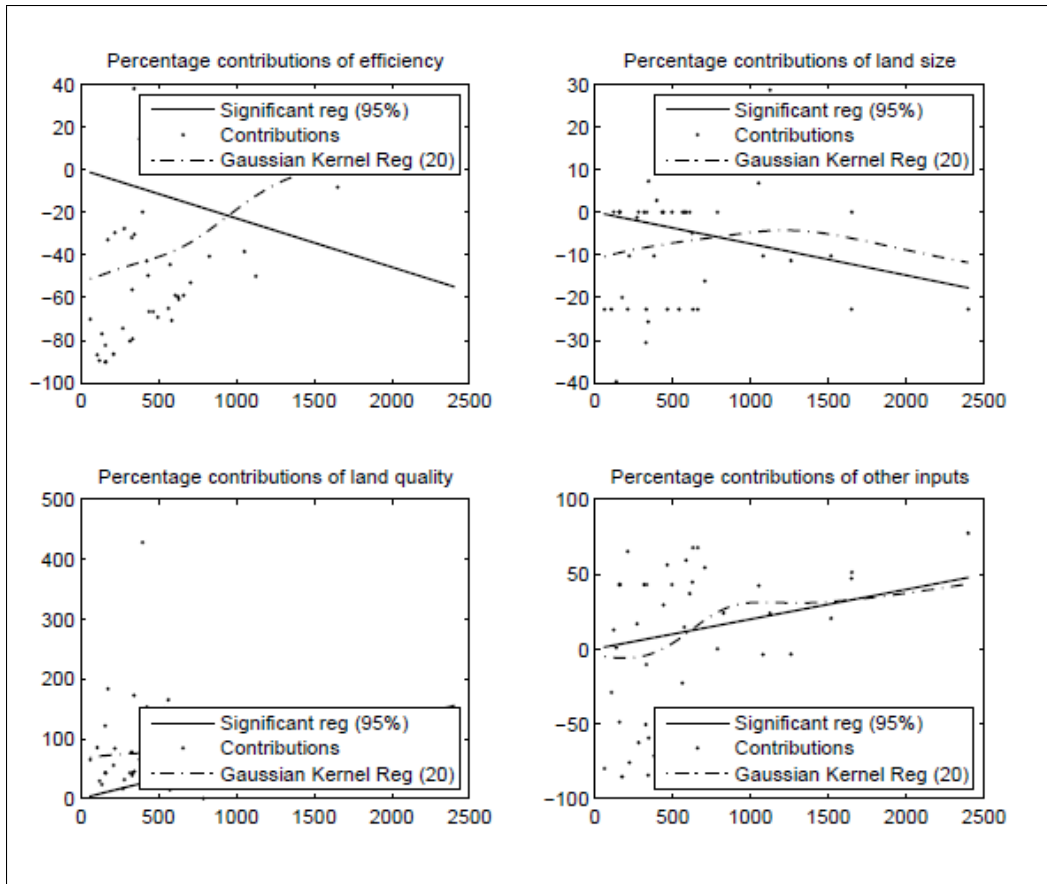


Figure 11. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with big land size, high yield, and low land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

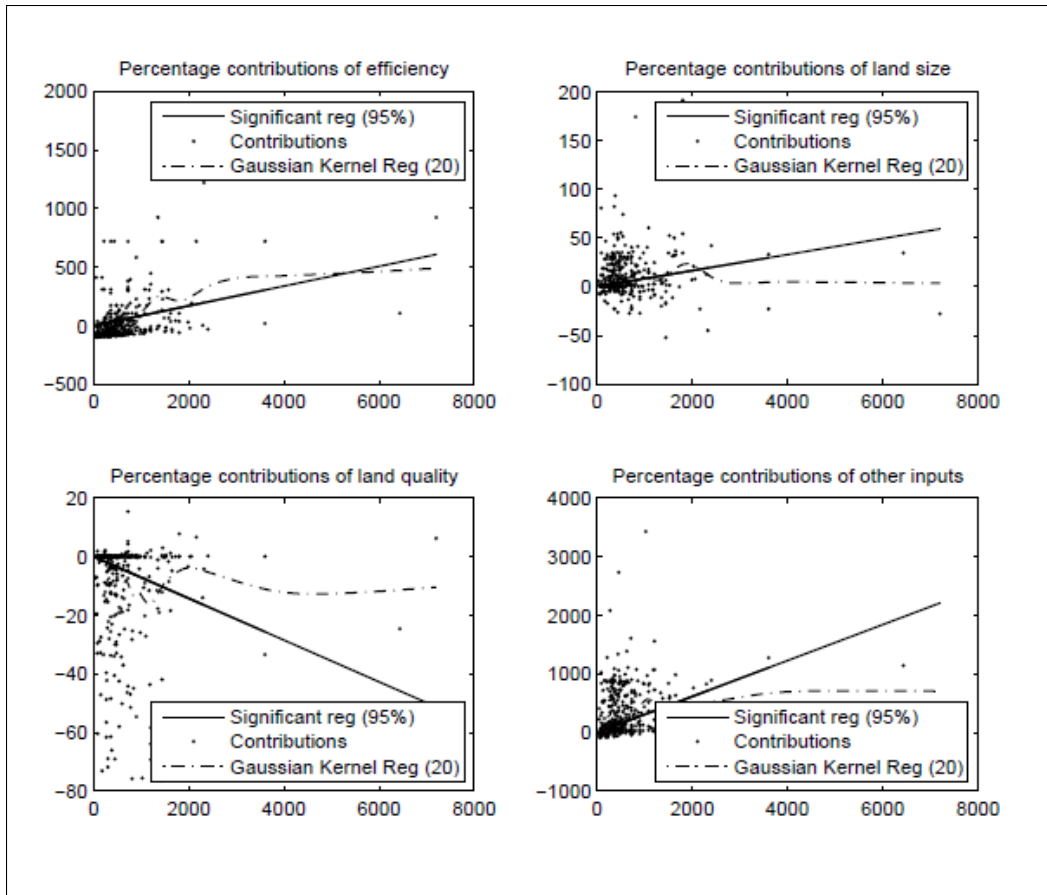


Figure 12. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with little land size, low yield, and high land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

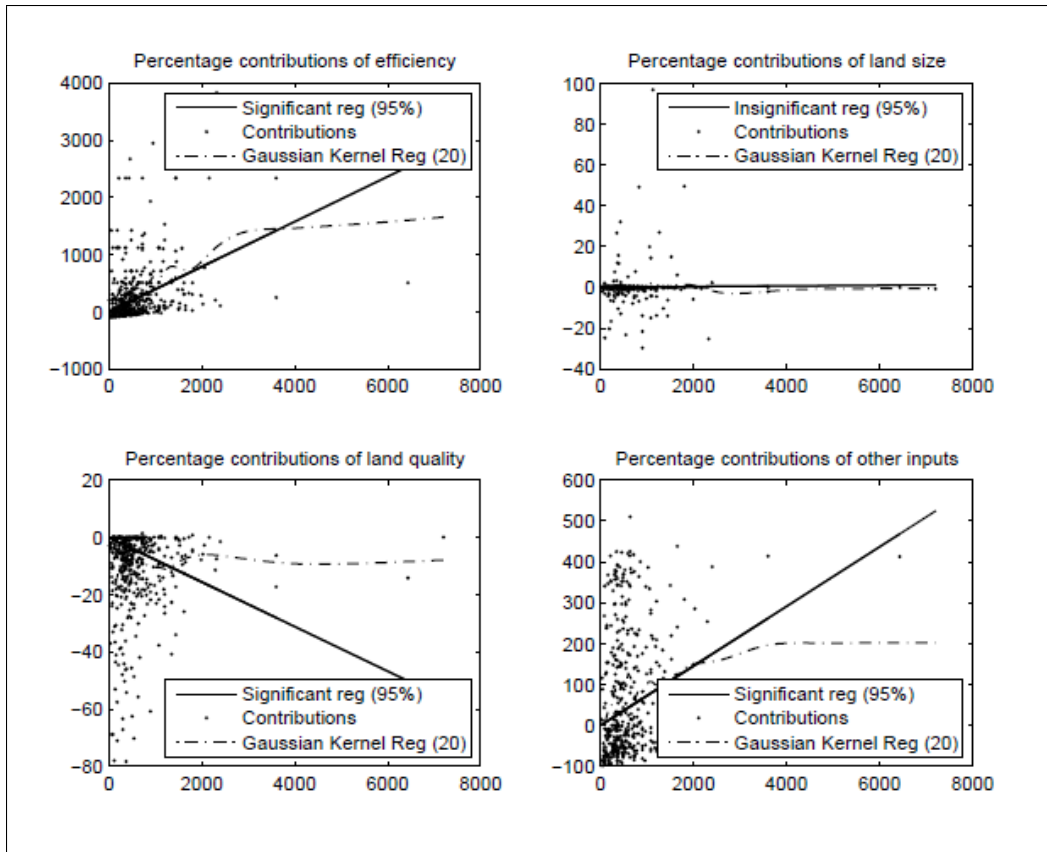


Figure 13. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with big land size, low yield, and high land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

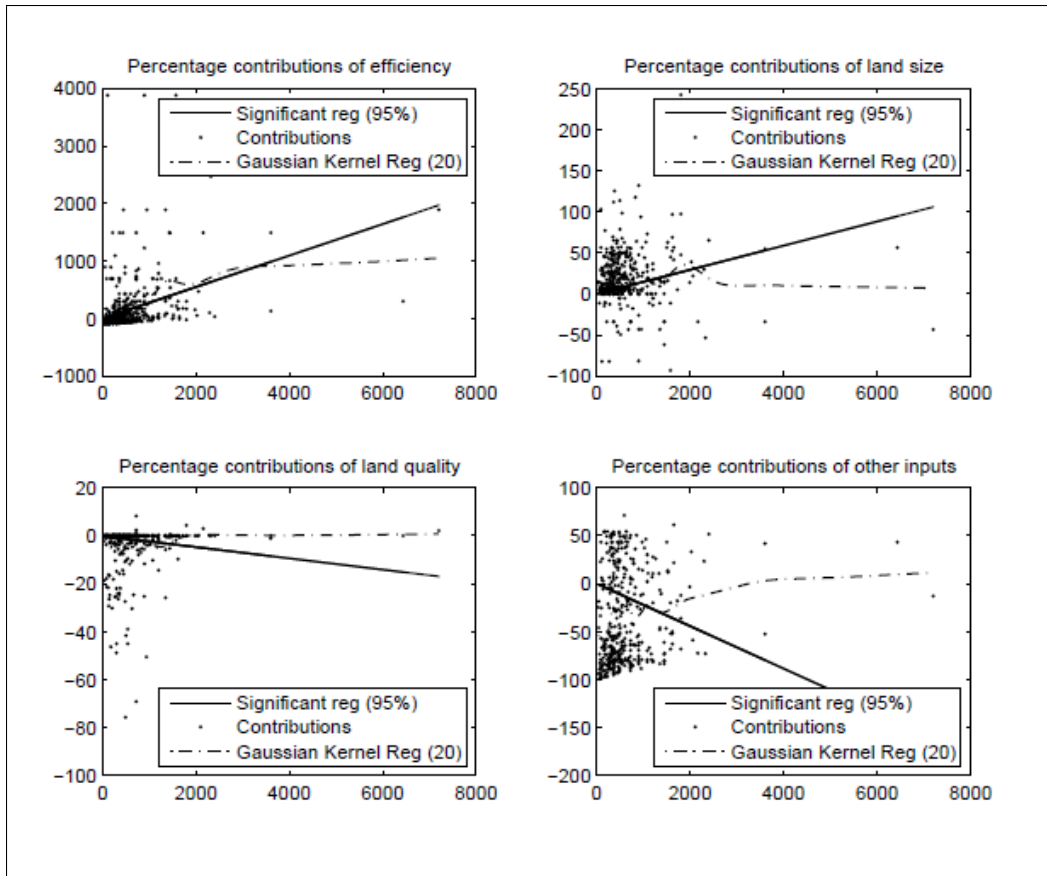


Figure 14. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with little land size, high yield, and high land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

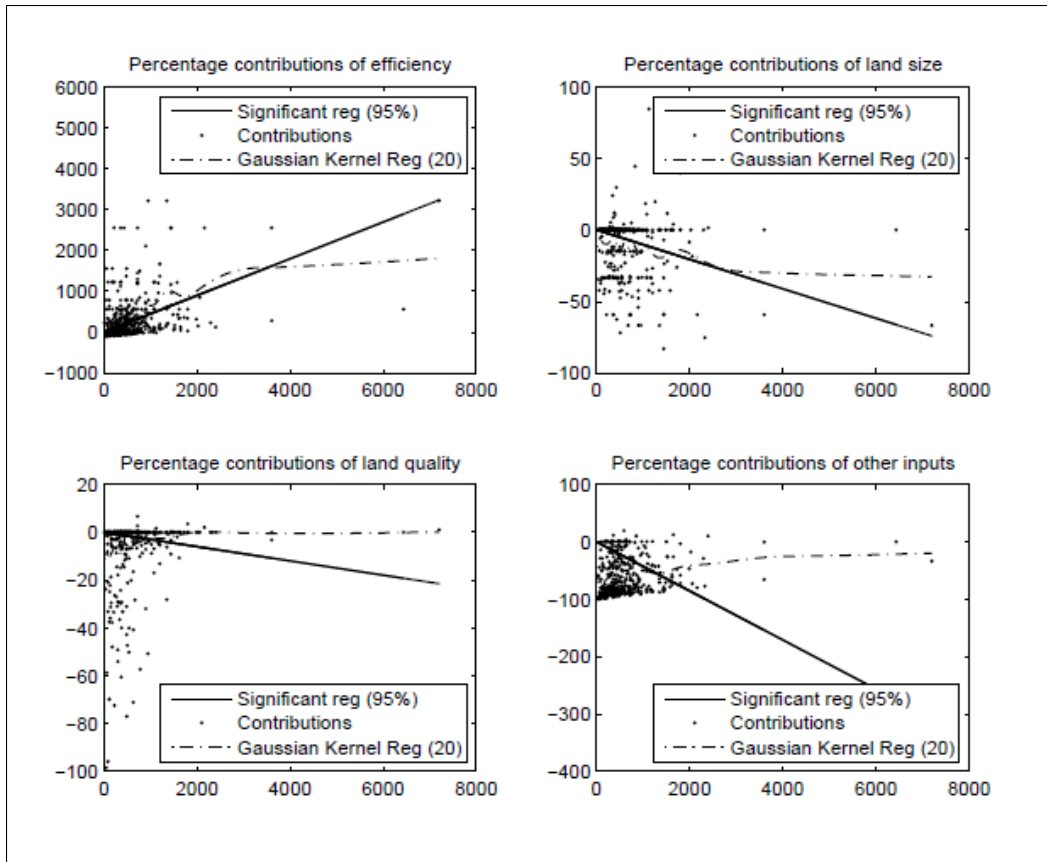


Figure 15. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with big land size, high yield, and high land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

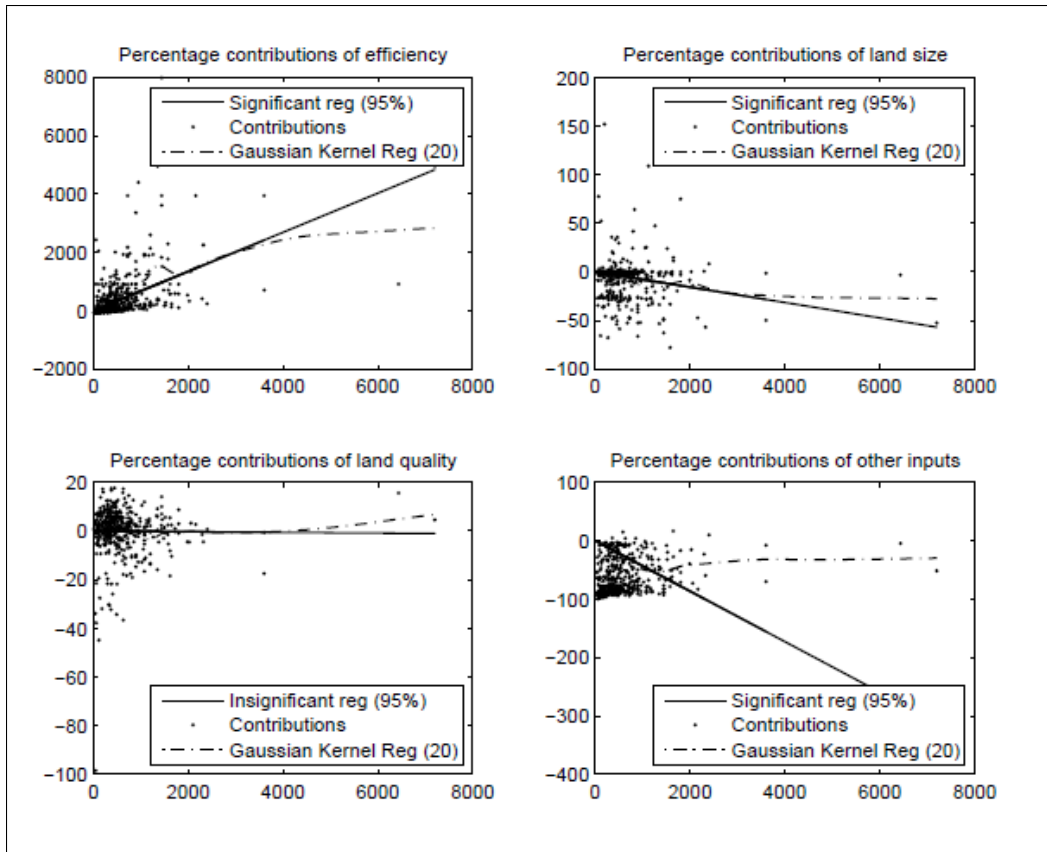


Figure 16. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with average land size, yield, and land quality under constant returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

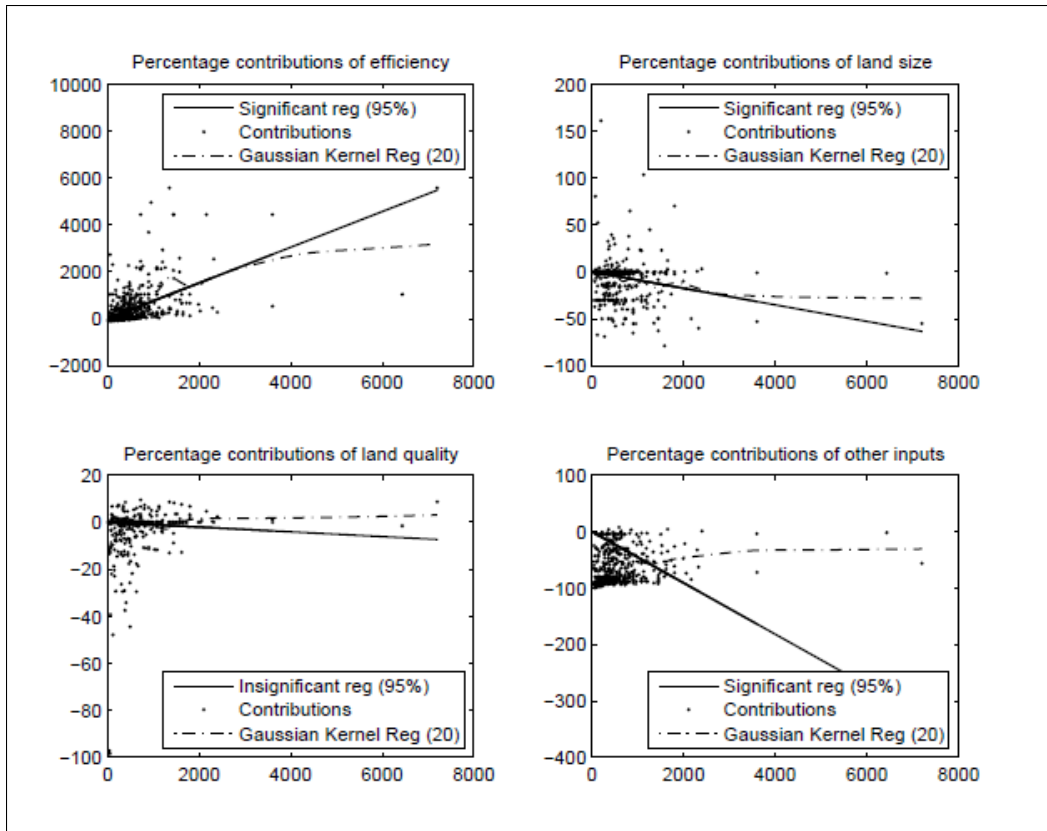


Figure 17. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with average land size, yield, and land quality under non-increasing returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

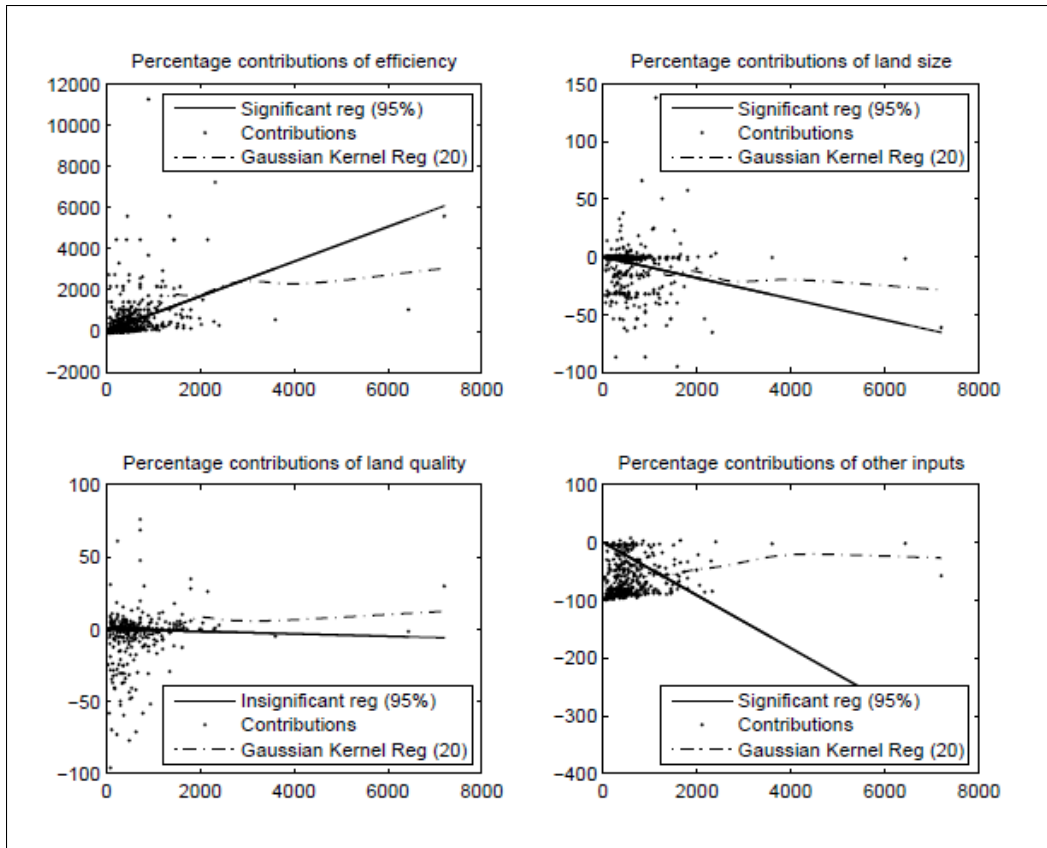


Figure 18. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with average land size, yield, and land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

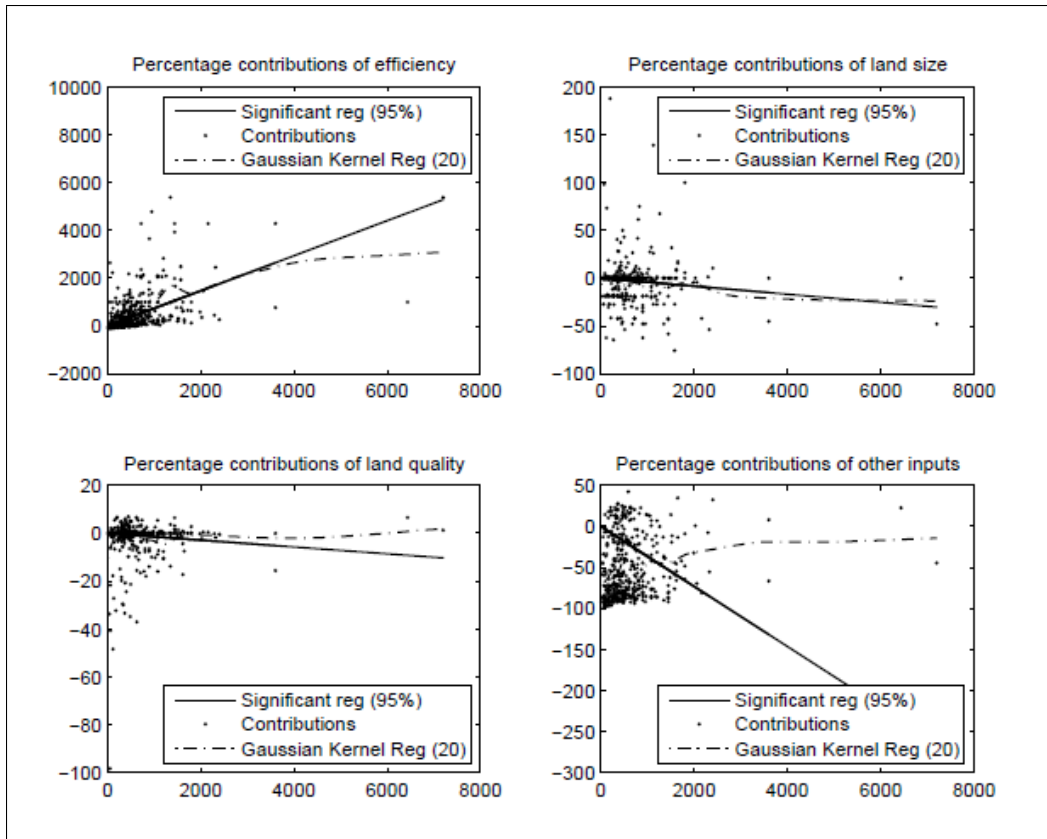


Figure 19. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with median land size, yield, and land quality under constant returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

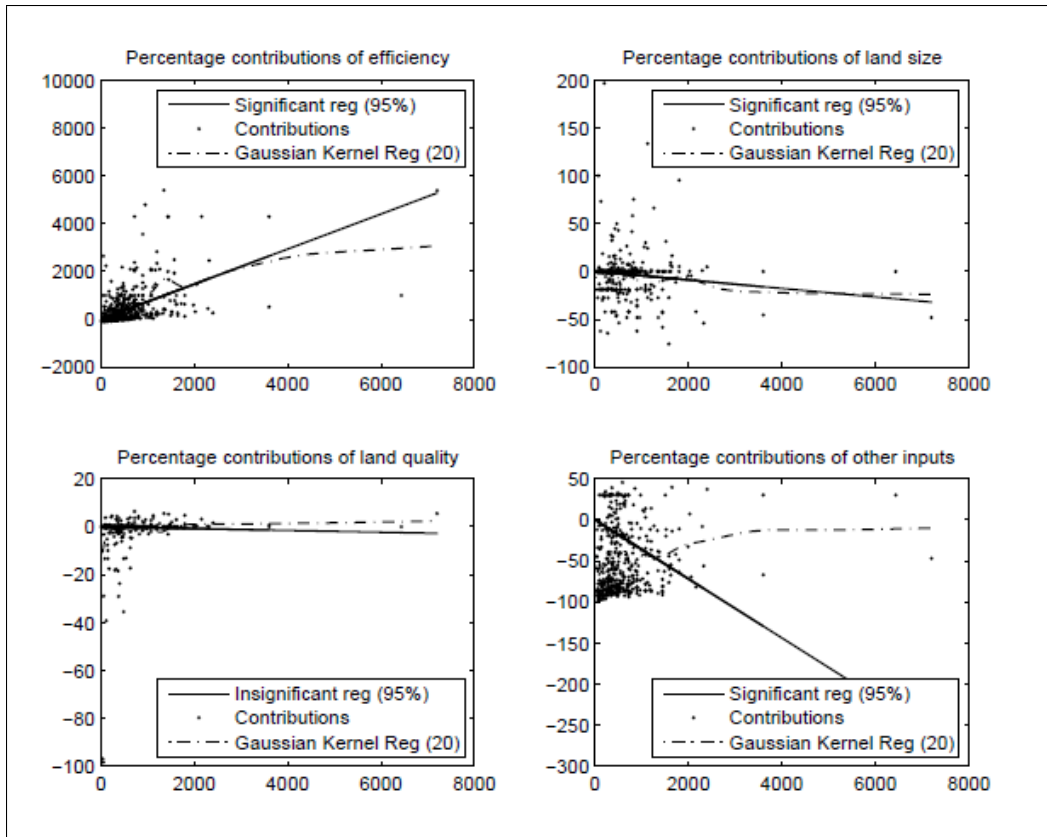


Figure 20. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with median land size, yield, and land quality under non-increasing returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

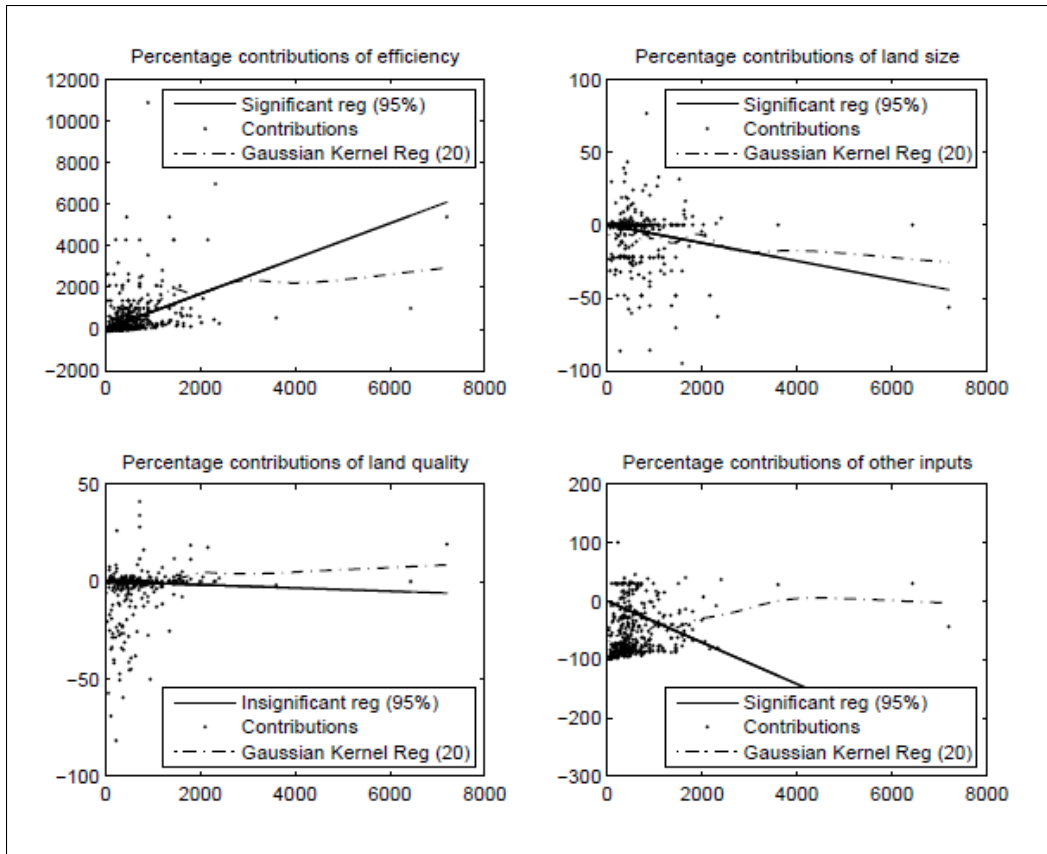


Figure 21. Percentage contributions of efficiency index, land quality, land size, and other inputs related to changes in observed yield when unit of reference is a household with median land size, yield, and land quality under variable returns to scale

Note: Percentage contributions are plotted as dots. Linear regression line is plotted with 95% significance level in the legend stating whether the coefficient is significant around the mean. Gaussian kernel line is also plotted to show the degree to which local regression (when done with 20 units at a time) differs from general regression.

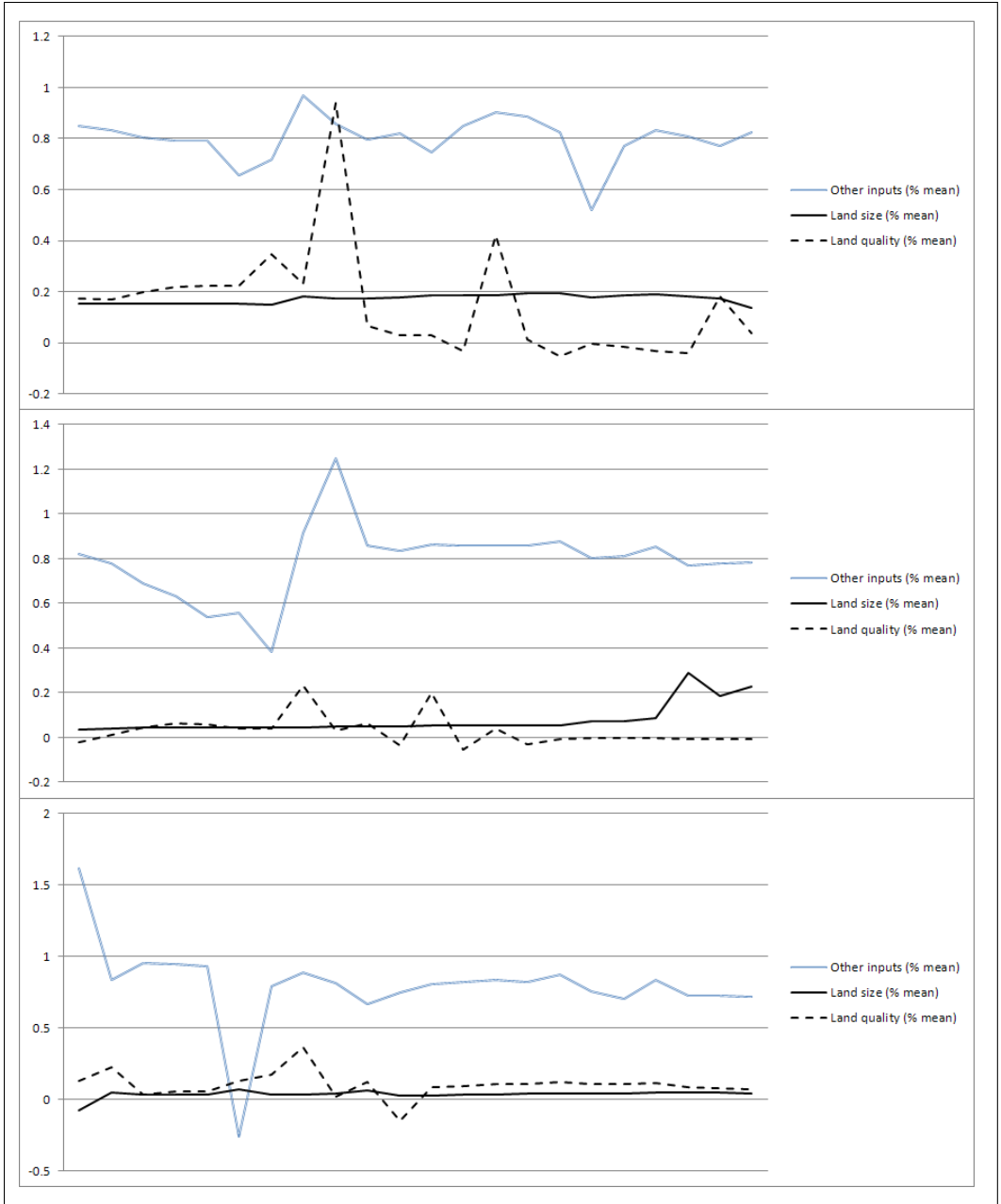


Figure 22. Average percentage contribution rates to yield difference for land quality, land size, and other inputs

Note: Measurements are presented for the last twenty percentiles (from the 80th to the 100th) of reference levels of inputs and outputs when calculating the land quality measure.

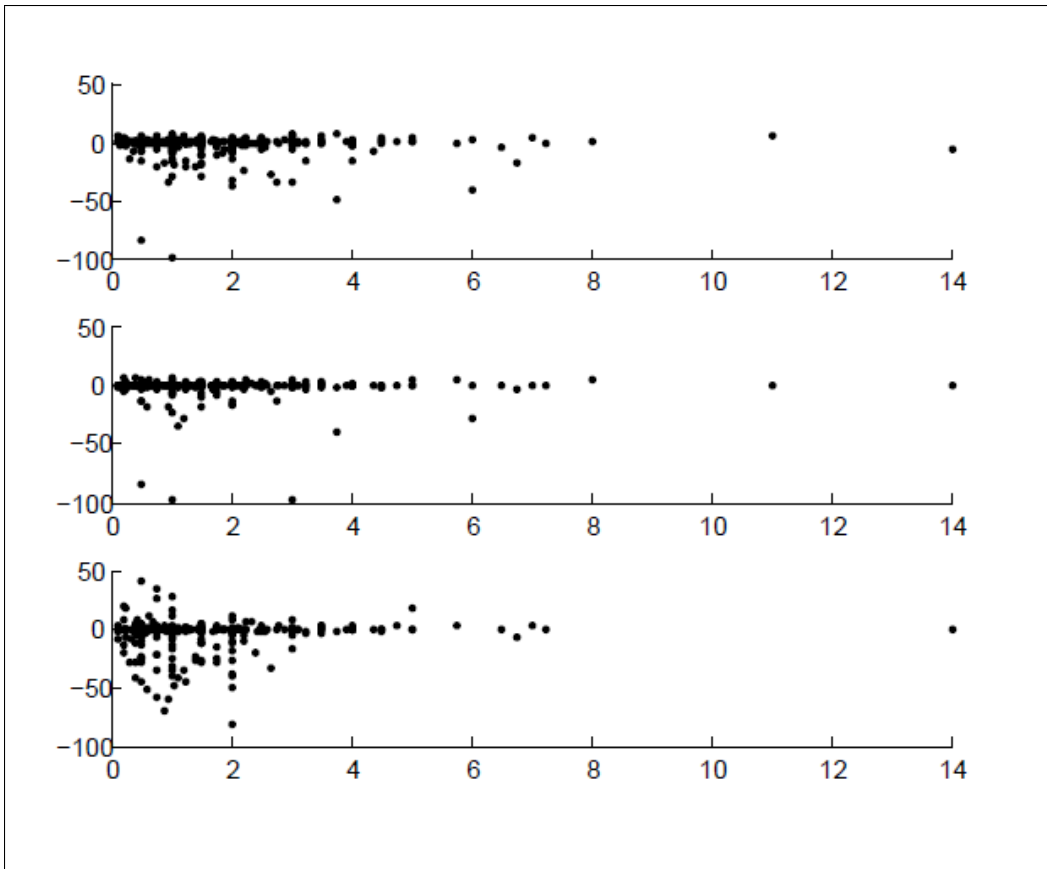


Figure 23. Percentage contributions of land quality under constant (upper), non-increasing (middle), and variable (lower) returns to scale against size of land

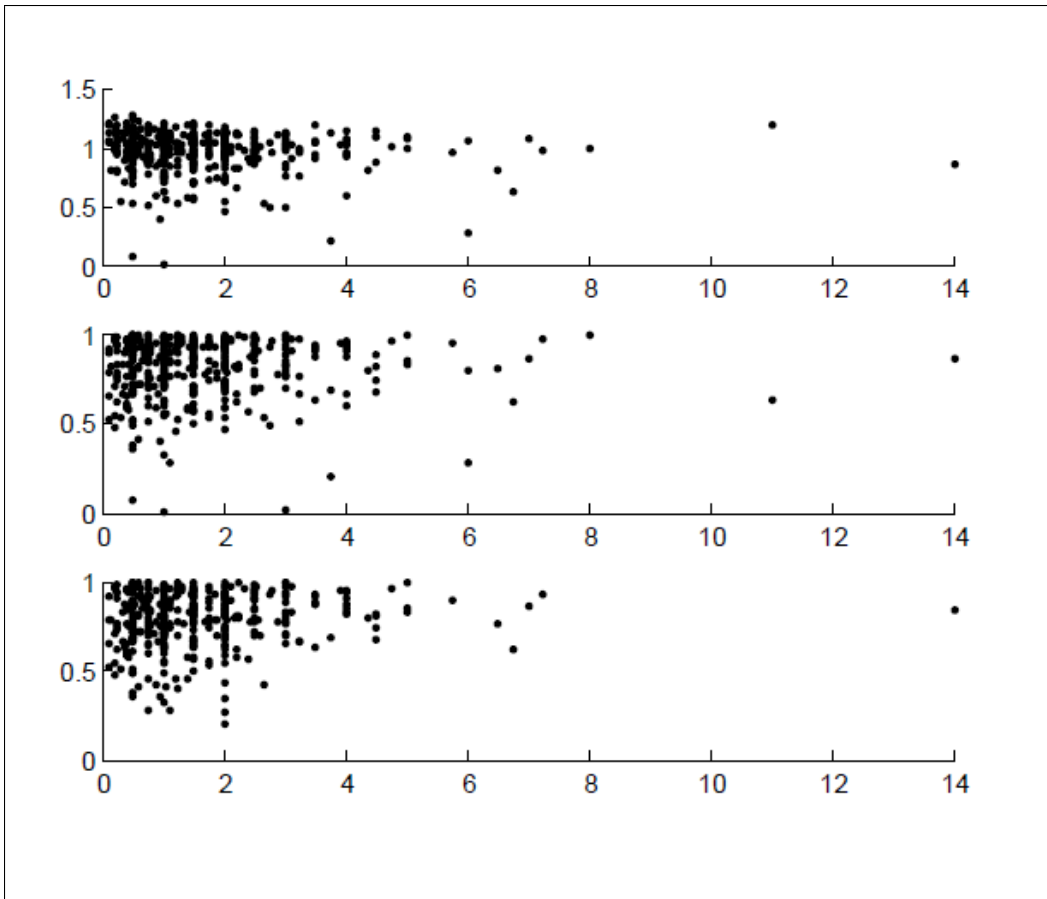


Figure 24. Land quality measurements under constant (upper), non-increasing (middle), and variable (lower) returns to scale against size of land

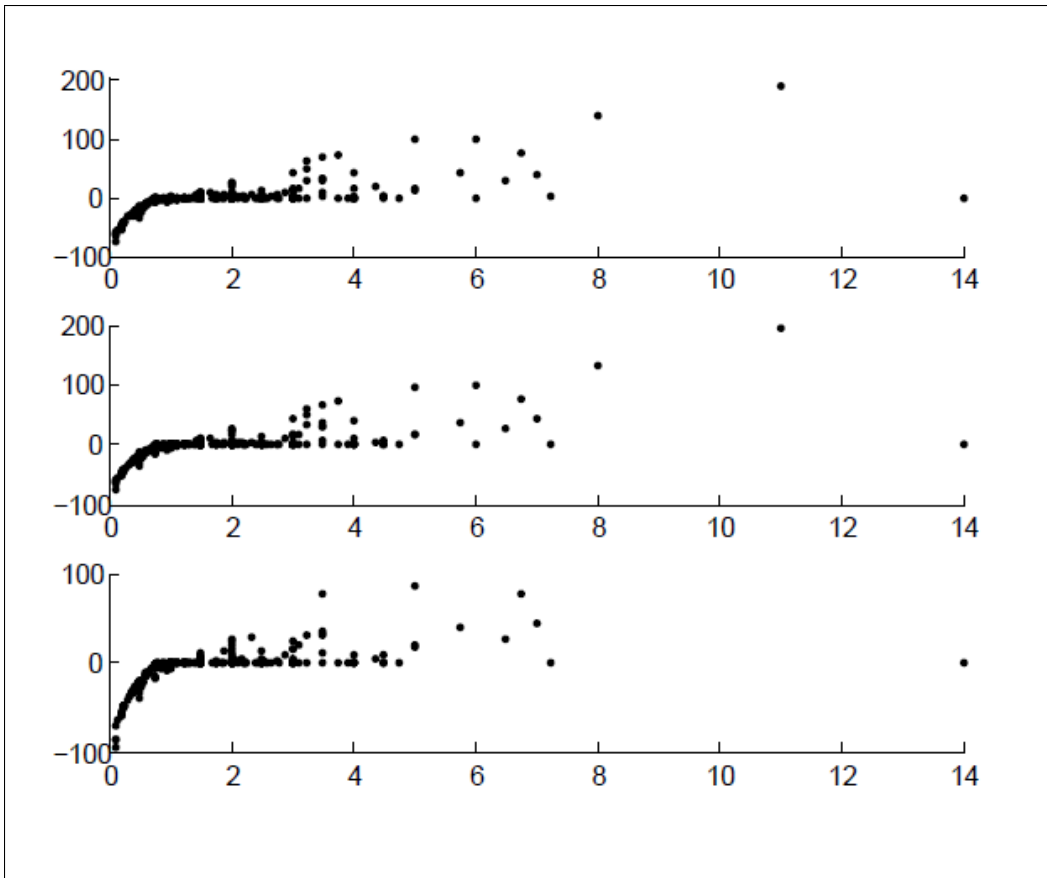


Figure 25. Percentage contributions of land size under constant (upper), non-increasing (middle), and variable (lower) returns to scale against size of land

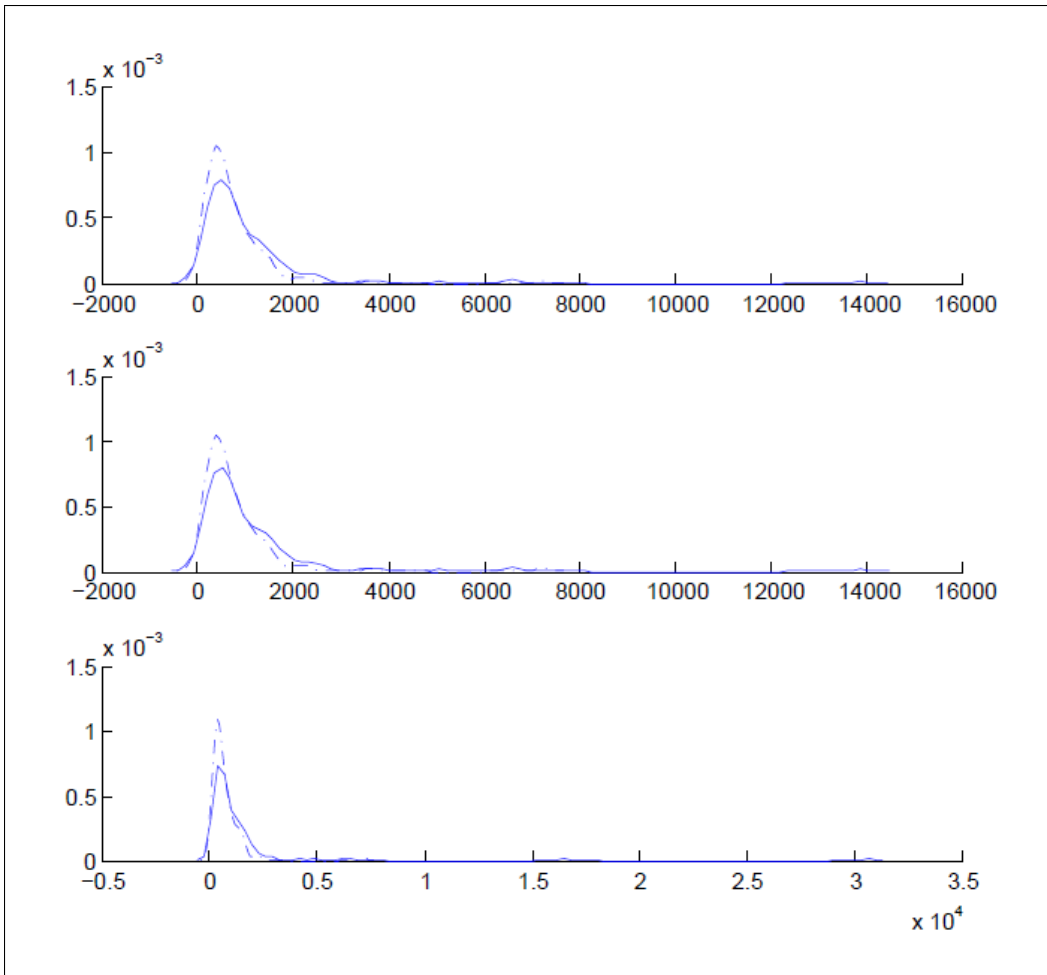


Figure 26. Land size contribution to yield distributions for smaller farmers

Note: Kernel smoothing probability densities from top to bottom under constant, non-increasing, and variable returns to scale up to 0.8 acres. Counterfactual distribution y_1^Q/l_1 is plotted as a solid line while observed yield distribution is the dashed line.

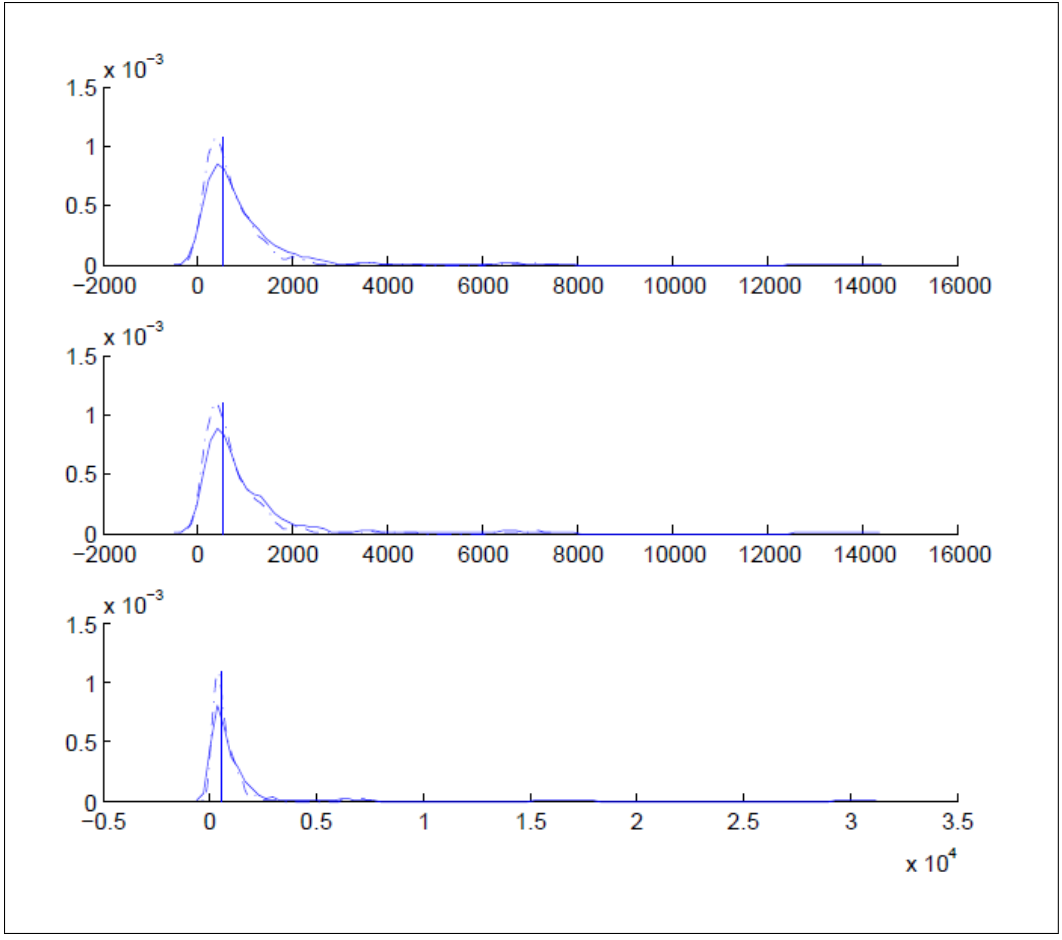


Figure 27. Negative land size contributions to observed yield distribution

Note: Kernel smoothing probability densities of land size contributions from top to bottom under constant, non-increasing, and variable returns for households with negative land size contributions. Counterfactual distribution y_1^Q/l_1 is plotted as a solid line while observed yield distribution is the dashed line.

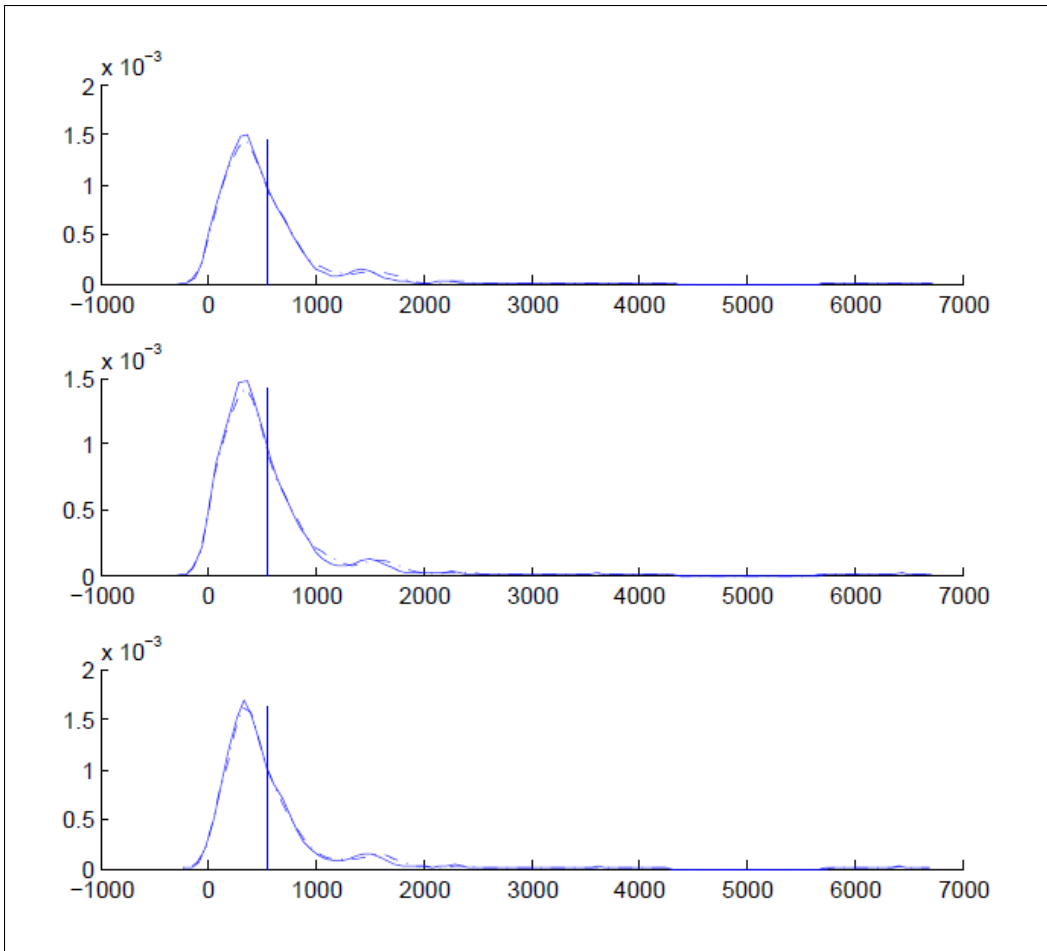


Figure 28. Non negative land size contributions to observed yield distribution

Note: Kernel smoothing probability densities of land size contributions from top to bottom under constant, non-increasing, and variable returns for households with non negative land size contributions. Counterfactual distribution y_1^Q/l_1 is plotted as a solid line while observed yield distribution is the dashed line.

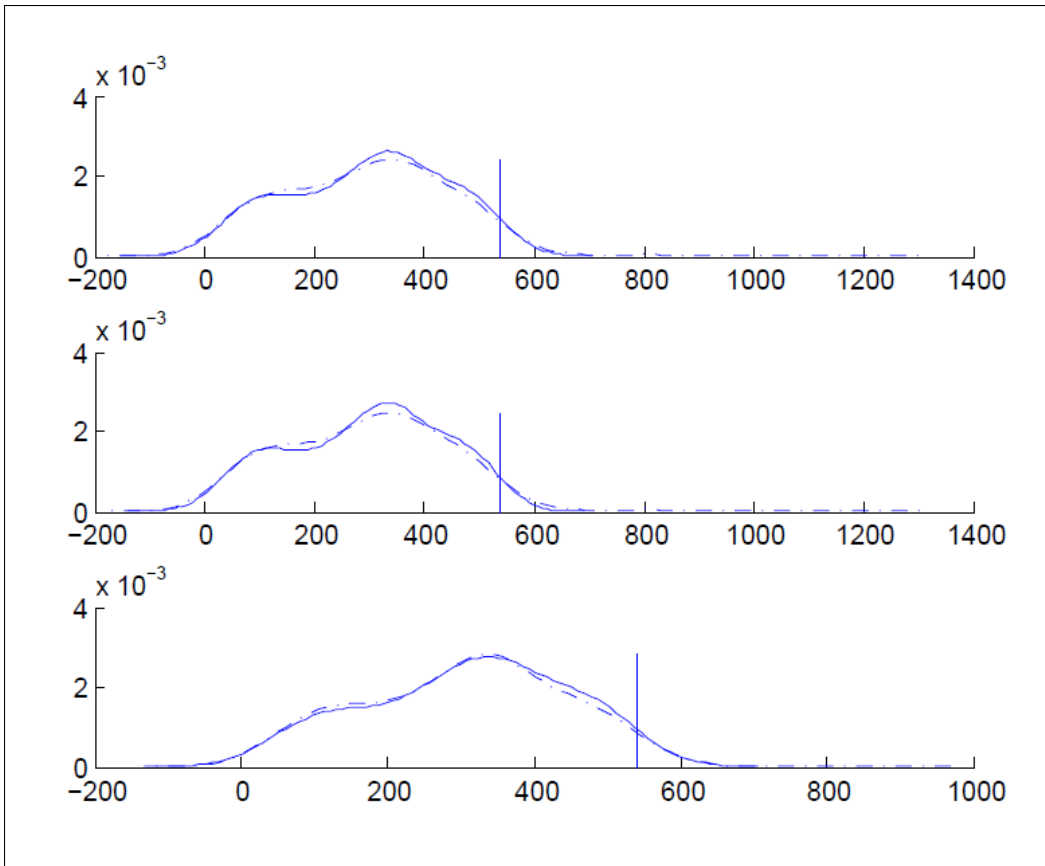


Figure 29. Land size contributions to yield distributions for farmers with yields lower than the median

Note: Kernel smoothing probability densities of land size contributions from top to bottom under constant, non-increasing, and variable returns to scale for farmers with yields lower than the median (540 Kg acre^{-1}). Counterfactual distribution y_1^Q/l_1 is plotted as a solid line while observed yield distribution is the dashed line.

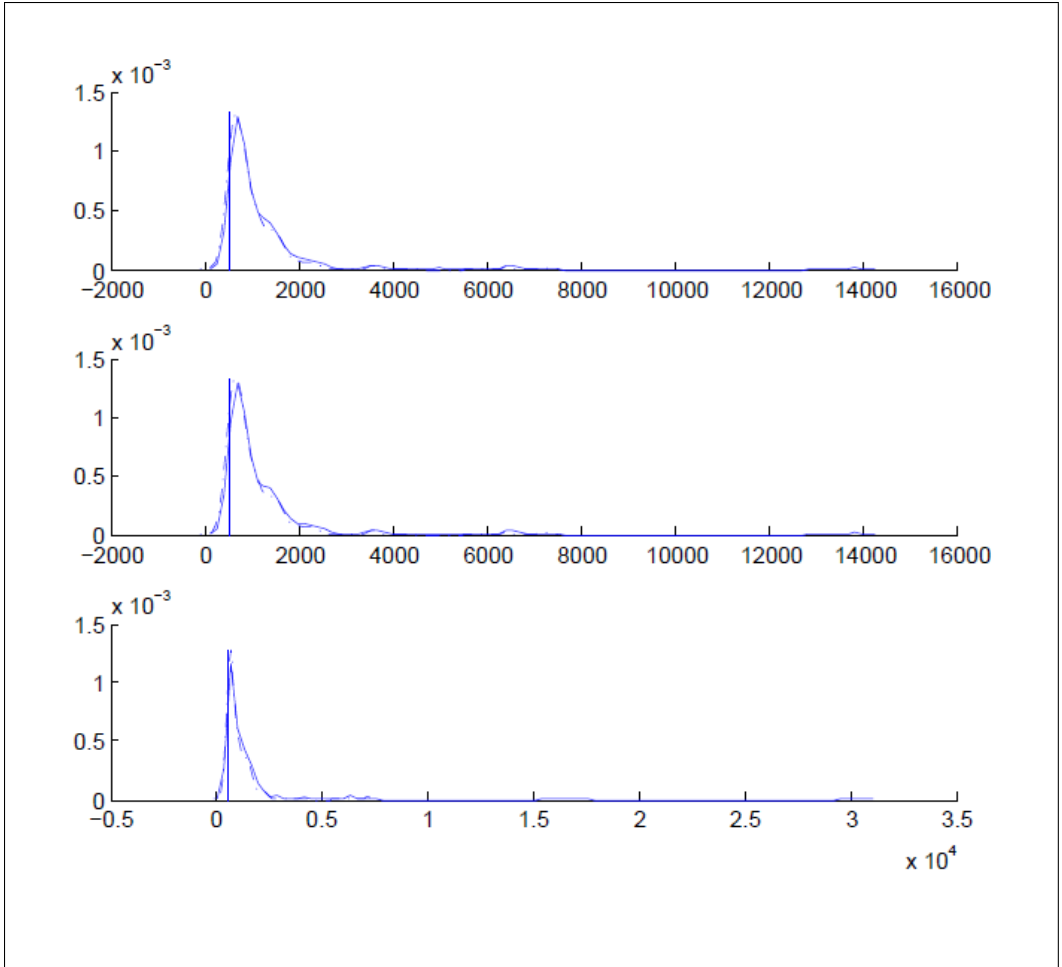


Figure 30. Land size contributions to yield distributions for farmers with yields higher than the median

Note: Kernel smoothing probability densities of land size contributions from top to bottom under constant, non-increasing, and variable returns to scale for farmers with yields higher than the median (540 Kg acre^{-1}). Counterfactual distribution y_1^Q/l_1 is plotted as a solid line while observed yield distribution is the dashed line.

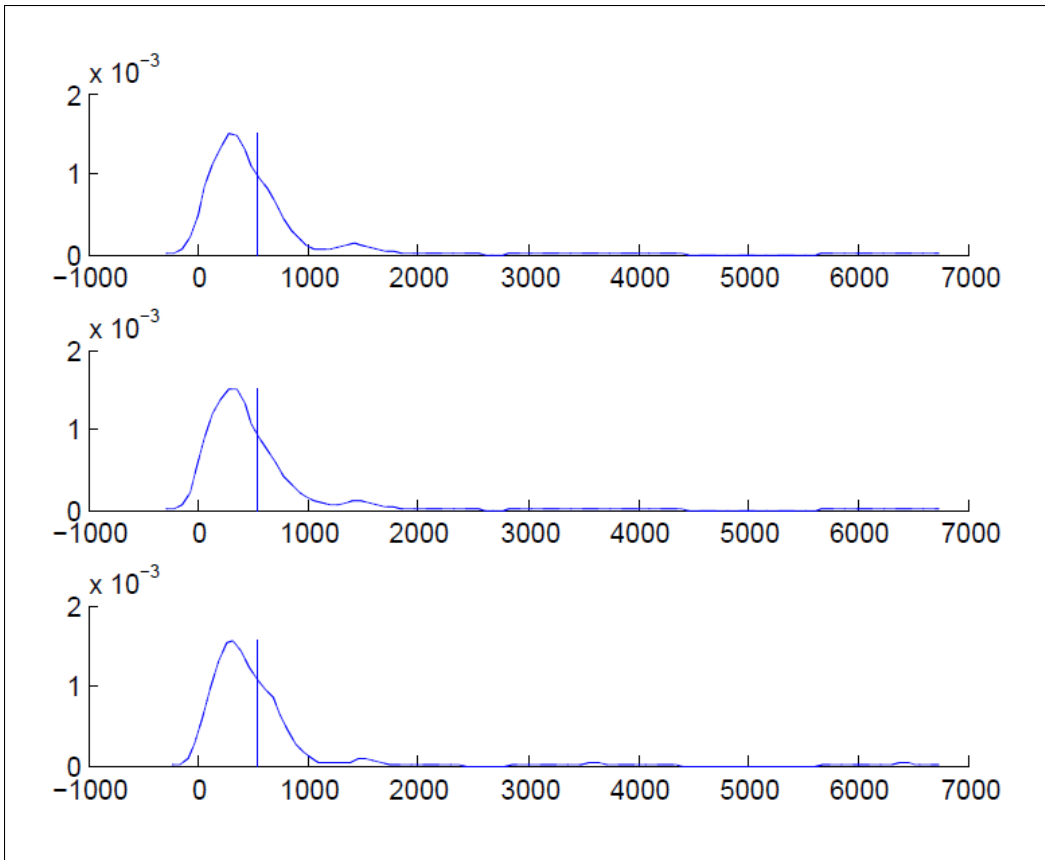


Figure 31. Observed yield distributions for farmers with zero land size contributions

Note: Kernel smoothing probability densities of yield distributions from top to bottom under constant, non-increasing, and variable returns to scale for farmers with zero land size contributions.

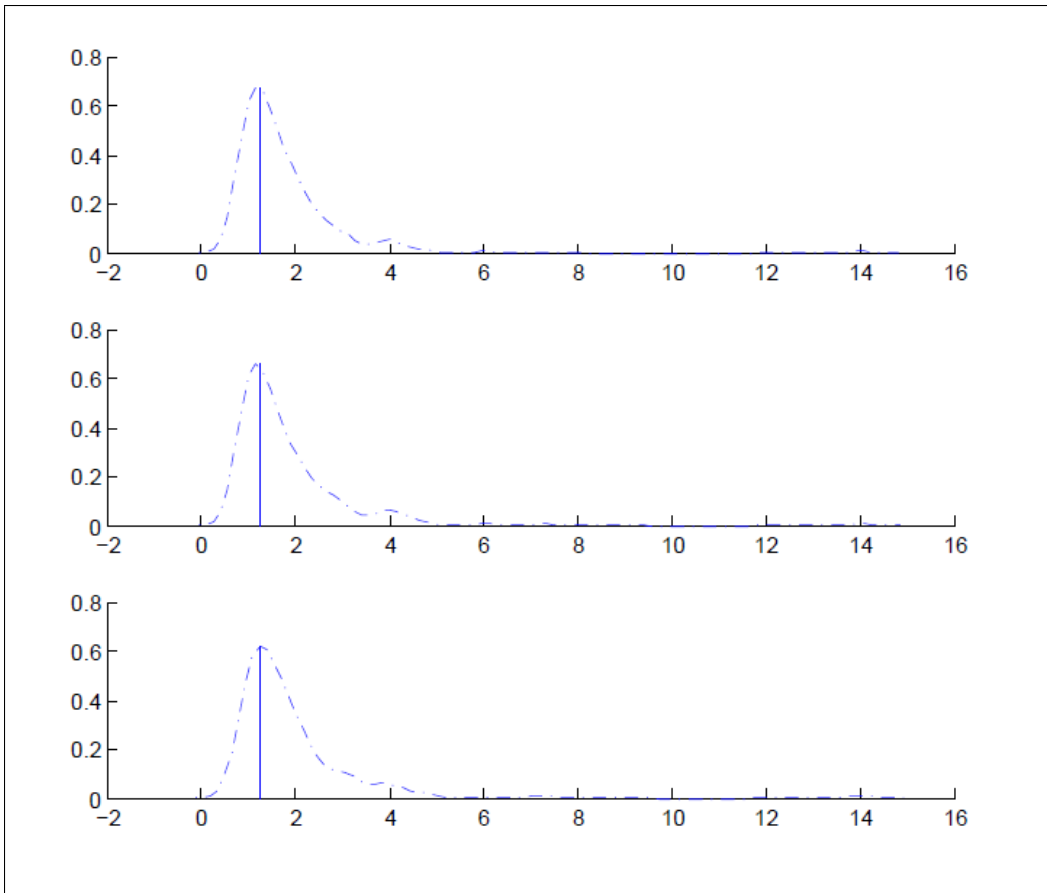


Figure 32. Observed land size distributions for farmers with zero land size contributions

Note: Kernel smoothing probability densities of land size distributions from top to bottom under constant, non-increasing, and variable returns to scale for farmers with zero land size contributions.

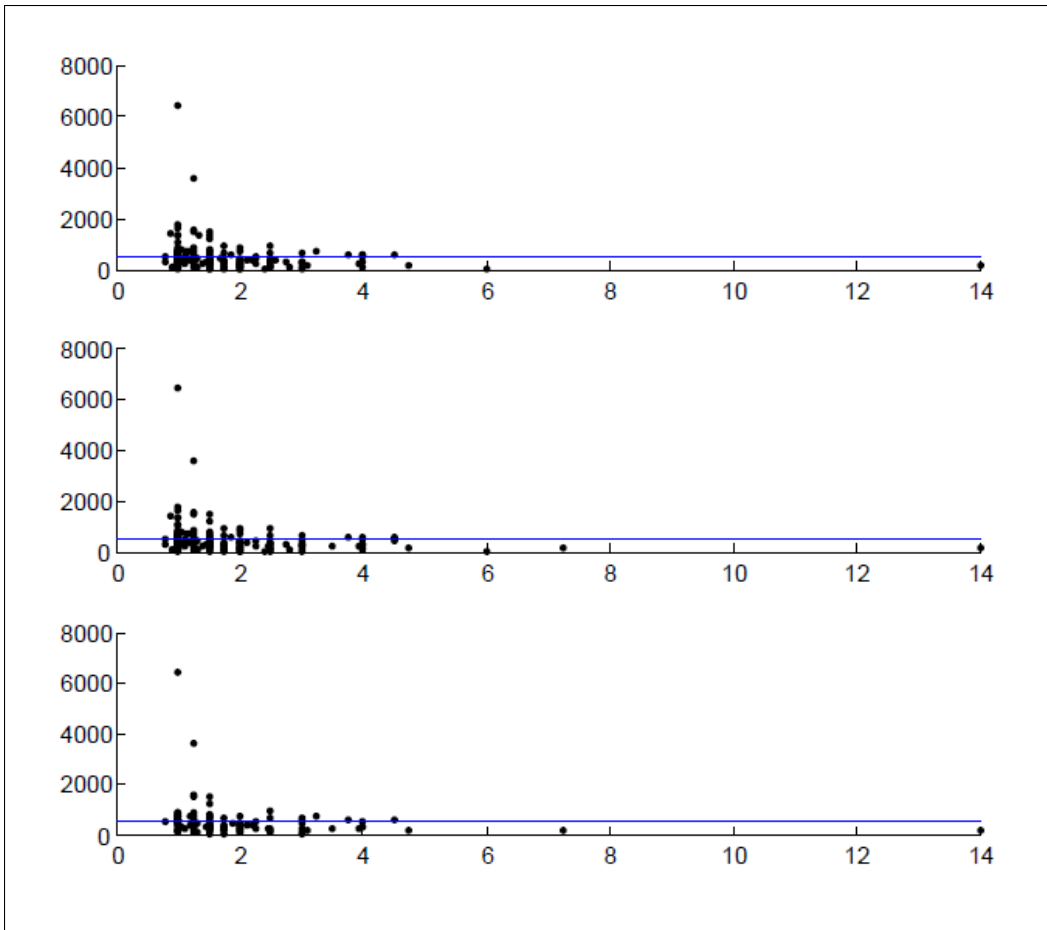


Figure 33. Yield and land size scatter diagrams for farmers with zero land size contributions

Note: Scatter diagrams of observed yields and land sizes from top to bottom under constant, non-increasing, and variable returns to scale for farmers with zero land size contributions.

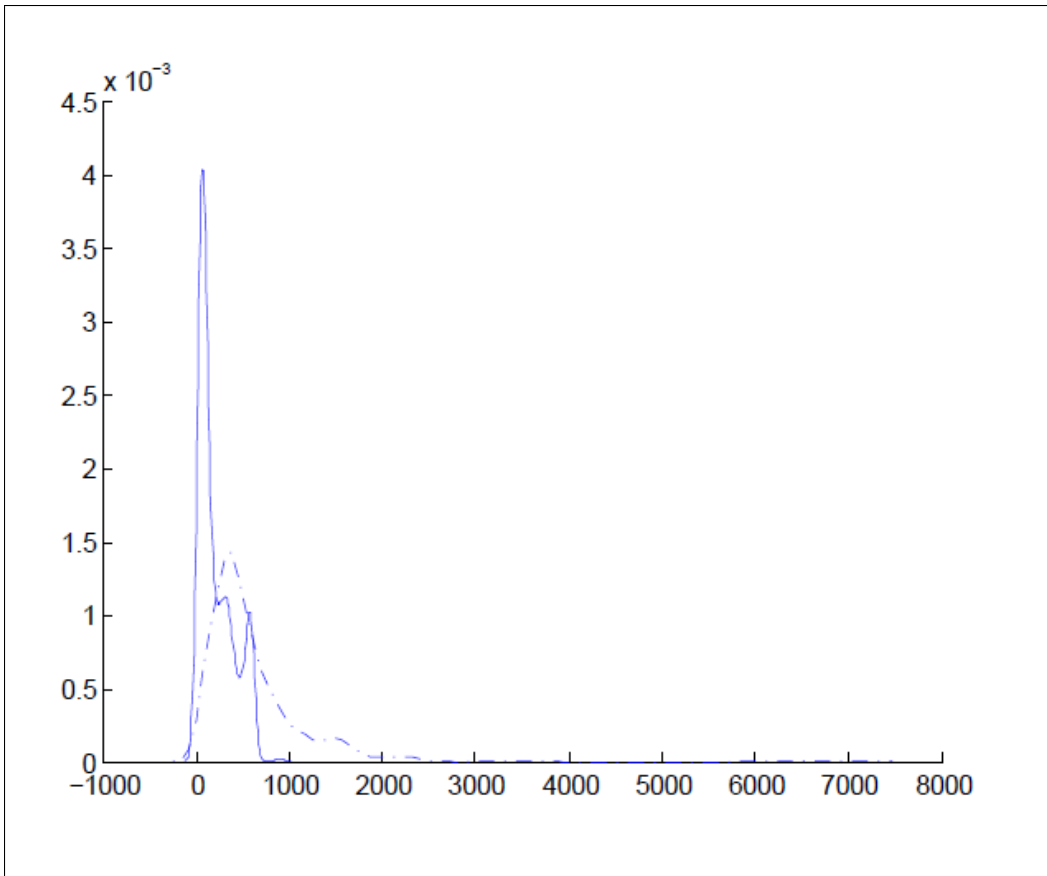


Figure 34. Counter factual distribution y_1^f/l_1 and observed yield distribution

Note: Counter factual distribution y_1^f/l_1 is plotted as a solid line while observed yield distribution is the dashed line.

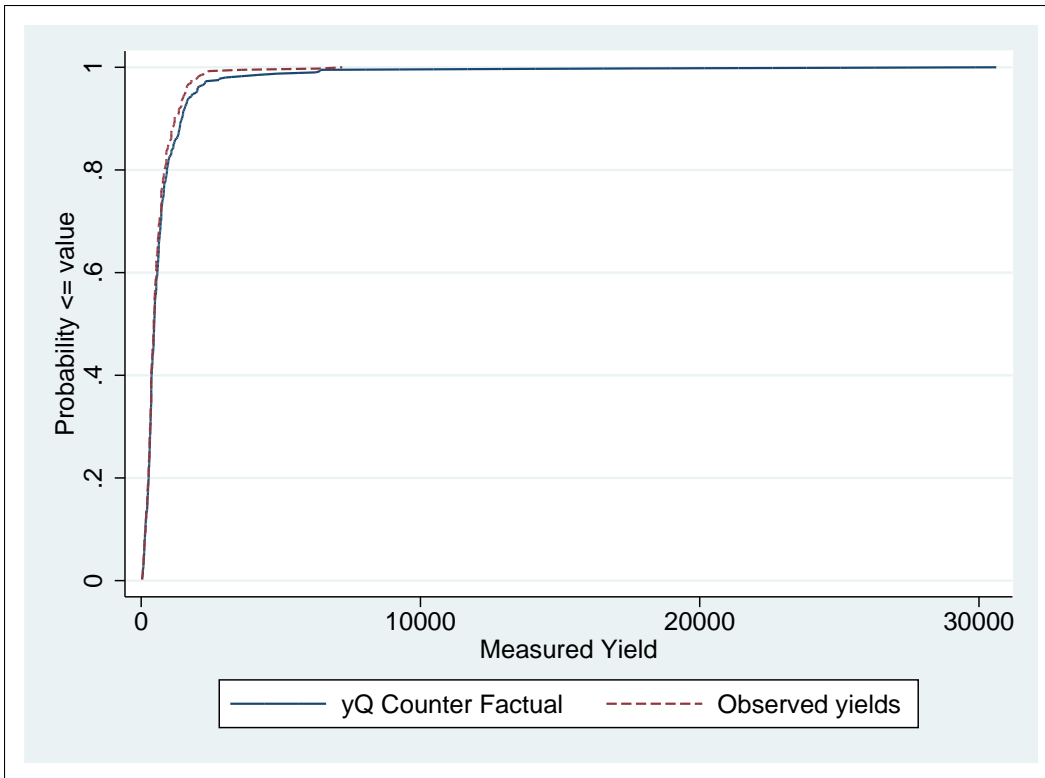


Figure 35. Cumulative empirical distribution of counterfactual distribution y_1^Q/l_1 and observed yield distribution: the effect of not adjusting for land size under variable returns to scale

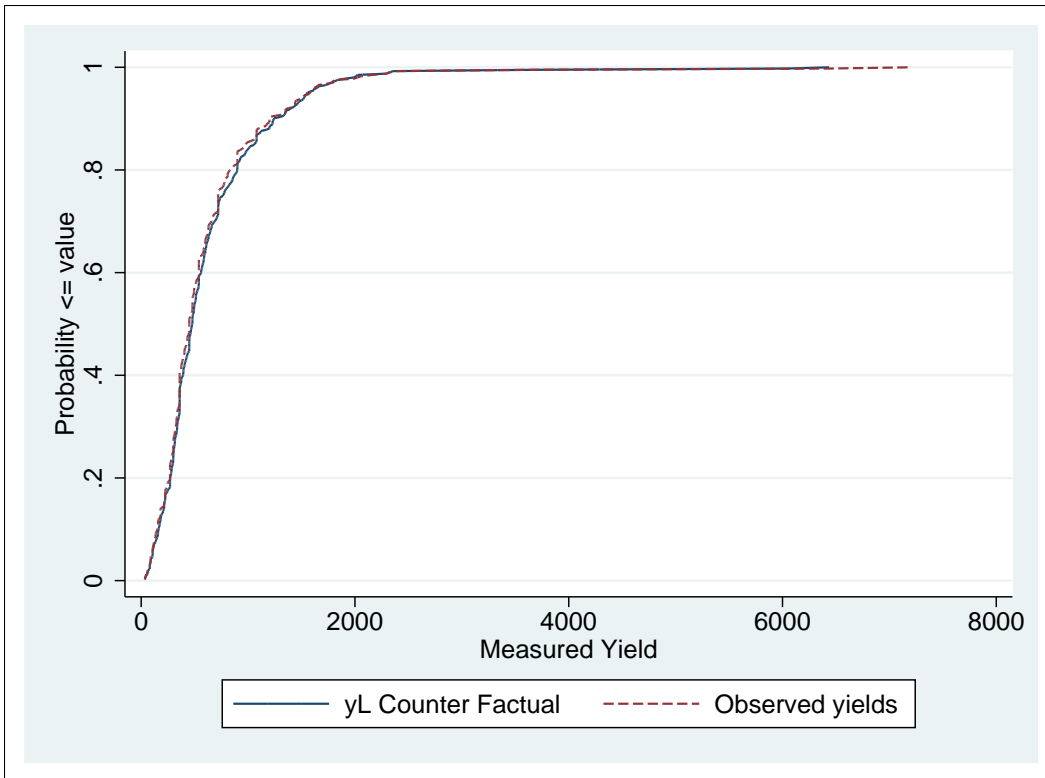


Figure 36. Cumulative empirical distribution of counterfactual distribution y_1^L/l_1 and observed yield distribution: the effect of not adjusting for land quality under variable returns to scale

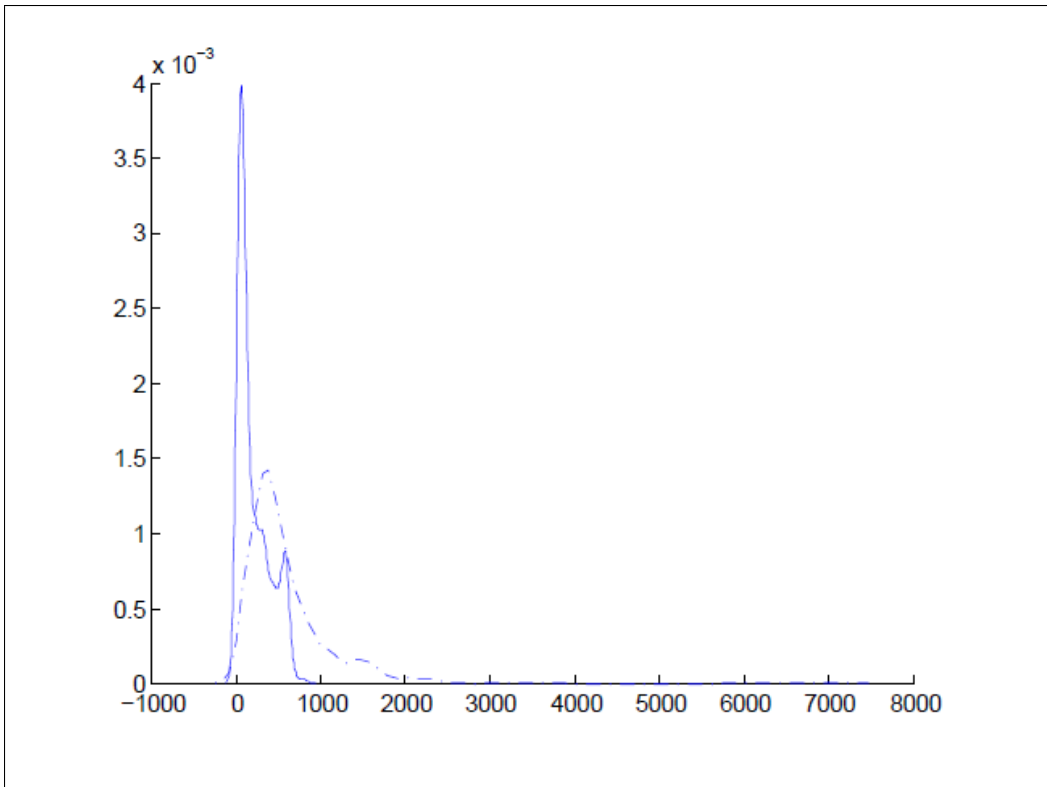


Figure 37. Counterfactual distribution y_1^{QL}/l_1 and observed yield distribution: the effect of not adjusting for efficiency under variable returns to scale

Note: Counterfactual distribution y_1^{QL}/l_1 is plotted as a solid line while observed yield distribution is the dashed line.

Tables

Table 1. Summary statistics of inputs, output, and land quality physical characteristics

Variable	Mean	Std.Dev.	Median	Min	Max
Inputs					
land area (acres)	1.6	1.4	1.25	0.1	14
quantity of seeds (kgs)	13.4	11	10	1	78
inorganic fertilizers (kgs)	46.8	74.4	22	0	650
organic fertilizers (kgs)	742.1	1214.6	300	0	9000
hired labor (cost in KSh)	2935.4	4911.9	1025	0	48160
family labor (hours)	431.5	510.9	287.5	0	4434.8
permanent and share labor (hours)	41.7	95.1	0	0	963
number of hand hoes	3.9	2.2	4	0	15
number of ploughs	0.1	0.3	0	0	2
number of spray-pumps	0.4	0.6	0	0	2
number of sickles	0.4	0.6	0	0	3
milking cows	1	0.9	1	0	5
Output					
total harvest dry maize (kg)	843.7	1122.8	540	0	9000
Land quality physical characteristics					
soil carbon content (% of soil weight)	2.6	1.5	2.18	0.7	15.2
soil clay content (% of soil weight)	28.3	3.9	28.6	15.5	44.9
Land quality ordinal index 100th % level					
constant returns to scale	0.9619	0.1697	0.9925	0.006	1.2696
non-increasing returns to scale	0.8204	0.1630	0.8694	0.006	1
variable returns to scale	0.7816	0.1906	0.8318	0	1
Observations	452				

Table 6. Summary statistics of land size contribution to yield difference under constant returns to scale in correspondence of different percentiles of reference vectors of inputs and outputs

Ref. percentile	Min	10%	20%	30%	40%	Median	60%	70%	80%	90%	Max	Mean	St.Dev.
31	-75.8212	-22.7953	-14.5604	-1.03219	0	0	0	0	1.490522	8.145222	183.9652	-2.37149	22.10468
32	-75.8226	-22.5544	-13.3592	-1.0342	0	0	0	0.000735	1.256817	7.398089	184.4233	-2.32423	22.12715
33	-75.8166	-22.384	-13.3311	-0.99672	0	0	0	0	0.873523	7.153378	185.1689	-2.39363	22.11403
34	-75.8172	-22.4325	-13.3482	-1.00404	0	0	0	0	0.933372	7.205832	185.417	-2.40363	22.12164
35	-75.8177	-22.482	-13.3509	-1.01173	0	0	0	0	0.965311	7.228081	185.5952	-2.40466	22.1265
36	-75.8311	-23.0486	-14.0002	-1.07765	0	0	0	0	0.762713	6.695958	193.4823	-2.63077	22.31757
37	-75.833	-23.058	-13.8826	-1.08632	0	0	0	0	0.780436	6.733732	194.1807	-2.65765	22.3373
38	-75.8339	-23.0593	-13.7775	-1.09108	0	0	0	0	0.710318	6.276778	194.8556	-2.71442	22.32782
39	-75.834	-23.0582	-13.7437	-1.08812	0	0	0	0	0.501729	6.276778	194.8723	-2.71353	22.32683
40	-75.8339	-23.0546	-13.878	-1.08567	0	0	0	0	0.522154	6.276778	194.9448	-2.71806	22.33126
41	-75.8325	-23.0476	-13.9031	-1.07698	0	0	0	0	0.369949	6.276778	194.8017	-2.71285	22.32572
42	-75.8375	-23.0693	-13.9788	-1.10529	0	0	0	0	0.319457	6.276778	195.4005	-2.73885	22.33903
43	-75.8391	-23.0663	-14.0356	-1.11553	0	0	0	0	0.309068	6.276778	195.977	-2.73513	22.34377
44	-75.8383	-23.0607	-14.0346	-1.11211	0	0	0	0	0.300195	6.265857	195.7319	-2.8924	22.07753
45	-75.8356	-23.043	-13.9253	-1.09283	0	0	0	0	0.303743	6.265857	195.0892	-2.87434	22.05507
46	-75.8334	-23.0331	-13.9264	-1.08206	0	0	0	0	0.310141	6.265857	194.7284	-2.88326	22.0425
47	-75.8316	-23.0249	-13.919	-1.07592	0	0	0	0	0.31805	6.265857	194.1211	-2.88088	22.02241
48	-75.8307	-23.0286	-13.9496	-1.07255	0	0	0	0	0.321308	6.265857	193.6185	-2.88939	22.00996
49	-75.8299	-23.0245	-13.9638	-1.06763	0	0	0	0	0.327249	6.265857	193.3326	-2.8881	22.00122
50	-75.8294	-23.0223	-13.9694	-1.06437	0	0	0	0	0.330298	6.265857	193.1757	-2.8865	21.99664
51	-75.8272	-23.012	-13.9918	-1.04926	0	0	0	0	0.337058	6.265857	192.4343	-2.88733	21.9696
52	-75.8263	-23.0087	-13.9985	-1.04246	0	0	0	0	0.340226	6.267029	192.1106	-2.88364	21.95942
53	-75.8254	-23.0047	-14.0049	-1.03521	0	0	0	0	0.347909	6.276778	191.8054	-2.87627	21.95116
54	-75.8226	-22.9644	-14.0118	-1.01679	0	0	0	0	0.364015	6.276778	190.8375	-2.85667	21.92254
55	-75.8213	-22.941	-14.0108	-1.00921	0	0	0	0	0.370858	6.276778	190.3851	-2.84753	21.90723
56	-75.8186	-22.7626	-13.9998	-0.99239	0	0	0	0	0.381486	6.29047	189.4241	-2.82869	21.87692
57	-75.8144	-22.5167	-13.9996	-0.97002	0	0	0	0	0.397413	6.379739	188.036	-2.78304	21.8433
58	-75.8113	-22.3668	-14.0012	-0.959	0	0	0	0	0.379351	6.276778	187.3477	-2.78277	21.81827
59	-75.8093	-22.331	-13.9904	-0.95244	0	0	0	0	0.38882	6.276778	186.9393	-2.77342	21.8054
60	-75.8048	-22.5125	-13.7849	-0.94399	0	0	0	0	0.367605	5.284342	185.8478	-3.08292	21.3648
61	-75.8008	-22.3096	-13.7602	-0.93487	0	0	0	0	0.368723	5.298897	184.9326	-3.05188	21.33866
62	-75.8001	-22.3253	-13.4297	-0.92557	0	0	0	0	0.300157	5.198984	184.6757	-3.04562	21.28497
63	-75.8003	-22.3909	-13.4183	-0.92761	0	0	0	0	0.300157	5.198984	184.345	-3.05428	21.2787
64	-75.8077	-22.4638	-13.3644	-0.95774	0	0	0	0	0.290158	5.648593	185.6039	-3.06888	21.3007
65	-75.808	-22.6046	-13.3312	-0.96356	0	0	0	0	0.240132	5.634555	185.2211	-3.08602	21.2882
66	-75.808	-22.6285	-13.3101	-0.96448	0	0	0	0	0.23003	5.623251	184.721	-3.09539	21.27446
67	-75.81	-22.9782	-13.5751	-0.97904	0	0	0	0	0.015921	5.516851	184.2376	-3.01414	21.54356
68	-75.8104	-22.9797	-13.7319	-0.98234	0	0	0	0	0.015921	5.446971	184.4348	-3.02037	21.54719
69	-75.8106	-22.9805	-13.8036	-0.98383	0	0	0	0	0.015921	5.417116	184.306	-3.02466	21.54572
70	-75.8109	-22.9819	-13.9091	-0.98603	0	0	0	0	0.015921	5.377845	184.2993	-3.03033	21.54679
71	-75.8116	-22.9843	-14.0762	-0.99084	0	0	0	0	0.015921	5.288721	184.452	-3.03841	21.5508
72	-75.8121	-22.9738	-14.1573	-0.99479	0	0	0	0	0.015269	5.198984	184.3691	-3.04965	21.54781
73	-75.8131	-22.9791	-14.1612	-1.00092	0	0	0	0	0.012265	5.198984	184.6532	-3.05564	21.55717
74	-75.8131	-22.9797	-14.1612	-1.00097	0	0	0	0	0.012296	5.198984	184.6462	-3.05641	21.55874
75	-75.813	-22.9794	-14.2002	-1.00055	0	0	0	0	0.012235	5.198984	184.6425	-3.05779	21.55902
76	-75.8143	-22.9867	-14.2452	-1.0097	0	0	0	0	0.012618	5.198984	184.7215	-3.07306	21.56164
77	-75.8148	-22.9906	-14.3634	-1.0144	0	0	0	0	0.015921	5.198984	184.667	-3.08703	21.55957
78	-75.8155	-22.9955	-14.3662	-1.01839	0	0	0	0	0.015921	5.198984	184.8113	-3.09348	21.56487
79	-75.817	-23.0041	-14.4177	-1.03075	0	0	0	0	0.016373	5.020374	184.6792	-3.11506	21.55939
80	-75.8171	-23.0053	-14.5296	-1.03147	0	0	0	0	0.017189	5.008489	184.7135	-3.11973	21.56166
81	-75.8169	-23.0053	-14.644	-1.03035	0	0	0	0	0.017858	5.031004	184.7177	-3.12279	21.56218
82	-75.8169	-23.0053	-14.6558	-1.03035	0	0	0	0	0.018118	5.031209	184.7189	-3.12317	21.56239
83	-75.8169	-23.0054	-14.6559	-1.03053	0	0	0	0	0.018512	5.028707	184.7235	-3.12333	21.56258
84	-75.818	-23.0094	-14.6611	-1.03882	0	0	0	0	0.027859	4.97799	184.8929	-3.13331	21.56678
85	-75.8186	-23.0114	-14.6893	-1.04254	0	0	0	0	0.027859	4.97799	185.1408	-3.13745	21.57252
86	-75.8196	-23.0162	-14.8711	-1.04979	0	0	0	0	0.027859	4.97799	185.7473	-3.14793	21.58735
87	-75.8198	-23.0214	-15.0953	-1.05047	0	0	0	0	0.031134	4.97799	185.9692	-3.15535	21.59258
88	-75.8204	-23.0267	-15.121	-1.05694	0	0	0	0	0.045602	4.97799	186.2986	-3.16756	21.59107
89	-75.8221	-23.0369	-15.128	-1.06691	0	0	0	0	0.04999	4.97799	187.0737	-3.18299	21.59891
90	-75.8237	-23.0448	-15.1612	-1.07685	0	0	0	0	0.053209	4.97799	187.8847	-3.19887	21.60543
91	-75.8243	-23.0468	-15.5608	-1.081	0	0	0	0	0.045143	4.97799	188.4766	-3.21493	21.61683
92	-75.8243	-23.0467	-15.7396	-1.081	0	0	0	0	0.042959	4.97799	188.797	-3.22764	21.62193
93	-75.8251	-23.0494	-15.7432	-1.08825	0	0	0	0	0.027859	4.97799	189.2518	-3.23831	21.62096
94	-75.8253	-23.0499	-15.7437	-1.08677	0	0	0	0	0.058535	4.97799	189.6504	-3.24858	21.62916
95	-75.8238	-23.0475	-15.737	-1.07605	0	0	0	0	0.027859	4.97799	189.9326	-3.25786	21.65085
96	-75.8252	-23.0523	-16.0551	-1.08658	0	0	0	0	0.030731	4.97799	190.7498	-3.27905	21.63982
97	-75.8266	-23.0609	-15.9102	-1.09357	0	0	0	0	0.032723	5.126552	191.1718	-3.28614	21.6493
98	-75.8257	-23.0526	-16.821	-1.08941	0	0	0	0	0.027859	5.198984	190.8879	-3.30787	21.61428
99	-75.8247	-23.049	-17.7428	-1.07825	0	0	0	0	0.027859	5.198984	190.5454	-3.31425	21.61192
100	-75.8148	-23.0182	-16.9921	-0.99717	0	0	0	0	0.023868	5.198984	187.9219	-3.18715	21.59904

Table 9. Summary statistics of land size contribution to yield difference under non-increasing returns to scale in correspondence of different percentiles of reference vectors of inputs and outputs

Ref. percentile	Min	10 %	20 %	30 %	40 %	Median	60 %	70 %	80 %	90 %	Max	Mean	St.Dev.
31	-75.8204	-21.4916	-14.5615	-1.04513	0	0	0	0	0.328334	8.143351	196.6479	-2.34972	22.21395
32	-75.8246	-21.3934	-13.2094	-1.02606	0	0	0	0	0.617087	8.157519	196.6479	-2.18325	22.24799
33	-75.829	-21.3309	-13.2255	-1.03614	0	0	0	0	0.219841	7.371395	196.6479	-2.28751	22.24859
34	-75.8307	-21.2994	-13.2314	-1.04447	0	0	0	0	0.243582	7.375145	196.6479	-2.29753	22.25453
35	-75.8311	-21.2632	-13.233	-1.04779	0	0	0	0	0.241054	7.375657	196.6479	-2.30006	22.25405
36	-75.8342	-23.0646	-13.582	-1.08498	0	0	0	0	0.049863	7.374466	196.6479	-2.58138	22.45848
37	-75.8349	-23.069	-13.5469	-1.09101	0	0	0	0	0.070293	7.126792	196.6479	-2.61663	22.46577
38	-75.8348	-23.0654	-13.5156	-1.09544	0	0	0	0	0.018555	6.276778	196.6479	-2.70976	22.4231
39	-75.8348	-23.0637	-13.7083	-1.09278	0	0	0	0	0.06552	6.276778	196.6479	-2.70297	22.42292
40	-75.8337	-23.0568	-13.7885	-1.08588	0	0	0	0	0.054409	6.276778	196.601	-2.70602	22.42352
41	-75.8309	-23.0459	-13.8028	-1.07069	0	0	0	0	0.030768	6.276778	195.89	-2.70213	22.39806
42	-75.8356	-23.0685	-13.9777	-1.09943	0	0	0	0	0.066913	6.276778	195.5402	-2.7233	22.38348
43	-75.8373	-23.065	-14.0223	-1.10614	0	0	0	0	0.04366	6.276778	196.6479	-2.71063	22.40241
44	-75.8363	-23.0585	-14.0206	-1.10064	0	0	0	0	0.034836	5.454741	196.2602	-2.86583	22.12964
45	-75.8332	-23.0366	-13.8536	-1.07991	0	0	0	0	0.042361	5.480384	196.4711	-2.84082	22.11753
46	-75.8321	-23.0275	-13.8619	-1.07757	0	0	0	0	0.042756	5.573008	196.4668	-2.84721	22.11181
47	-75.8308	-22.9118	-13.8636	-1.06994	0	0	0	0	0.037688	5.66165	196.3612	-2.83765	22.09917
48	-75.83	-22.7713	-13.8633	-1.06462	0	0	0	0	0.036146	5.684626	196.2094	-2.84576	22.10264
49	-75.829	-22.5662	-13.858	-1.05865	0	0	0	0	0.036553	5.692846	195.976	-2.84227	22.09232
50	-75.8285	-22.4749	-13.8545	-1.05579	0	0	0	0	0.036959	5.762518	195.8528	-2.83981	22.08691
51	-75.8263	-22.4494	-13.845	-1.04208	0	0	0	0	0.025344	6.196833	195.2446	-2.82708	22.05831
52	-75.8257	-22.458	-13.8252	-1.03708	0	0	0	0	0.018309	6.276778	195.0832	-2.81976	22.051
53	-75.8244	-22.4663	-13.8157	-1.02956	0	0	0	0	0.016729	6.275785	194.71	-2.81128	22.03679
54	-75.8215	-22.5261	-13.8014	-1.01217	0	0	0	0	0.018309	6.251428	193.8046	-2.78812	22.00314
55	-75.82	-22.5536	-13.8342	-1.00213	0	0	0	0	0.018309	6.196308	193.311	-2.77622	21.98466
56	-75.8171	-22.3072	-13.8062	-0.9834	0	0	0	0	0.018309	6.114593	193.0993	-2.75038	21.96588
57	-75.8135	-22.0179	-13.7941	-0.96176	0	0	0	0	0.021042	6.276778	192.8052	-2.71511	21.94755
58	-75.8108	-21.8217	-13.7869	-0.94982	0	0	0	0	0.02113	6.276778	192.7295	-2.69312	21.93742
59	-75.8087	-21.6925	-13.7584	-0.94256	0	0	0	0	0.020728	6.276778	192.7049	-2.67869	21.93106
60	-75.8042	-21.0794	-14.0562	-0.92769	0	0	0	0	0.025282	6.276778	192.5711	-2.63924	21.90891
61	-75.7999	-20.849	-13.9728	-0.86381	0	0	0	0	0.031573	6.335232	192.5337	-2.60307	21.89698
62	-75.7994	-21.3676	-13.9847	-0.88471	0	0	0	0	0.02217	6.377788	191.5751	-2.61364	21.88273
63	-75.7997	-21.4435	-13.9885	-0.89753	0	0	0	0	0.021813	6.382039	191.4304	-2.61995	21.88367
64	-75.807	-21.7154	-13.9604	-0.94251	0	0	0	0	0.028986	6.41425	192.6076	-2.67562	21.90637
65	-75.8081	-22.0858	-13.9031	-0.94846	0	0	0	0	0.018143	6.405903	192.5168	-2.69218	21.90843
66	-75.8087	-22.2394	-13.8797	-0.95155	0	0	0	0	0.013744	6.402045	192.4905	-2.70614	21.90784
67	-75.8128	-22.6223	-13.8306	-0.97587	0	0	0	0	0	5.453126	193.685	-2.95051	21.73596
68	-75.8139	-22.6682	-13.4858	-0.98373	0	0	0	0	0	5.392213	194.0337	-2.95348	21.76387
69	-75.8144	-22.777	-13.2396	-0.9863	0	0	0	0	0.003939	5.352411	194.1875	-2.953	21.7667
70	-75.8154	-22.8802	-13.0779	-0.98951	0	0	0	0	0.007896	5.287229	194.436	-2.96085	21.77645
71	-75.816	-22.977	-13.0799	-0.99323	0	0	0	0	0.00733	5.217981	194.6465	-2.96921	21.78122
72	-75.8386	-23.0968	-19.019	-1.12151	0	0	0	0	0	4.97799	194.8866	-3.32868	21.95961
73	-75.8176	-22.9856	-13.201	-1.0043	0	0	0	0	0.004924	5.211115	195.2082	-2.98667	21.79775
74	-75.8177	-22.9863	-13.2199	-1.00455	0	0	0	0	0.007435	5.211333	195.2261	-2.9875	21.79965
75	-75.8177	-22.9871	-13.2318	-1.00455	0	0	0	0	0.007436	5.211153	195.217	-2.98998	21.7998
76	-75.8194	-22.9931	-13.2812	-1.01586	0	0	0	0	0.014636	5.206796	195.7908	-3.00253	21.81311
77	-75.8212	-22.9987	-13.4146	-1.02581	0	0	0	0	0.012409	5.204171	196.3323	-3.01475	21.82716
78	-75.8221	-23.002	-13.4184	-1.03277	0	0	0	0	0.016585	5.161648	196.6479	-3.01871	21.83713
79	-75.8248	-23.0109	-13.4773	-1.05168	0	0	0	0	0.007786	5.199483	196.6479	-3.03757	21.83908
80	-75.8255	-23.0127	-13.6028	-1.05478	0	0	0	0	0.006792	5.199907	196.6479	-3.04453	21.84094
81	-75.8258	-23.013	-13.7311	-1.05529	0	0	0	0	0.003838	5.200193	196.6479	-3.05078	21.84105
82	-75.8258	-23.0131	-13.7448	-1.05542	0	0	0	0	0.00362	5.200089	196.6479	-3.05162	21.84108
83	-75.8258	-23.0132	-13.7469	-1.05561	0	0	0	0	0.003501	5.199944	196.6479	-3.0519	21.84136
84	-75.8273	-23.0183	-13.753	-1.06648	0	0	0	0	0.005032	5.163049	196.6479	-3.06059	21.84224
85	-75.828	-23.0206	-13.786	-1.07139	0	0	0	0	0.009286	5.142199	196.6479	-3.06509	21.84208
86	-75.8306	-23.0529	-13.9832	-1.08139	0	0	0	0	0.017223	5.141475	196.6479	-3.07915	21.8472
87	-75.8317	-23.0563	-14.0531	-1.08581	0	0	0	0	0.01557	5.084079	196.6479	-3.09266	21.84769
88	-75.8335	-23.0653	-14.0743	-1.09784	0	0	0	0	0.015005	4.989345	196.6479	-3.10238	21.84578
89	-75.8363	-23.08	-14.0885	-1.11203	0	0	0	0	0.016296	4.98296	196.6479	-3.11504	21.84326
90	-75.8382	-23.0891	-14.1349	-1.12151	0	0	0	0	0.00837	4.927395	196.6479	-3.11951	21.81382
91	-75.8386	-23.0913	-14.2761	-1.12151	0	0	0	0	0.002433	4.87494	196.6479	-3.14396	21.81022
92	-75.8386	-23.0937	-14.332	-1.12151	0	0	0	0	0.004713	4.822737	196.6479	-3.16567	21.8045
93	-75.8386	-23.0937	-14.3687	-1.12151	0	0	0	0	0.009122	4.735602	196.6479	-3.16881	21.79942
94	-75.8386	-23.0961	-14.3218	-1.12151	0	0	0	0	0.00577	4.668985	196.6479	-3.18764	21.79187
95	-75.8386	-23.0968	-16.3071	-1.12151	0	0	0	0	2.22E-14	4.80745	196.6479	-3.24436	21.8066
96	-75.8386	-23.0968	-16.4924	-1.12151	0	0	0	0	0.00134	4.816078	196.6479	-3.25486	21.78282
97	-75.8386	-23.0968	-17.5131	-1.12151	0	0	0	0	0.001414	4.807302	196.6479	-3.27582	21.76344
98	-75.8386	-23.0968	-17.5324	-1.12151	0	0	0	0	0.002478	4.714627	196.6479	-3.27985	21.74613
99	-75.8386	-23.0968	-17.5324	-1.12151	0	0	0	0	0.00407	4.700253	196.6479	-3.27238	21.75133
100	-75.8386	-23.0968	-17.5324	-1.12151	0	0	0	0	0.007375	4.803176	196.6479	-3.25606	21.72564

Table 12. Summary statistics of land size contribution to yield difference under variable returns to scale in correspondence of different percentiles of reference vectors of inputs and outputs

Ref. percentile	Min	10%	20%	30%	40%	Median	60%	70%	80%	90%	Max	Mean	St.Dev.
31	-86.2322	-6.63908	-2.23892	0	0	0	0	0	0.539429	8.273663	217.3474	0.346979	17.07911
32	-85.6027	-9.86747	-2.89623	-0.00068	0	0	0	0.025473	1.285431	8.610118	250.5537	0.019169	19.55267
33	-100	-15.8383	-6.43586	-0.43473	-0.15716	0	0	0	0.751009	7.774249	273.7238	-2.01541	23.4606
34	-100	-16.5834	-7.5668	-0.62104	-0.20794	0	0	0.003417	0.705421	7.978218	272.4398	-2.01279	23.14941
35	-100	-16.7465	-7.73579	-0.6536	-0.17413	0	0	0.006683	0.765968	7.978218	271.8382	-2.08424	23.1722
36	-100	-36.4112	-14.8092	-1.64091	0	0	0	0	0.822813	7.18194	318.8213	-7.25669	32.66301
37	-100	-33.6254	-14.4919	-1.64091	0	0	0	0	0.684846	7.315341	318.8635	-7.59697	33.11537
38	-100	-33.0118	-13.1577	-1.70589	-0.06606	0	0	0	0.656835	7.430752	319.423	-7.42999	33.10915
39	-100	-32.7488	-12.9428	-1.73625	-0.10138	0	0	0.002504	0.700346	7.857657	319.8392	-7.18033	32.97475
40	-100	-32.177	-12.9444	-1.65478	0	0	0	0.000275	0.75443	7.864747	319.8747	-6.87147	32.35764
41	-100	-30.593	-12.688	-1.78538	0	0	0	0	0.754046	7.451363	319.6151	-6.22448	31.09535
42	-100	-30.428	-11.9831	-1.78538	0	0	0	0.001501	0.812103	8.78352	319.8009	-5.63226	30.96877
43	-100	-30.189	-11.9116	-1.78538	0	0	0	0	0.798933	8.78352	319.7623	-5.52852	30.36912
44	-100	-30.3238	-11.9785	-1.87232	-0.12905	0	0	0.0062	0.683413	7.419919	104.2582	-6.37626	24.56085
45	-100	-30.3887	-12.1313	-1.92497	-0.20433	0	0	0.008371	0.695145	7.731624	103.6977	-6.23018	24.08075
46	-100	-29.7446	-9.6813	-0.54596	0	0	0	0	0.468565	7.112683	102.3223	-5.95794	24.03765
47	-100	-26.429	-11.2184	-0.60001	0	0	0	0	0.479735	7.317288	102.241	-5.72086	23.54232
48	-100	-26.0868	-11.1615	-0.53518	0	0	0	0	0.474392	7.656955	102.144	-5.62025	23.55577
49	-100	-25.7256	-11.1472	-0.51656	0	0	0	0	0.423937	7.117791	102.0047	-5.56565	23.46931
50	-100	-25.6037	-11.1465	-0.50377	0	0	0	0	0.420758	7.125361	102.0047	-5.53669	23.43032
51	-100	-28.3883	-12.0285	-0.61652	0	0	0	0	0.324485	7.122473	102.0047	-5.76109	23.71936
52	-100	-28.645	-12.2419	-0.62393	0	0	0	0	0.377691	6.878073	109.638	-5.76051	23.74167
53	-100	-28.4095	-12.0106	-0.62317	0	0	0	0	0.378355	6.869535	109.638	-5.7415	23.71324
54	-100	-27.8724	-11.8669	-0.62498	0	0	0	0	0.345224	6.86855	109.638	-5.7001	23.66188
55	-100	-25.535	-12.1461	-1.04312	0	0	0	0	0.273385	7.116047	109.638	-5.44116	22.92228
56	-100	-25.299	-12.0145	-0.62786	0	0	0	0	0.265487	7.114813	109.638	-5.27735	22.54898
57	-100	-25.087	-11.9675	-0.8801	0	0	0	0	0.275985	7.154866	109.638	-5.24141	22.52591
58	-100	-26.5932	-11.8838	-1.08879	0	0	0	0	0.270595	6.974647	109.638	-5.31899	22.56504
59	-100	-22.3277	-7.38359	0	0	0	0	0	0.028059	3.068636	91.25484	-4.20841	17.24694
60	-100	-23.6518	-8.8995	-0.00989	0	0	0	0	0.033344	3.075817	90.79714	-4.68496	17.68881
61	-100	-23.3558	-8.88773	-0.00317	0	0	0	0	0.038298	2.937494	90.51217	-4.64434	17.6187
62	-100	-24.1535	-9.58337	-0.01563	0	0	0	0	0.032004	2.916889	90.48344	-4.68358	17.65877
63	-100	-24.2009	-9.5721	-0.02575	0	0	0	0	0.03712	3.226025	90.29227	-4.58238	17.8151
64	-100	-22.8007	-9.11889	-0.09045	0	0	0	0	0.236539	3.492877	89.22319	-4.3365	18.03346
65	-100	-22.9046	-9.50742	-0.2474	0	0	0	0	0.272579	3.874074	88.87562	-4.37284	18.08129
66	-100	-23.6368	-10.24	-0.38549	0	0	0	0	0.066553	3.414296	88.78868	-4.68569	18.21905
67	-100	-26.0258	-12.1072	-0.71628	0	0	0	0	0.091988	5.003512	89.18003	-5.11785	18.83622
68	-100	-25.5263	-11.8298	-0.73756	0	0	0	0	0.084297	4.467375	90.19728	-5.06849	18.7142
69	-100	-24.3257	-11.2363	-0.50465	0	0	0	0	0.082417	4.062927	90.36757	-4.77132	18.30826
70	-100	-23.8425	-10.8063	-0.60518	0	0	0	0	0.09829	3.827399	91.71974	-4.68169	18.15386
71	-100	-23.7063	-9.75141	-0.25161	0	0	0	0	0.101734	3.800277	92.28768	-4.61958	18.10958
72	-100	-23.4506	-10.7231	-0.6312	0	0	0	0	0.105589	4.270609	92.361	-4.74872	18.612
73	-100	-23.2014	-10.5264	-0.59079	0	0	0	0	0.100888	4.159103	92.43397	-4.48978	18.02335
74	-100	-22.8791	-10.3451	-0.51859	0	0	0	0	0.098113	4.159791	92.63949	-4.42642	17.99344
75	-100	-22.4995	-10.026	-0.39048	0	0	0	0	0.114079	4.734366	92.38836	-4.26251	17.91563
76	-100	-22.3085	-10.0307	-0.34382	0	0	0	0	0.098663	4.789629	91.30389	-4.26747	17.83831
77	-100	-22.097	-10.193	-0.31704	0	0	0	0	0.085857	3.819927	91.13566	-4.27798	17.72282
78	-100	-22.2089	-10.9603	-0.21987	0	0	0	0	0.073937	4.081461	90.15081	-4.67064	18.27818
79	-100	-22.3163	-11.5897	-0.26406	0	0	0	0	0.090961	3.906074	90.02947	-4.74035	18.31556
80	-100	-22.3072	-11.6113	-0.27523	0	0	0	0	0.053167	3.458774	89.14279	-4.7744	18.29875
81	-100	-22.1924	-11.0715	-0.18373	0	0	0	0	0.057705	3.464061	90.00412	-4.67338	18.19026
82	-100	-22.1659	-11.0571	-0.17352	0	0	0	0	0.05647	3.454864	90.12848	-4.67161	18.19263
83	-100	-22.1661	-11.0582	-0.17373	0	0	0	0	0.056468	3.452247	90.13	-4.67471	18.19413
84	-100	-22.1603	-11.1765	-0.17229	0	0	0	0	0.079067	3.250601	89.89535	-4.70142	18.21929
85	-100	-22.1663	-11.6855	-0.17695	0	0	0	0	0.082067	3.138558	89.99506	-4.71609	18.23397
86	-100	-22.4401	-12.0471	-0.39129	0	0	0	0	0.118235	3.226226	88.83326	-4.80003	18.38353
87	-100	-22.4331	-12.0016	-0.64896	0	0	0	0	0.111067	3.453336	88.48078	-4.63457	18.6739
88	-100	-22.5887	-12.078	-0.35607	0	0	0	0	0.119953	3.423741	88.83603	-4.70109	18.68088
89	-100	-22.6608	-12.0621	-0.27451	0	0	0	0	0.104723	3.333063	88.25337	-4.8681	19.12524
90	-100	-23.1494	-12.0403	-0.54205	-0.02447	0	0	0	0.060999	3.591342	87.74283	-5.07805	19.37699
91	-100	-23.3802	-12.958	-0.25337	-0.00941	0	0	0	0.061594	3.617823	87.46837	-5.14074	19.29244
92	-100	-23.4292	-12.9593	-0.26396	-0.00775	0	0	0	0.045981	3.337498	86.1332	-5.17592	19.25106
93	-100	-23.4462	-12.9734	-0.28575	-0.00267	0	0	0	0.046647	3.413399	84.73347	-5.19069	19.25523
94	-100	-23.1867	-12.3591	-0.24327	0	0	0	0	0.021734	3.340624	83.63361	-5.15964	19.22084
95	-100	-24.0735	-13.0302	-0.26117	0	0	0	0	0.008523	3.244198	89.69686	-5.07533	19.56389
96	-100	-24.1234	-13.0302	-0.36822	0	0	0	0	0.008523	3.194495	89.44347	-5.10117	19.51714
97	-100	-24.0621	-13.0316	-0.35045	0	0	0	0	0.008523	3.4365	87.90394	-5.02025	19.06019
98	-100	-24.1147	-13.685	-0.44583	0	0	0	0	0.008523	3.31787	86.21387	-5.12268	19.18436
99	-100	-24.5647	-13.7136	-0.44951	0	0	0	0	0.012579	2.940412	83.07275	-5.11079	19.21809
100	-100	-24.8133	-13.8923	-0.47428	0	0	0	0	0.004962	2.942763	86.76323	-5.28783	18.8497

Table 14. Tests for equality of observed yields distribution and counter factual distributions: test by Li, Maasoumi, and Racine (2009) of integrated squared density difference under different returns to scale at the median level

Null hypothesis	CRS Tn	CRS p-value	NIRS Tn	NIRS p-value	VRS Tn	VRS p-value
$y_1/l_1 = y_1^I/l_1$	72.6824	0.000	58.4876	0.000	92.7071	0.000
$y_1/l_1 = y_1^E/l_1$	-28.0267	0.279	-26.9965	0.322	-60.1376	0.089
$y_1/l_1 = y_1^L/l_1$	-9.2267	0.724	-0.7495	1	0.5935	0.9475
$y_1/l_1 = y_1^O/l_1$	-39.1165	0.354	-39.0964	0.35	4.2780	0.901
$y_1/l_1 = y_1^{IL}/l_1$	72.2510	0.000	69.4647	0.000	89.6234	0.000
$y_1/l_1 = y_1^{IO}/l_1$	75.2886	0.000	60.4098	0.000	62.7195	0.000
$y_1/l_1 = y_1^{OL}/l_1$	78.8345	0.000	73.946	0.000	82.5379	0.000

Note: The tests statistics are Tn. They are performed on the observed yield against counter factual distributions indicated. For constant and non-increasing returns to scale the amount of units useful for this exercise is 443 while for variable returns to scale the amount of units is 403 with reference the median level. Equality is rejected if p-value is smaller than the significance level desired.

Table 15. Tests across returns to scale for equality of distributions (counter factual and observed yields): test by Li, Maasoumi, and Racine (2009) of integrated squared density difference at the median level

Distribution	CRS vs NIRS Tn	CRS vs NIRS p-value	CRS vs VRS Tn	CRS vs VRS p-value	NIRS vs VRS Tn	NIRS vs VRS p-value
y_1/l_1	-0.9707	1	-5.833	0.927	-5.833	0.927
y_1^f/l_1	3.8122	0.1325	-3.9411	0.051	-9.014	0.656
y_1^E/l_1	1.9715	0.925	-59.9877	0.3955	-60.889	0.347
y_1^I/l_1	8.7414	0.763	5.061	0.599	-4.3919	0.862
y_1^O/l_1	-0.3074	0.999	52.5195	0.403	52.5174	0.404
y_1^H/l_1	1.8888	0.1365	-1.8827	0.017	-6.6776	0.699
y_1^{IO}/l_1	2.2412	0.771	2.6959	0.2475	-2.0262	0.897
y_1^{OL}/l_1	0.7957	0.8435	-0.2356	0.218	-3.2456	0.7905

Note: The tests statistics are Tn. For constant and non-increasing returns to scale the amount of units useful for this exercise is 443 while for variable returns to scale the amount of units is 403. Equality is rejected if p-value is smaller than the significance level desired.

Table 16. Tests for equality of observed yields distribution and counter factual distributions: test by Li, Maasoumi, and Racine (2009) of integrated squared density difference under different returns to scale at the mean level

Null hypothesis	CRS Tn	CRS p-value	NIRS Tn	NIRS p-value	VRS Tn	VRS p-value
$y_1/l_1 = y_1^I/l_1$	60.3772	0.000	62.2517	0.000	65.1694	0.000
$y_1/l_1 = y_1^E/l_1$	-26.9913	0.1345	-24.6082	0.055	-58.0400	0.013
$y_1/l_1 = y_1^L/l_1$	-11.3779	0.731	-7.8012	0.6745	-9.6793	0.5095
$y_1/l_1 = y_1^O/l_1$	-39.4264	0.139	-1.2041	0.4955	2.0790	0.582
$y_1/l_1 = y_1^{IL}/l_1$	63.9487	0.000	67.8881	0.000	74.8112	0.000
$y_1/l_1 = y_1^{IO}/l_1$	70.1965	0.000	66.8617	0.000	68.184	0.000
$y_1/l_1 = y_1^{OL}/l_1$	74.5181	0.000	75.3272	0.000	80.6204	0.000

Note: The tests statistics are Tn. They are performed on the observed yield against counter factual distributions indicated. For constant and non-increasing returns to scale the amount of units useful for this exercise is 443 while for variable returns to scale the amount of units is 411. Equality is rejected if p-value is smaller than the significance level desired.