Norwegian University of Life Sciences
School of Economics and Business

Maize Productivity and Input Subsidies in Malawi: A State-Contingent Stochastic Production Frontier Approach

Stein T. Holden and Christopher J. O'Donnell

Working Papers No. 2 / 2015

ISSN: 2464-1561

Maize Productivity and Input Subsidies in Malawi: A State-Contingent Stochastic Production Frontier Approach

S.T. Holden[a] and C. J. O'Donnell[b]

[a] *School of Economics and Business, Norwegian University of Life Sciences, Norway*
[b] *School of Economics, The University of Queensland, Australia*

Abstract

We make cross-sectional comparisons of productivity in a risky agricultural setting. To make meaningful comparisons, we find it necessary to define a new productivity index that satisfies important axioms from index number theory (e.g., transitivity). The index can be computed without any information on output or input prices. However, it cannot be computed without an estimate of a state-contingent production frontier. We use maximum likelihood methods to estimate a state-contingent stochastic production frontier that explicitly allows for variations in input quality. We find that differences in productivity are mainly due to differences in environment and scale-mix efficiency. In turn, we conjecture that differences in scale-mix efficiency are partly driven by variations in access to input subsidies. The maximum likelihood estimator appears to do a poor job of disentangling the effects of technical inefficiency and statistical noise.

Key words: agricultural productivity; risky environment; imperfect factor markets; state-contingent analysis; total factor productivity; input subsidies.

JEL codes: C4; O13; Q1.

---

1. Introduction

Low agricultural productivity has frequently been blamed for high levels of poverty and vulnerability in developing countries in sub-Saharan Africa. The Green Revolution that stimulated agricultural productivity in Asian and Latin-American countries appears to have been less effective in raising productivity in developing countries elsewhere. Population growth and climate change are now placing increased pressure on governments to raise agricultural productivity in Africa. Unfortunately, in African countries, poor infrastructure, high dependence on rain fed production, climatic variability, the spatial nature of production, and imperfect information contribute to pervasive market imperfections (Binswanger and Rosenzweig, 1986). These conditions impose constraints on productivity analysis due to missing and biased price data (Feder, 1985). Most productivity analyses in such environments have thus focused on returns to land or labor without attempting to estimate total factor productivity (TFP). Many studies in the efficiency literature have also failed to take environmental variables into account. This can lead to upward bias in the estimation of frontiers and downward bias in estimates of technical efficiency (O'Donnell and Griffiths 2006). In this paper, we aim to illustrate how the two problems of missing prices and a stochastic production environment can be handled when analysing variations in technical efficiency and TFP.

We develop a general state-contingent stochastic frontier model and apply it in a small farmer risky environmental setting. We break the agricultural production period into two sub-periods: in the first sub-period, the firm chooses inputs (land, labor, seed and fertilizer) in the face of uncertainty about characteristics of the production environment (e.g., rainfall); in the second sub-period, Nature resolves this uncertainty by choosing a value from a set $\Omega = \{1, \ldots, S\}$. Hereafter, the elements of Ω are referred to as *states of Nature* (or simply *states*). Different farmers may experience different states of Nature and may face different input prices depending on their access to subsidised inputs. Access to subsidised inputs stimulates input demand. However, the consequences for productivity are uncertain because such access can affect the efficiency of input use.

To illustrate our modelling approach, we use a cross-section dataset comprising observations on small maize producers in Malawi. This is a convenient illustration because maize is the main crop produced by almost all rural farms. Maize is also the most important staple food for rural as well as for urban consumers in the country. Maize production has therefore for many years been stimulated by a national input subsidy program aiming to improve household and national food security. Very high subsidy levels are used in the program causing large variations in factor prices depending on access to targeted subsidies. The high costs of the program in combination with the vulnerability of maize production to droughts makes the analysis of maize productivity under different states of Nature (drought, no drought) and access to subsidized inputs highly relevant.

The paper is structured as follows. In Part 2 the technology assumptions are specified. A general productivity index is presented in Part 3 and a new way of decomposing it is presented in Part 4. Firm behaviour is discussed in Part 5 before the econometric model and estimation method are explained in Part 6. The data are described in Part 7. We present and discuss the main results in Part 8 before we briefly conclude in Part 9.

2. Technologies and Metatechnologies

In O'Donnell (2015), a technology is defined as "a technique, method or system for transforming inputs into outputs. .... For all practical intents and purposes, it is convenient to think of a technology as a book of instructions." Furthermore, "the set of technologies available a given period is referred to as a *metatechnology*. If we think of a technology as a book of instructions, then we should think of a metatechnology as a library." In this paper, we analyse cross-section data, so we are dealing with a single metatechnology.

Metatechnologies can be represented using various sets and functions. In this paper, we represent the metatechnology using state-contingent output sets and distance functions. These sets and functions are said to be *state-contingent* because they are explicitly conditioned on states of Nature. This is one of the distinguishing features of our paper.

A second distinguishing feature of our paper is the way we deal with variations in input quality. In practice, it is common to deal with this issue by disaggregating inputs into homogeneous groups (e.g., by dividing the number of employees into numbers of skilled and unskilled employees; or by disaggregating the capital input into land, buildings and equipment). One problem with this approach is that the number of inputs can become very large, and this can make it difficult to obtain reliable estimates of economic quantities of interest (in econometric jargon, we get *multicollinearity*). In this paper, we avoid this problem by introducing separate measures of input quality directly into the definitions of state-contingent output sets.

To make these concepts more concrete, let $x = (x_1, \ldots, x_{M^*})'$, $q = (q_1, \ldots, q_{N^*})'$ and $a = (a_1, \ldots, a_{G^*})'$ denote vectors of input quantities, output quantities, and input attributes respectively. In this paper, the set of outputs that can be produced in state s using inputs x having attributes a is formally defined as $P(x, a, s) \equiv \{q : x$ with attributes a can produce q in state $s\}$. This is a state-contingent output set that explicitly accounts for variations in input quality. In this paper, we make assumptions about the metatechnology by way of assumptions about this set. To be specific, we make the following assumptions:

O1 $0 \in P(x,a,s)$ for all $x \geq 0$ (inactivity);

O2 $P(x,a,s)$ is bounded for all $x \geq 0$;

O3 $q \geq 0 \Rightarrow q \notin P(0,a,s)$ (weak essentiality; no free lunch);

O4 $q \in P(x,a,s)$ and $0 \leq \tilde{q} \leq q \Rightarrow \tilde{q} \in P(x,a,s)$ (outputs strongly disposable);

O5 $q \in P(x,a,s)$ and $\tilde{x} \geq x \Rightarrow q \in P(\tilde{x},a,s)$ (inputs strongly disposable); and

O6 $P(x,a,s)$ is closed for all $x \geq 0$ and $L(q,a,s) \equiv \{x : q \in P(x,a,s)\}$ is closed for all $q \geq 0$.

If these assumptions are true, then equivalent representations of $P(x,a,s)$ include (state-contingent) output and input distance functions. In this paper, the output distance function is formally defined as $D_O(x,q,a,s) \equiv \inf\{\rho > 0 : q/\rho \in P(x,a,s)\}$. The input distance function is defined as $D_I(x,q,a,s) \equiv \sup\{\theta > 0 : q \in P(x/\theta,a,s)\}$. If input attributes do not change and there is only one state of Nature, then these distance functions collapse to the distance functions of Shephard (1970, pp. 206, 207).

If assumptions O1–O6 are true, then the output (resp. input) distance function is nonnegative (NN), non-decreasing (ND) and homogeneous of degree one (HD1) in outputs (resp. inputs). As we shall see in Part 3, these three properties mean that distance functions can be used to construct sensible output, input and TFP indexes. Output and input distance functions can also be used for another purpose: the value of the output distance function is a measure of output-oriented technical efficiency (OTE), and the reciprocal of the value of the input distance function is a measure of input-oriented technical efficiency (ITE). These concepts of technical efficiency can be traced back at least as far as Debreu (1951) and Farrell (1957).

Finally, if there is only one *ex post* output, as in this paper, then an equivalent representation of $P(x,a,s)$ is the (state-contingent) production function. In this paper, the production function is formally defined as $F(x,a,s) \equiv 1/D_O(x,1,a,s)$. If assumptions O1–O6 are true, then the production function is NN and ND in inputs. An example is the following Cobb-Douglas (CD)[1] function:

$$F(x,a,s) = \exp\left(\alpha_s + \sum_{g=1}^{G^*} \psi_{gs} a_g + \sum_{m=1}^{M^*} \beta_{ms} \ln x_m\right) \tag{1}$$

where $\beta_{ms} \geq 0$ is a state-contingent output elasticity and $r_s = \sum_m \beta_{ms} \geq 0$ is a state-contingent elasticity of scale. The state-s production function exhibits decreasing returns to scale (DRS), constant returns to scale (CRS) or increasing returns to scale (IRS) as r_s is less than, equal to, or greater than one.

3. A General Productivity Index

It is convenient at this point to introduce firm subscripts into the notation so that, for example, $q_i = (q_{1i}, \ldots, q_{N^*i})'$ now denotes the output vector of firm i. In O'Donnell (2012), an output quantity index that compares q_i with q_k is defined as any variable of the form $QI_{ki} \equiv Q(q_i)/Q(q_k)$ where $Q(.)$ is an NN, ND and HD1 scalar aggregator function. On the input side, an input quantity index that compares x_i with x_k is any variable of the form $XI_{ki} \equiv X(x_i)/X(x_k)$ where $X(.)$ is an NN, ND and HD1 scalar aggregator function. Output and input indexes of this form satisfy important axioms from index number theory, including identity, transitivity and circularity. The transitivity axiom, for example, says that if firm B produces (resp. uses) twice as much as firm A, and firm

[1]The CD terminology derives from the fact that if input attributes do not change, there is only one state of Nature, and there are only two inputs involved in the production process, then (1) collapses to the production function of Cobb and Douglas (1928).

C produces (resp. uses) twice as much as firm B, then firm C produces (resp. uses) four times as much as firm A. Indexes that do not satisfy this axiom include the well-known Fisher and Törnqvist indexes.

Any NN, ND and HD1 aggregator functions can be used for purposes of constructing output and input quantity indexes. If assumptions O1–O6 are true, then the menu of suitable aggregator functions includes $Q(q) \propto D_O(\bar{x}, q, \bar{a}, \bar{s})$ and $X(x) \propto D_I(x, \bar{q}, \bar{a}, \bar{s})$ where $\bar{s}$ is a representative state of Nature, $\bar{q}$ is a representative vector of outputs, and $\bar{x}$ and $\bar{a}$ are representative vectors of input quantities and attributes. The associated output and input quantity indexes are $QI^G_{ki} = D_O(\bar{x}, q_i, \bar{a}, \bar{s})/D_O(\bar{x}, q_k, \bar{a}, \bar{s})$ and $XI^G_{ki} = D_I(x_i, \bar{q}, \bar{a}, \bar{s})/D_I(x_k, \bar{q}, \bar{a}, \bar{s})$. These indexes are general (G) in the sense that they nest several other indexes as special cases. For example, if input attributes do not change and there is only one state of Nature, then they collapses to the output and input quantity indexes defined by Färe and Primont (1995, pp. 36, 38).

Finally, a TFP index is a measure of output change divided by a measure of input change. For example, the index that compares the TFP of firm i with the TFP of firm k is a variable of the form $TFP_{ki} \equiv QI_{ki}/XI_{ki}$. Again, any output and input quantity indexes can be used for purposes of constructing a TFP index. Dividing QI^G_{ki} by XI^G_{ki}, for example, yields the following general TFP index:

$$TFPI^G_{ki} = \frac{D_O(\bar{x}, q_i, \bar{a}, \bar{s})}{D_O(\bar{x}, q_k, \bar{a}, \bar{s})} \frac{D_I(x_k, \bar{q}, \bar{a}, \bar{s})}{D_I(x_i, \bar{q}, \bar{a}, \bar{s})}. \quad (2)$$

Different assumptions concerning metatechnologies have important practical implications for the form of this index. For example, if there is only one *ex post* output involved in the production process and the production function is given by (1), then:

$$TFPI_{ki}^{G} = \left(\frac{q_i}{q_k}\right) \prod_{m=1}^{M^*} \left(\frac{x_{mk}}{x_{mi}}\right)^{\lambda_{m\bar{s}}} \tag{3}$$

where $\lambda_{m\bar{s}} \equiv \beta_{m\bar{s}}/r_{\bar{s}} \geq 0$ and $\sum_m \lambda_{m\bar{s}} = 1$.

4. The Components of Productivity Change

In theory, any TFP index can be decomposed into measures of environmental change and efficiency change.[2] In practice, the number and type of components depends in part on the assumed properties of the metatechnology. For example, if assumptions O1–O6 are true and there is only one *ex post* output involved in the production process, as in this paper, then we can write $q_i = F(x_i, a_i, s)\exp(-u_{si})$ where $u_{si} \equiv -\ln D_O(x_i, q_i, a_i, s) \geq 0$ is a technical inefficiency effect.[3] Similarly, in the case of firm k, we can write $q_k = F(x_k, a_k, r)\exp(-u_{rk})$. Substituting these equations into (3) yields:

$$TFPI_{ki}^{G} = \left[\frac{F(x_i, a_i, s)}{F(x_k, a_k, r)} \prod_{m=1}^{M^*} \left(\frac{x_{mk}}{x_{mi}}\right)^{\lambda_{m\bar{s}}}\right] \left[\frac{\exp(-u_{si})}{\exp(-u_{rk})}\right]. \tag{4}$$

The first term in brackets on the right-hand side of this equation is an environment and output-oriented scale-mix efficiency index (EOSMEI). The second term is an output-oriented technical efficiency index (OTEI). Whether or not the EOSMEI component can be further decomposed into separate measures of environmental change, scale efficiency change and/or mix efficiency change depends on whether the production function is

[2] In a time-series or panel data context, any TFP index can be decomposed into a measure of technical change as well as measures of environmental change and efficiency change. Details can be accessed from O'Donnell (2015).

[3] If $N^* = 1$, then $D_O(x_i, q_i, a_i, s) = q_i D_O(x_i, 1, a_i, s) = q_i/F(x_i, a_i, s) \Rightarrow q_i = F(x_i, a_i, s) D_O(x_i, 1, a_i, s) \Rightarrow q_i = F(x_i, a_i, s)\exp(-u_{si})$.

multiplicatively separable in inputs, input attributes and state variables.

5. Firm Behaviour

It is useful at this point to say something about firm behaviour. In this paper, we suppose that firms maximise welfare. The exact form of the welfare-maximisation problem depends on the variables that the firm can and cannot choose. In the state-contingent literature, it is common to assume that firms can choose inputs and *ex ante* outputs freely, and that prices are known at the time these choices are made (e.g., because they are fixed under a contract). Again suppose there is only one *ex post* output. In this case, the welfare-maximisation problem of firm i is:

$$W(w_i, p_i, a_i) = \max_{x \geq 0, q \geq 0} \{W(p_i q_1 - w_i' x, \ldots, p_i q_S - w_i' x) : q_s \leq F(x, a_i, s) \text{ for all } s \in \Omega\}$$

where w_i is a vector of input prices, p_i is a scalar output price, $q = (q_1, \ldots, q_S)'$ is a vector of *ex ante* outputs, and $W(.)$ is a utility function that is nondecreasing in *ex ante* outputs and nonincreasing in inputs. The input vector that solves this problem is an input demand correspondence of the form $x(w_i, p_i, a_i)$. To say that the firm maximises welfare is to say that $x_i = x(w_i, p_i, a_i)$. Among other things, this means that, if necessary, input and output prices can be used as instrumental variables when estimating production frontiers. Access to subsidized inputs implies lower input prices for those with access and this may stimulate input demand. Ricker-GIlbert et al. (2011) estimated that one extra unit of fertilizer input was associated with 0.78 units increase in total fertilizer use, implying a crowding out effect of 0.22 unit.

6. *Ex Post* Econometrics

If there is only one *ex post* output, then the relationship between the variables involved in the production process can be written as $y_i = \ln F(x_i, a_i, s) - u_{si}$ where $y_i \equiv \ln q_i$ is the *ex post* log-output and $u_{si} \equiv -\ln D_O(x_i, q_i, a_i, s) \geq 0$ is the technical inefficiency effect introduced in Part 4. If the functional form of $F(.)$ is unknown, as in this paper, then we can write $y_i = f(x_i, a_i, s) + v_{si} - u_{si}$ where $f(.)$ is an approximating function of our own choosing and $v_{si} = \ln F(x_i, a_i, s) - f(x_i, a_i, s)$ is an unobserved error that is commonly referred to as statistical noise. In this paper, we choose a CD *approximating* function and write:

$$y_i = \alpha_s + \sum_{g=1}^{G} \psi_{gs} a_{gi} + \sum_{m=1}^{M} \beta_{ms} \ln x_{mi} + v_{si} - u_{si}. \tag{5}$$

Here, v_{si} accounts for functional form errors [i.e., the possibility that the true production function is not, in fact, given by (1)], omitted variables (i.e., the possibility that $G \leq G^*$ and/or $M \leq M^*$) and other sources of noise (e.g., errors in the measurement of output). In this paper, to estimate the unknown parameters in (5), we make the following assumptions:[4]

A1 v_{si} is an independent $N(0, \sigma_v^2)$ random variable, and

A2 u_{si} is an independent $N^+(0, \sigma_u^2)$ random variable.

If A1 and A2 are true, then ordinary least squares (OLS) estimators of the slope parameters in (5) are consistent, and maximum likelihood (ML) estimators of the *intercept and* slope parameters are consistent, asymptotically normal and asymptotically efficient.

[4]In this context, the term *independent* means, *inter alia*, that $Cov(a_{gi}, v_{si} - u_{si}) = Cov(\ln x_{mi}, v_{si} - u_{si}) = 0$ for all g, s, m, and i. If $Cov(a_{gi}, v_{si} - u_{si}) \neq 0$ and/or $Cov(\ln x_{mi}, v_{si} - u_{si}) \neq 0$, then the explanatory variables are said to be *endogenous*.

Finally, if the relationship between the variables involved in the production process is written as $y_i = f(x_i, a_i, s) + v_{si} - u_{si}$, then the TFP index (3) can be decomposed as:

$$TFPI_{ki}^{G} = \left[\frac{\exp[f(x_i, a_i, s)]}{\exp[f(x_k, a_k, r)]} \prod_{m=1}^{M^*} \left(\frac{x_{mk}}{x_{mi}} \right)^{\lambda_{m\bar{s}}} \right] \left[\frac{\exp(-u_{si})}{\exp(-u_{rk})} \right] \left[\frac{\exp(v_{si})}{\exp(v_{rk})} \right]. \tag{6}$$

The first term in brackets on the right-hand side of this equation is still an environment and output-oriented scale-mix efficiency index (EOSMEI). The second term is still an output-oriented technical efficiency index (OTEI). The last term is a statistical noise index (SNI). Thus, (6) says that TFPI = EOSMEI $\times$ OTEI $\times$ SNI. If assumptions A1 and A2 are true, then (i) a consistent estimator of the TFP index itself can be obtained by replacing the unknown state-contingent output elasticities with their ML estimators, (ii) a consistent estimator of the OTEI component can be obtained using the OTE estimator of Jondrow et al. (1982), (iii) a consistent estimator of the SNI component can be obtained by netting the OTEI component out of the (change in) the ML residuals, and (iv) a consistent estimator of the EOSMEI component can be obtained using the fact that EOSMEI = TFPI/(OTEI $\times$ SNI). If $\sigma_v^2 = 0$, then there is no statistical noise and (6) collapses to (4).

7. Data

The data come from a household-farm survey in Malawi. Stratified random sampling was used in six districts in Central and Southern Regions of the country including some of the most densely populated rural areas in the country in the south, with varying market access and tenure systems (Lunduka, 2009). While the population has been surveyed four times (2006, 2007, 2009 and 2012), we only use the data for 2012 as an illustrative example. One advantage of these data is the strong dominance of a single crop, maize, in the farming system. Maize is also the dominant staple food of the rural small-scale producers, many of whom are net buyers of maize due to the small size of farms. A second advantage of these data is that maize area and farm size were measured with GPS, so errors in the measurement of the land input should be small. A third advantage is that many households in the sample were exposed to a quite severe drought in form of a dry spell in the early rainy season in 2011/12, making a state-contingent frontier analysis relevant.

In the process of assembling the data for econometric analysis, some outlier observations, some missing input data, and some inconsistencies in the data were discovered. Consequently, 15 observations were deleted from the dataset, leaving 287 observations for the final analysis. There were $S = 2$ states of Nature: a normal state (227 observations) and a drought state (60 observations). The variables used in the analysis were: q = total maize output (kg); x_1 = land (ha); x_2 = labor (days); x_3 = seed (kg) (all varieties, including seed used for replanting); x_4 = fertilizer (kg); a_1 = land slope index; and a_2 = soil fertility index. In this dataset, lower values of the soil fertility index indicate higher levels of fertility. Descriptive statistics for these variables are reported in Table 1.

8. Results and Discussion

We used ML to estimate the state-contingent stochastic frontier model given by (5) and assumptions A1 and A2. A Hausman test was use to test the null hypothesis that the explanatory variables are uncorrelated with the composite error term (the discussion in Part 5 led us to use output and input prices as instruments). We found no evidence that the explanatory variables are endogenous (the null hypothesis could not be rejected at the 10% level of significance). Assumption O5 (strong disposability of inputs) implies that the production function is ND in inputs, so we constrained all output elasticities to be nonnegative. Inequality-constrained ML estimates of the unknown parameters are reported in Table 2. The estimates of β_{31} and β_{22} are zero, with standard errors of zero, because the inequality constraints on these parameters are binding (this is a well-known shortcoming of the sampling theory approach to inference).

Table 2 shows that the land area is the most important determinant of maize output. Land scarcity is also therefore the most important explanation for small farm production. The average maize farm size is only 0.72 ha (Table 1). Fertilizer application also contributes to a substantial increase in output in a normal year but less so in a drought year as seen by the smaller coefficient (estimated output elasticity). The impact of the drought is particularly evident when comparing the estimated intercept terms in the normal year and drought year, indicating that the state of Nature can give rise to a large change in expected output in addition to changes in output elasticities.

The other estimates reported in Table 2 are plausible insofar as they indicate the following: returns to scale in a normal (resp. drought) season are greater (resp. less) than one; more land, more labor, more seed, more fertilizer and higher levels of soil fertility all lead to more output; and the estimate of $\lambda \equiv \sigma_u/\sigma_v$ is significantly different from zero indicating that there is technical inefficiency in this dataset. The fact that some coefficients are statistically insignificant indicates that our approach to dealing with input quality may not have fully eliminated the multicollinearity problem.

Selected indexes of firm performance are reported in Table 3. These indexes compare the performance of a subset of farm households with the performance of household 300 (hereafter HH 300). The results in Table 3 have been ordered from the most productive household in the sample to the least productive. Thus, the most productive household in the sample is HH 389, and the least productive is HH 74. The interpretation of the indexes is straightforward. For example, the results in the first few columns of the first row indicate that HH 389 produced 6.2% less output and used 82.4% less input than HH 300 (QI $= 1 - 0.052 = 0.938$ and XI $= 1 - 0.824 = 0.176$). Consequently, HH 389 was found to be 5.3 times more productive than HH 300 (TFPI $=$ QI$/$XI $= 0.938/0.176 = 5.319$). The last three columns in Table 3 attribute differences in TFP to differences in efficiency and statistical noise. Again, the interpretation of these results is straightforward. For example, the entries in the first row indicate that HH 389 was 5.3 times more productive than HH 300 due to the combined effects of (i) technical efficiency (TI $= 1.950 > 1 \Rightarrow$ a positive effect), (ii) the production environment and/or economies of scale and scope (EOSMEI $= 11.544 \gg 1 \Rightarrow$ a large positive effect), and (iii) other factors that we cannot identify (SNI $= 0.236 \ll 1 \Rightarrow$ a large negative effect). Finally, the indexes reported in Table 3 are transitive, which means it is meaningful to make comparisons across rows and columns. For example, the first and last entries entries in the TFPI column indicate that HH 389 was more than 85 times more productive than HH 74 (TFPI $= 5.319/0.062 = 85.86$).

To obtain further insights into the sources of TFP change, we will look more closely at selected households reported at the top, middle and bottom of Table 3. We start with HH 300 (the reference, or benchmark, household). This household produced 800 kg of maize (improved variety) on an area of 0.73 ha, using 49 days of labor (= 67 days/ha in hoe-based farming), 2 kg of seed (= 2.7 kg/ha), and 80 kg of fertiliser (= 109.4 kg/ha) in a normal season (no drought) on land with slope index = 1 and fertility index = 1. Half of the fertilizer (40 kg) was obtained through the subsidy program. Among the highly productive households, HH 389 produced 750 kg of maize (OPV) on an area of 0.21 ha, using 50 kg fertilizer obtained through the subsidy program. This farm household was land-poor and a net buyer of maize and had casual off-farm jobs (ganyu) to make ends meet. HH 289 is another highly productive household producing 1800 kg of hybrid maize on 0.66 ha (slightly smaller than the benchmark), using 51 days of labor (= 77.8 days/ha, similar to the benchmark), 3 kg of seed (= 4.6 kg/ha, much higher than the benchmark), and 50 kg of fertiliser (= 76.2 kg/ha, much lower than the benchmark) in a normal season (same as the benchmark) on land with slope index = 1 (same as the benchmark) and fertility index = 2 (less fertile than the benchmark). Thus, EOSMEI = 2.737 because of the combined effects of different input mix (higher seeding rate and lower fertiliser rate = a positive effect) and less fertile soil (a negative effect). This household was a net seller of maize but also had casual off-farm jobs to make ends meet.

Households 389 and 289 obtained good maize yields partly because they were located in places that were not badly affected by the 2011/12 drought. On the other hand, HH 77 was affected by the drought and the maize had to be replanted using a drought tolerant short duration maize variety. HH 77 produced 1500 kg of maize on 0.44 ha using 100 kg fertilizer of which 50 kg was received through the subsidy program. The household had a treadle pump and could thus to some extent irrigate the crop (our failure to include the treadle pump as an input is a source of statistical noise). This household therefore produced substantially more maize than other households in this drought affected area. The household was therefore a net seller of maize also in this drought year. HH 229 was affected by drought but has still been able to get a reasonable production output of 1350 kg maize from 0.48 ha, using 100 kg fertilizer (obtained through the subsidy program), and was one of the few net sellers of maize in its area in the drought year.

At the bottom of the table, we find HH 74 which has 0.31 ha of land with maize (local maize) but produced only 20 kg of maize on this area in 2012 due to the severe drought (it produced 250 kg maize in 2011 and 350 kg maize in 2010 showing the severity of the drought). It used 12.5 kg fertilizer on the maize but did not get any inputs from the subsidy program. HH 74 produced some cassava and pigeon pea (more drought tolerant crops) and had casual off-farm employment to make ends meet. The next unproductive household is HH 175 which produced only 200 kg of maize on 2.1 ha while using 100 kg fertilizer of which 50 kg was received from the subsidy program. This household also had casual off-farm work to make ends meet and was a net buyer of maize. The drought may have been an important reason for the low maize yields but poor management may also play an important role as yields in previous years with more favourable weather conditions were also low (OTE scores are relatively low). HH 163 produced 200 kg maize on 0.65 ha while using as much as 400 kg fertilizer of which 50 kg was received through the subsidy program. This farm household was therefore badly affected by the drought. It still managed to produce quite a bit of tomatoes for the market as the main source of cash income. HH 58 produced 350 kg of maize on 1.56 ha while using 80 kg fertilizer of which 50 kg was from the subsidy program. The drought was so severe that the maize had to be replanted. Some of the replanted area was planted with drought tolerant early maturing maize and some with another type of hybrid maize. This farm household also had additional income from casual off-farm employment as well as a non-agricultural business (having a sewing machine).

To summarize, these brief farm household stories reveal that all maize farms are small and they have to varying degrees been affected by the drought. Access to irrigation water, ability to replant the crop after the drought with an early maturing variety, late arrival of subsidized inputs and managerial ability appear as important reasons for variation in the TFPI and EOSMEI.

A clearer picture of the variation and drivers of TFP change is given in Figures 1 to 4. These figures summarise results for all 287 households, again ordered from most productive (HH 389) to least productive (HH 74). Figure 1 presents indexes of output, input and TFP change; these are the types of results we would expect to see reported in the growth accounting literature. Figure 1 tells us that input levels are fairly similar across firms, and that large difference in TFP are due to large differences in output.[5] Figure 2 presents TFP and EOSME indexes together with an output-oriented technical efficiency and statistical noise index (OTESNI $=$ OTEI $\times$ SNI); these are the types of results we would expect to see from SFA researchers who estimate the model using least squares (LS) methods (LS methods cannot disentangle technical inefficiency and noise). Figure 2 is also where the contribution of our paper lies: it indicates that differences in TFP are mainly due to differences in the environment and scale-mix efficiency (the EOSMEI component). Figure 3 is the type of analysis we might get from SFA researchers who estimate the model using ML (assumptions A1 and A2 allow us to disentangle the OTEI and SNI components). Unfortunately, Figure 3 suggests that the (standard) ML assumptions we have used in this paper may not be correct. Assumption A2, for example, says that all noise errors have a mean of zero. However, Figure 3 suggests that the noise errors for more productive households have lower means than the noise errors for unproductive households. Figure 4 presents estimates of OTE levels (not indexes). This figure reveals that levels of OTE ranged from less than 0.1 to approximately 0.8. This indicates that even the most technically efficient household in the sample could increase its output by as much as 25% by making better use of available technologies.

Finally, in an attempt to isolate the state of Nature effect, we predict output levels

[5] Growth accountants, who have interesting concepts of causality, might say that large differences in output were due to large differences in TFP. If prices were constant, then this would equivalent to saying that large differences in sales were due to large differences in profits. In our case, there are large differences in factor prices depending on the access to subsidized inputs (i.e. inputs with which have highly subsidized prices.

in two different states of Nature holding the inputs fixed for all observations in the dataset. To be more specific, our measure of the environmental effect is ENV $= \text{Q1}/\text{Q0}$ where Q1 is predicted output in a drought season and Q0 is predicted output in a normal season. In the case of HH 389, for example, ENV $= 0.955$, indicating that, all other things being equal, output in a drought is only 95.5% of what it is in a normal season (assuming that households do not change their input mix when they are facing drought). The average value of ENV across all households is 0.63, indicating that, all other things being equal, output in a drought is on average 37.1% lower than in a normal season. Figure 5 summarises these results. Most of the bars in Figure 5 are below the horizontal line at 1, indicating that drought generally leads to lower output (all other things being equal). However, for about 15 households, their inputs are such that drought would have led to higher output (in one case output would have more than doubled; the maximum value of ENV is 2.13). This is partly because of the estimated output elasticities: the estimated seed (resp. labor) elasticity is zero in a normal season (resp. drought).

9. Conclusion

We have made cross-sectional comparisons of total factor productivity in a risky agricultural setting with imperfect input markets. A new productivity index is constructed that satisfies important axioms from index number theory. This index can be computed without any information about input or output prices and is based on the estimated coefficients from state-contingent production frontiers. A maximum likelihood approach was used to estimate frontiers for two states of Nature, drought and no drought.

We find that most of the cross-sectional variation in productivity can be explained by variation in environment and output-oriented scale-mix efficiency (EOSME). Further dissagretion of this measure by state of Nature indicated that average output was 37.1% lower in the drought state of Nature than the normal state of Nature, all other things being equal (i.e., holding inputs and input attributes fixed).

We found significant technical inefficiency in the data (efficiency scores ranged from 0.1 to 0.8). Variation in access to input subsidies and late arrival of subsidized inputs in addition to variation in managerial skills may explain this. However, there are indications that the maximum-likelihood estimation methodology has failed to properly separate output-oriented technical efficiency from statistical noise.

The data also show that the average maize farm was small (0.72 ha) and that land scarcity was a limiting factor causing many maize farmers to be deficit producers. Maize is their main staple food, and a large share of the small farmers therefore depend on access to casual off-farm work (*ganyu*) to make ends meet. The results suggest that households are vulnerable to drought, and that subsidized inputs can, to a limited extent, protect them against drought shocks. This suggests that complementary policies and livelihood strategies are needed in years with more extreme weather conditions.

Table 1: Descriptive Statistics

	Variable	Mean	St. Dev.	Min.	Max.
q	output	754.07	940.42	20.00	7920.00
x_1	land	0.72	0.66	0.01	8.45
x_2	labor	55.05	41.13	4.00	276.00
x_3	seed	21.28	28.77	0.20	360.00
x_4	fertilizer	76.66	79.01	1.00	501.00
a_1	slope	0.29	0.61	0.00	3.00
a_2	soil fertility	0.38	0.78	0.00	3.00

Table 2: ML Parameter Estimates

		Estimate	St. Err.	p-value
Normal Season ($s = 1$)				
α_1	constant	6.8839^{***}	0.7362	< 0.001
β_{11}	log-land	0.7714^{***}	0.1207	< 0.001
β_{21}	log-labor	0.0207	0.1470	0.8880
β_{31}	log-seed	0.0000	0.0000	n.a.
β_{41}	log-fertilizer	0.2449^{***}	0.0672	0.0003
ψ_{11}	slope	-0.1064	0.1753	0.5440
ψ_{21}	soil fertility	-0.0013	0.1708	0.9940
Drought Season ($s = 2$)				
α_1	constant	-0.1850	0.7889	0.8146
β_{11}	log-land	0.4349^{***}	0.0695	< 0.001
β_{21}	log-labor	0.0000	0.0000	n.a.
β_{31}	log-seed	0.0204	0.0504	0.6854
β_{41}	log-fertilizer	0.1957^{***}	0.0282	< 0.001
ψ_{11}	slope	0.1824^{**}	0.0797	0.0220
ψ_{21}	soil fertility	-0.3990^{***}	0.0971	< 0.001
$\lambda \equiv \sigma_u/\sigma_v$		1.9153^{***}	0.2506	< 0.001
$\sigma \equiv \sqrt{\sigma_v^2 + \sigma_u^2}$		1.0133^{***}	0.0028	< 0.001

Table 3: Selected Indexes of Household Performance

Household	QI	XI	TFPI	OTEI	EOSMEI	SNI
389	0.938	0.176	5.319	1.950	11.544	0.236
289	2.250	0.849	2.649	1.629	2.737	0.594
77	1.875	0.794	2.360	1.849	3.522	0.362
229	1.688	0.841	2.008	1.742	2.384	0.483
2	1.719	0.884	1.944	1.751	2.346	0.473
:	:	:	:	:	:	:
300	1.000	1.000	1.000	1.000	1.000	1.000
:	:	:	:	:	:	:
271	0.375	1.216	0.308	0.449	0.533	1.288
58	0.438	1.638	0.267	0.639	0.358	1.168
163	0.250	1.339	0.187	0.550	0.279	1.219
175	0.250	2.057	0.122	0.447	0.211	1.290
74	0.025	0.404	0.062	0.196	0.196	1.616

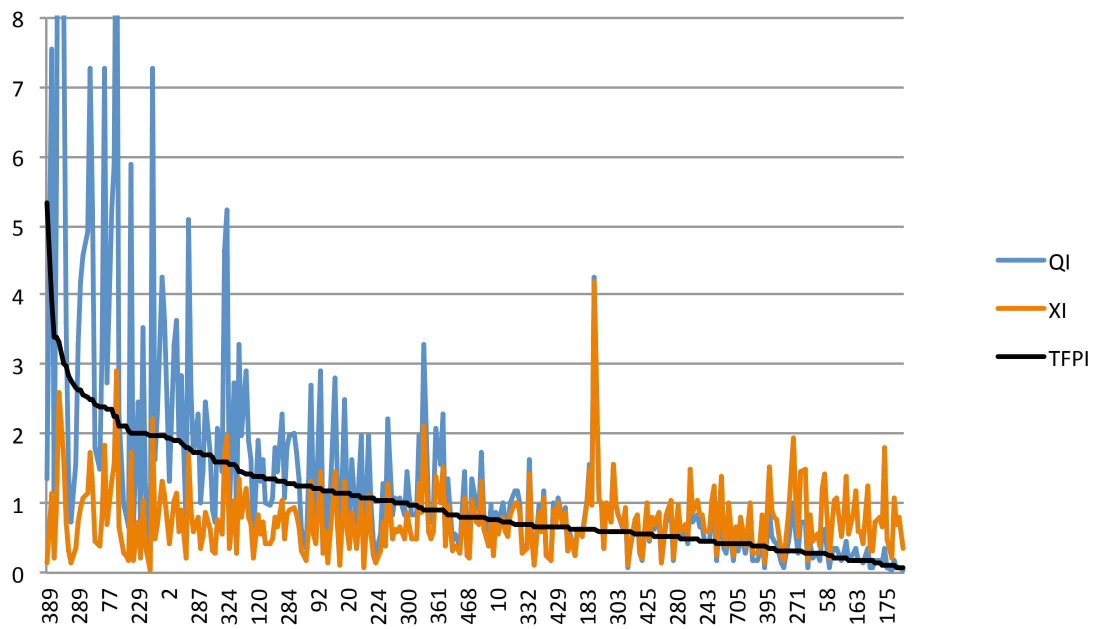

Figure 1: Output, Input and TFP Change

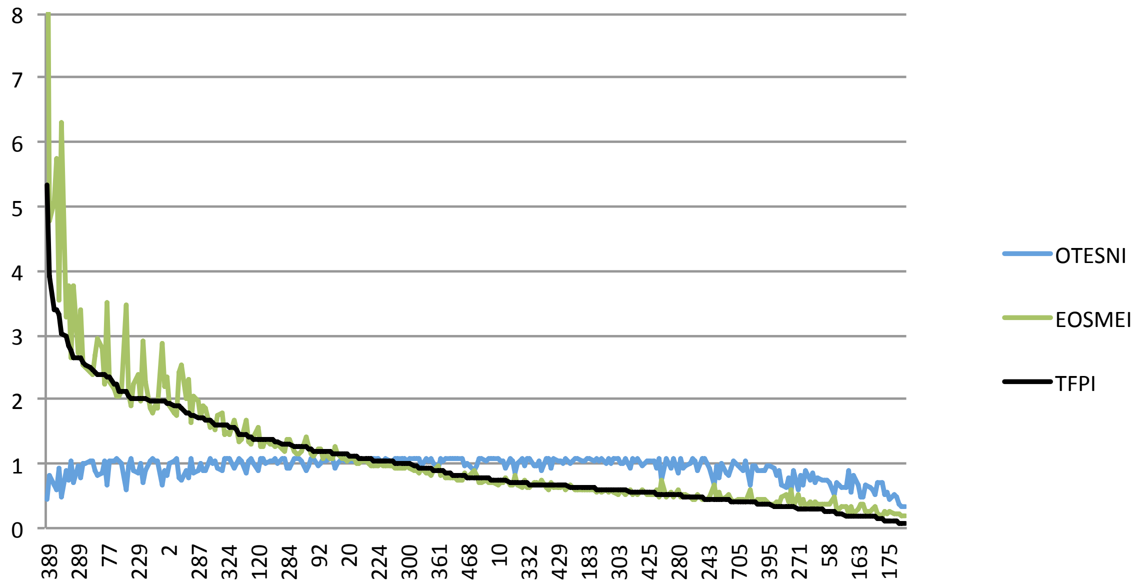

Figure 2: The Economic Components of TFP Change

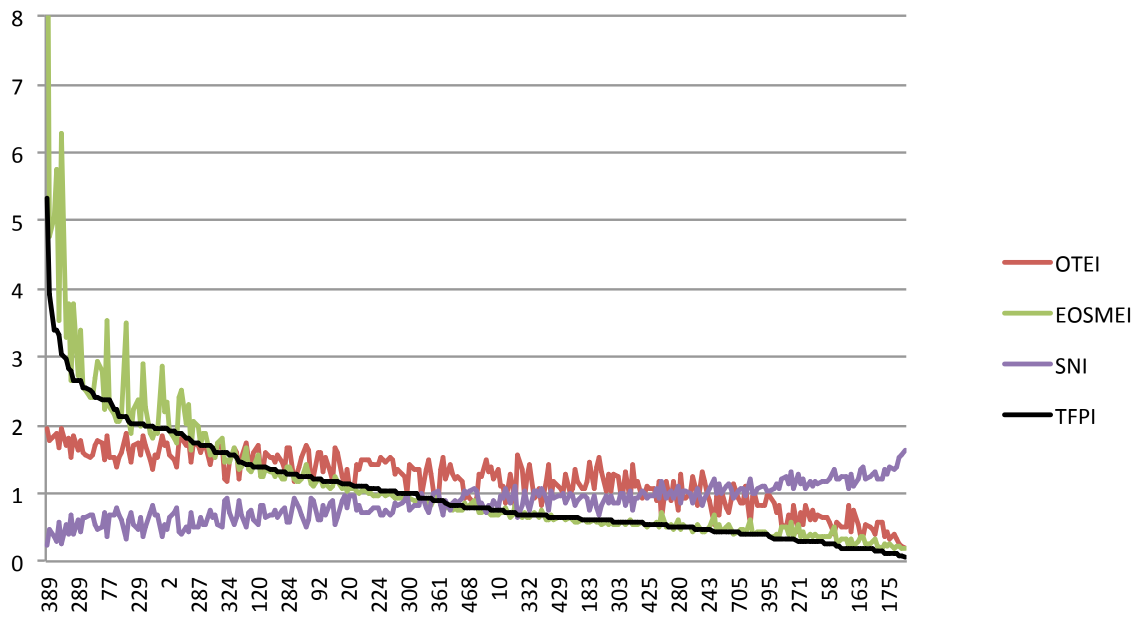

Figure 3: The Economic Components of TFP Change

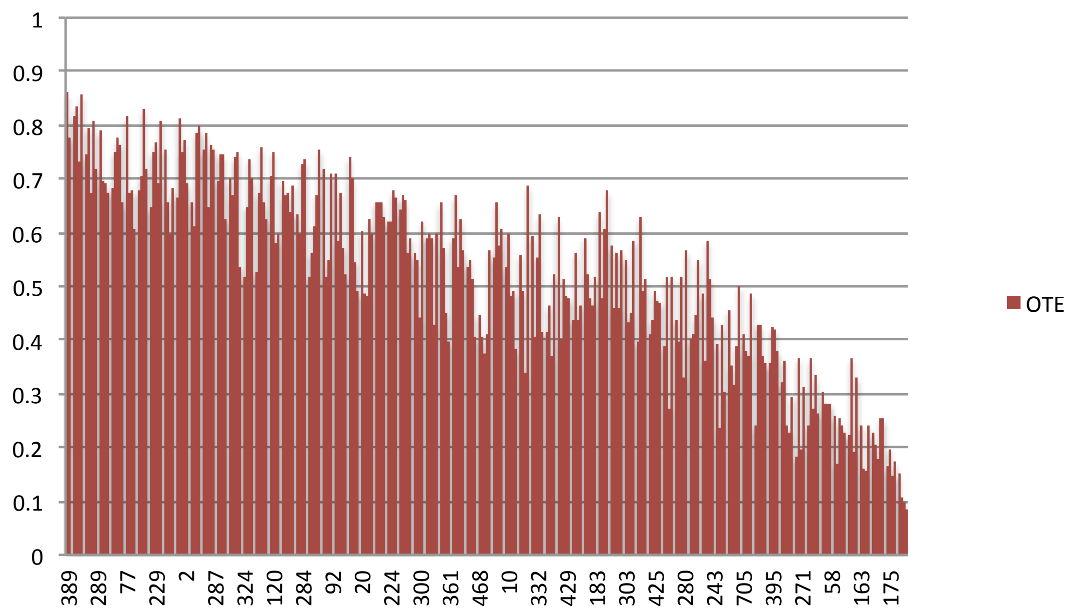

Figure 4: Levels of Output-Oriented Technical Efficiency

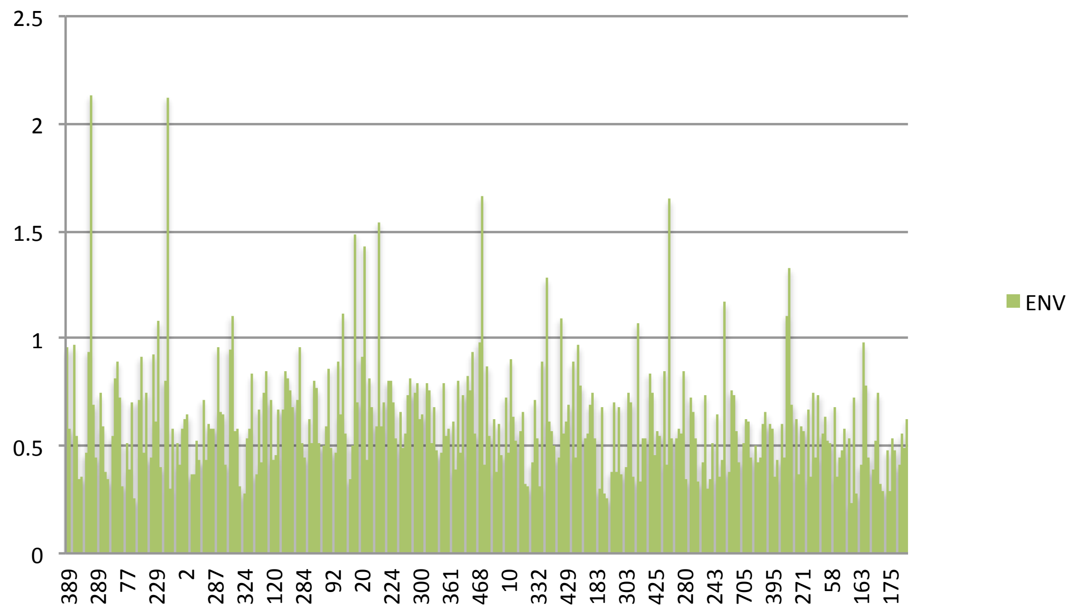

Figure 5: Environmental Effects

References

Binswanger, H. and Rosenzweig, M. (1986), 'Behavioural and material determinants of production relations in agriculture', *Journal of Development Studies* **22**(3), 503–539.

Cobb, C. and Douglas, P. (1928), 'A theory of production', *The American Economic Review* **18**(1), 139–165.

Debreu, G. (1951), 'The coefficient of resource utilization', *Econometrica* **19**(3), 273–292.

Färe, R. and Primont, D. (1995), *Multi-output Production and Duality: Theory and Applications*, Kluwer Academic Publishers, Boston.

Farrell, M. (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society, Series A (General)* **120**(3), 253–290.

Feder, G. (1985), 'The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints.', *Journal of Development Economics* **18**(2), 297–313.

Jondrow, J., Lovell, C., Materov, I. and Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics* **19**(2-3), 233–238.

Lunduka, R. (2009), Land rental markets, investment and productivity under customary land tenure systems in Malawi, PhD thesis, Department of Economics and Resource Management, Norwegian University of Life Sciences.

O'Donnell, C. (2012), 'An aggregate quantity framework for measuring and decomposing productivity change', *Journal of Productivity Analysis* **38**(3), 255–272.

O'Donnell, C. (2015), 'Using information about technologies, markets and firm behaviour to estimate and decompose a proper productivity index', *Journal of Econometrics* **forthcoming**.

Ricker-GIlbert, J., Jayne, T. and Chirwa, E. (2011), 'Subsidies and crowding out: A double-hurdle model of fertilizer demand in Malawi', *American Journal of Agricultural Economics* **93**(1), 26–42.

Shephard, R. (1970), *The Theory of Cost and Production Functions*, Princeton University Press, Princeton.