WP 98-15
December 1998

Smallholder Technical Efficiency With Stochastic Exogenous Production Conditions

Shane M. Sherlund, Christopher B. Barrett, and Akinwumi A. Adesina

Smallholder Technical Efficiency With Stochastic Exogenous Production Conditions

Shane M. Sherlund, Christopher B. Barrett, and Akinwumi A. Adesina

December 1998 first draft
Comments appreciated

Corresponding author: Dr. Christopher B. Barrett, 351 Warren Hall, Dept. Of Agr., Res. and Mgrl. Economics, Cornell University, Ithaca, NY 14853-7801; email: cbb2@cornell.edu; telephone: 607-255-4489; fax: 607-255-9984; http://www.cals.cornell.edu/dept/arme/staff/cbb2/

The authors are graduate student, University of Wisconsin-Madison, associate professor, Department of Agricultural, Resource, and Managerial Economics, Cornell University, and senior scientist, Rockefeller Foundation, respectively. We thank the West Africa Rice Development Association for making the data available, and David Aadland, Tim Dalton, Quinn Weninger, and seminar audiences at Utah State and the 1998 annual meeting of the American Agricultural Economics Association for helpful discussions. The usual disclaimer applies.

Smallholder Technical Efficiency With Stochastic Exogenous Production Conditions

Abstract

There is a large literature on the estimation of frontier production functions, much of it applied to low-income agriculture. However, much of this literature largely ignores nature's role in agricultural production. Because exogenous, natural production conditions (e.g., rainfall, soil quality, pest infestation, plant disease, weed growth) are rarely uniform or symmetrically distributed within a population or a sample thereof, this omission generally leads to downward bias in producers' estimated efficiency and to biased estimates of both the parameters of the production frontier and the correlates of true technical inefficiency. Using panel data from 464 traditional rice plots in Côte d'Ivoire, we show that controlling for stochastic, exogenous, natural production conditions in estimating the production frontier significantly increases smallholder rice farmers' estimated efficiency, whether estimated using parametric, stochastic or nonparametric, nonstochastic methods. The resulting frontier parameter estimates are also more consistent with theoretical predictions than are those of a frontier estimated without controlling for exogenous production conditions. Conventional estimates of technical efficiency may then mislead policymakers' perceptions of overall efficiency levels and of the sources of such inefficiency.

Key words: Africa (Sub-Saharan), Ivory Coast, production frontiers, agricultural productivity, rice.

JEL Codes: O12, Q12, D2

1. Introduction

In his classic "poor but efficient" hypothesis, Schultz (1964) argued that traditional farmers, given a long period of time to learn their production processes, will identify their respective optimal input and output bundles. Schultz then suggested that agricultural development policy adopt an expansionary approach with respect to smallholder production frontiers as the most cost-effective means to increase the welfare of low-income farmers around the world. This vision helped guide the Green Revolution and much ongoing research on improving crop production technologies in the developing world. But countless empirical studies have refuted Schultz's claim, finding widespread technical inefficiency among smallholder producers and consequently recommending that policy makers reallocate scarce resources toward redressing apparent obstacles to farmer technical efficiency through improved extension work, farmer education, land tenure reforms, etc. Today, several international agricultural research centers and national agricultural research systems face funding crises, and there appears less consensus than there was a quarter century ago within the development community about the relative importance of advances in crop production technologies to the improvement of smallholder welfare.

This paper speaks to this issue by reconsidering the estimation of production frontiers in low-income agriculture heavily dependent on stochastic, exogenous production conditions. In particular, we show that the omission of measurable exogenous arguments to a production function (e.g., pest and weed infestation, plant disease, rainfall) leads to inflated estimates of plot-specific technical inefficiency, and to biased estimates of both the parameters of the production frontier and of the correlates of technical inefficiency, which might be used for targeting extension interventions. We demonstrate these results empirically using a panel data set of rice farmers in the west African nation of Côte d'Ivoire to estimate production frontiers using both stochastic, parametric and nonstochastic, nonparametric estimation methods.

2. Natural Variability, Production Frontiers, and Technical Efficiency

There are industries (e.g., banking, semiconductors) in which firms have considerable, or even complete control over the physical production environment. Such is not the case in traditional smallholder agriculture, which responds strongly to natural conditions that vary markedly over time and space. The stochastic, natural environment conditions the results of farmers' production decisions. Otherwise identical producers — same technology, same ability — will produce different quantities of grain if faced with different rainfall, plant disease, weed, or other natural production conditions.[1]

This fundamental feature of smallholder agriculture should inform the estimation of production frontiers, but rarely does. We attribute this oversight primarily to the absence from most farm production data sets of detailed, farm- or plot-specific information on the natural conditions facing producers. Lack of data causes analysts to omit potentially relevant exogenous, natural variables. But with what consequence? If the econometrician does not control explicitly for environmental variability, he or she is ultimately comparing *all* producers against the (probably small) subset of producers facing the best realized exogenous production environment. Since natural variables are likely important to smallholder productivity, it is reasonable to suspect the result will be omitted variables bias in the estimated parameters of the production frontier.[2] We will show that the omission of relevant stochastic exogenous production inputs also yields downward bias in the estimated technical efficiency of production units and bias in the estimated relationship between technical efficiency and various
indicators that might be used to target remedial interventions.

In slightly more formal terms, suppose farmer i generates output, Y_i, from a production frontier

[1] It is also likely that exogenous conditions influence input allocation of land, labor, fertilizer, etc. In this paper, for the sake of degrees of freedom in estimation, we maintain the hypothesis of separability between traditional and natural inputs. An obvious extension of this work is to relax that assumption.

[2] Furthermore, since these omitted variables are likely correlated with variables included as regressors, there will likely also be an endogeneity problem. We do not address this problem directly in the paper.

defined over endogenous inputs, $\mathbf{X}_i$, and exogenous states of nature, $\mathbf{W}_i$, adjusted for the farmer's technical inefficiency, u_i ($u_i \leq 0$) where $i = 1, \ldots, \hat{N}$. Output is strictly monotonically increasing in both $\mathbf{X}$ and $\mathbf{W}$[3]. This relationship may be estimated as either a nonstochastic frontier, Y # = # f ($\mathbf{X}$;$\mathbf{W}$) + $\mathbf{u}$, or, given mean zero, symmetric sampling and measurement error, $\mathbf{v}$, in the data, as a stochastic frontier, $\mathbf{Y} = f(\mathbf{X};\mathbf{W}) + \mathbf{u} + \mathbf{v}$. (The production unit is said to be technically efficient only if $u_i = 0$.) Because the literature has generally paid little attention to the exogenous shocks affecting output, the relation typically estimated is instead $\mathbf{Y} = g(\mathbf{X}) + \hat{\mathbf{u}} + \hat{\mathbf{v}}$. The first problem to recognize is that if $\mathbf{Y}$ indeed is at least partly a function of $\mathbf{W}$, then the $g(\cdot)$ estimates suffer obviously from omitted variable bias. Note that $f(\mathbf{X};\mathbf{W}) = g(\mathbf{X})$ if and only if $\mathbf{Y}$ and $\mathbf{W}$ are uncorrelated.

The second problem arises from the stochasticity and potentially asymmetric distribution of $\mathbf{W}$. If there is variation in sample in $\mathbf{W}$, then a nonstochastic frontier fitted without controlling for $\mathbf{W}$ will necessarily generate $\hat{u}_i < u_i$ (recall $u_I \leq 0$) for any i^{th} production unit for which $W_i<W^*$, where W^* is the optimal realized value of any element of $\mathbf{W}$ in the sample. So when there exists variation in sample in natural conditions, there will be downward bias in the estimated technical efficiency because $E[\hat{u}_i] < u_i$ for those production units experiencing suboptimal realized environmental conditions.

This problem exists even when estimating a stochastic frontier, albeit under slightly less general conditions. Assume for the moment that $\mathbf{X}$ and $\mathbf{W}$ are uncorrelated, that $\mathbf{W}$ has distribution $\Phi(\mathbf{w})$ in sample, and that $\mathbf{v}$ has distribution $\Upsilon(\mathbf{v})$ (usually the normal). When $\mathbf{W}$ is omitted from the regressors in estimating $f(\cdot\cdot)$—hence in estimating $g(\cdot)$—its effects on $\mathbf{Y}$ will then be picked up in the composite error term, $\hat{\mathbf{u}} + \hat{\mathbf{v}}$. A necessary — but not sufficient — condition for the totality of the effect of W to be captured by the

[3] In this analytical section, for the sake of clarity we assume $\mathbf{W}$ represents states of nature ordered from worst to best, hence the monotonicity of $\mathbf{Y}$ in $\mathbf{W}$. In practice, such an ordering may, however, require a nonmonotonic transform of the raw, underlying data (e.g., for rainfall) since moderate measures may be optimal. In the empirical section to follow, we will work with polynomial functions of the raw data, implicitly relaxing the assumption of

statistical residual, $\hat{\mathbf{v}}$, and not to affect the technical efficiency estimates, $\hat{\mathbf{u}}$, is that the Φ and Υ distributions differ from one another only by location and scale parameters. In general, $\Phi(\mathbf{w})$ can be represented as a mixture of two distributions, Γ and Υ, the latter potentially transformed by a location parameter, α, a scale parameter, β, or both, such that $\Phi(\mathbf{w}) = \gamma\Gamma(\mathbf{w}) + (1 - \gamma)\, \Upsilon(\alpha + \beta\mathbf{w})$, for $\gamma \in [0,1]$. If $\gamma > 0$, then the efficiency parameter estimates, $\hat{\mathbf{u}}$, will capture part of the effect of the omitted exogenous variables, **W** because of the deviation from location and scale differences introduced through Γ. If $\Phi(\mathbf{w})$ is asymmetric and $\Upsilon(v)$ is the conventional normal (or another symmetric distribution, perhaps student-t), then it must be the case that $\gamma>0$, since there is more than a location-scale difference between the two distributions. Under the standard assumption of symmetrically distributed measurement and sampling error, v, an asymmetric distribution for $\Phi(\mathbf{w})$ implies $\hat{\mathbf{u}}$ is a biased and inconsistent estimator of **u** because $E[\hat{u}_i] \neq u_i$ for some production units (unless $f(\mathbf{X}_i, \mathbf{W}_i) = g(\mathbf{X}_i)$, implying that **Y** and **W** are uncorrelated). In particular, the bias will be downward, $E[\hat{u}_i] < u_i$, i.e., technical efficiency will be understated.

The third problem arises with respect to identifying the correlates of true technical inefficiency, u. While knowing the extent of technical inefficiency prevailing in a sector is useful, policy makers often also like to know the correlates of technical inefficiency so as to target interventions appropriately and thereby reduce inefficiency in the sector. The second-stage relation to be estimated is thus

$\mathbf{u} = h(\mathbf{Z}) + \boldsymbol{\tau}$, where **Z** is a vector of farmer characteristics and practices, and $\boldsymbol{\tau}$ is a white noise error term. However, the second-stage relationship is commonly estimated as $\hat{\mathbf{u}} = j(\mathbf{Z}) + \hat{\boldsymbol{\theta}}$. If the efficiency parameter estimates generated by a model omitting **W**, $\hat{\mathbf{u}}$, are biased estimates of the true **u** in population, then $j(\cdot)$ will likewise yield biased and inconsistent estimates of the relationship of interest, $h(\cdot)$.

Since, in general, $f(\mathbf{X};\mathbf{W}) \neq g(\mathbf{X})$, $\mathbf{u} \neq \hat{\mathbf{u}}$, and $h(\mathbf{Z}) \neq j(\mathbf{Z})$, misleading policy implications may be

---

monotonicity with respect to **W**.

drawn from estimated frontier production functions that omit stochastic exogenous conditions. One clear implication is that this omission is likely to lead to an overstatement of the prevalence and degree of technical inefficiency among smallholders highly dependent on stochastic exogenous natural conditions to agricultural production. That overstatement may lead to underemphasis on the need for research designed to shift outward the production possibility frontiers currently facing traditional producers.

In the remainder of this paper, we show how the omission of measurable stochastic exogenous conditions, **W**, alters the frontier production function estimation results in one particular case: rice in the west African nation of Côte d'Ivoire. We estimate the primal production function rather than dual cost or profit functions primarily because of the myriad inferential problems associated with using observed market prices in estimating the production behaviors of smallholders most of whose labor, land, and animal traction allocation decisions do not involve market transactions (Barrett 1997).

3. Data

The data used come from the farm management and household survey (FMHS) fielded by the West Africa Rice Development Association (WARDA). WARDA FMHS tracked 120 randomly selected rice-producing households in Côte d'Ivoire, 1993-95, encompassing 1,218 individual plots, 589 of which were planted in rice. Twenty-two surveys were administered annually and are described in detail in WARDA (1997). Due to nonsystematically missing data or mechanization (we study only traditional rice farmers, the subject of Schultz's "poor but efficient" hypothesis), 464 of the 589 rice plots are used in the estimation reported in the next section.

One comparative advantage of the WARDA FMHS is its inclusion of quarterly plot-level measurements of exogenous stochastic conditions, such as rainfall, pests, weeds, plant disease, plot slope, and soil quality. Sample descriptive statistics are presented in Table 1 for the key explanatory variables:

land, adult family labor, adult hired labor, child labor, animal traction, fertilizer, erosivity, soil fertility, soil aptitude for rice, pest infestation, weed density, weed height, plant disease, topographic location, plot slope, number of days of rain, and rain volume. As is evident, there is relatively little use of either animal traction or chemical fertilizers—this is prototypical smallholder, traditional cropping—and considerable variability in area and labor use patterns, as well as in natural environmental conditions.

The previous section pointed out that if the stochastic natural environmental conditions of production are not symmetrically distributed, then their omission will lead to biased estimates of plot-specific technical efficiency. As reflected in Table 1 and shown graphically for a few variables in Figure 1, these exogenous variables are asymmetrically distributed, with statistically significant positive skewness. So the problems identified in the previous section appear relevant in this data set, which affords an uncommon opportunity to check the consequences of the omission of measurable exogenous states of nature.

4. Stochastic, Parametric Production Frontier and Technical Efficiency Estimation

Much of the frontier production function estimation literature follows the stochastic, parametric approach pioneered by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977). In this approach, one specifies *a priori* a functional form for the production frontier and probability density functions for the asymmetric technical inefficiency parameter and the symmetric error term. We employ a modified form of the generalized Leontief production frontier and a half-normal probability density function for the technical inefficiency parameter, $\mathbf{u}$.[4] In order to demonstrate the consequence of omitting exogenous production conditions, we estimate the frontier with and without those $\mathbf{W}$ variables. The traditional, or "short" specification, omitting $\mathbf{W}$ (i.e., $\mathbf{Y} = g(\mathbf{X}) + \hat{\mathbf{u}} + \hat{\mathbf{v}}$), may be written:

---

[4] The presence of many zero-valued observations precludes use of the transcendental logarithmic form, and the standard generalized Leontief, using square rather than cube roots, failed diagnostic tests. Hence the choice of this particular functional form. We also estimated this using exponential and truncated normal distributions for the technical efficiency parameter and obtained qualitatively identical results. For more detail on the specification testing, see Sherlund (1998).

$$\sqrt[3]{\mathbf{Y}} = \beta_0 + \sum_{k=1}^{K} \beta_k \left(\sqrt[3]{\mathbf{X}_k} \right) + \frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{K} \gamma_{kj} \left(\sqrt[3]{\mathbf{X}_k \mathbf{X}_j} \right) + \hat{\mathbf{u}} + \hat{\mathbf{v}} \;, \tag{1}$$

where β_0, β_k, γ_{kj} (j,k = 1,...,K) are parameters to be estimated. Output (**Y**) is rice production, in kilograms. The inputs (**X**) are land, measured in acres; familial labor, measured in hours; hired labor, measured in hours; child labor, measured in hours; animal traction, measured in hours; and chemical fertilizer usage, measured in kilograms. The error term, $\hat{\mathbf{v}}$, is assumed to be iid, normal.

We assume separability of **X** and **W** in order to conserve degrees of freedom in estimating the "full" specification, also as a modified generalized Leontief production function. This may be written:

$$\sqrt[3]{\mathbf{Y}} = \beta_0 + \sum_{k=1}^{K} \beta_k \left(\sqrt[3]{\mathbf{X}_k} \right) + \frac{1}{2} \sum_{k=1}^{K} \sum_{j=1}^{K} \gamma_{kj} \left(\sqrt[3]{\mathbf{X}_k \mathbf{X}_j} \right) + \sum_{r=1}^{R} \delta_r \mathbf{W}_r + \mathbf{u} + \mathbf{v} \;, \tag{2}$$

where δ_r (r = 1,...,R) are parameters to be estimated in addition to those listed above. The exogenous **W** variables included are categorical variables for erosion, where one indicates erosion is a problem, zero, that it is not; fertility, where one is good fertility and three is poor; soil aptitude, rated one to three, as with fertility; slope, measured in percent incline; pest infestation, ranked from one, no pest problems, to seven, severe pest problems; weed density, where one is no weed density problems and five is severe; weed height, where one is no weed height problems and five is severe; plant disease, where one is no plant disease and nine is severe;[5] topographic location dummies, where the base is upland; annual days of rain; and rainfall volume, in centimeters per year.

Parameter estimates for both the short and full specifications are reported in Table 2 (t-ratios presented in parentheses). The statistical superiority of the full specification is apparent statistically in the likelihood ratio test statistic, 233.9, which has a p-value equal to 0.000 against the $\chi^2(20)$ distribution.

---

[5] The insect infestation, plant disease, and weed density variables are based on decile categorical observations, i.e., a "2" score means 11-20% of the plot was affected, a "7" means that 61-70% was affected, etc.

Following the work of Jondrow, et al. (1982), we calculated plot-specific output measures of technical efficiency, $\theta_i \in [1,\infty)$, where θ_i represents the multiple by which the i^{th} plot's output could be expanded using the same inputs on the production frontier. The empirical distribution of these θ_i are plotted in Figure 2a for both the short and full specification, and descriptive statistics are reported in the two leftmost columns of Table 3. The unambiguous reduction in estimated technical inefficiency appears as first-degree stochastic dominance of the full over the short specification's distribution and as a nontrivial reduction in the median and the moments of the θ_i distributions. This demonstrates our claim that the omission of relevant measurable exogenous production conditions biases upward estimates of technical inefficiency when the **W** are assymmetrically distributed. The policy implications are significant, because the less the apparent technical inefficiency, the greater the need for research to generate outward shifts in the production possibilities frontier.

The omission of exogenous production conditions not only biases estimates of technical efficiency, it also affects the estimates of the parameters of the production frontier. This is somewhat apparent in Table 1's raw parameter estimates, but appears more readily in Table 4, where output elasticity estimates are reported for the six discretionary (land, familial labor, hired labor, child labor, animal traction, and fertilizer) at the sample means. The qualitative picture is similar across both specifications. Land is the key input to production, animal traction and chemical fertilizers contribute little, and output is significantly more responsive to family adult labor than to either hired adult or child labor. But in this sample, omission of relevant measurable exogenous shocks to production increases output elasticities nearly 30 and 80 percent with respect to familial and hired labor, respectively. And the elasticity estimates for both child labor and animal traction are negative in the short specification, but positive—albeit of low magnitude—in the full specification.

In these data, the improved parameterization of the production frontier has the added advantage of eliminating some inconsistencies between the parameter estimates and basic hypotheses of economic theory. When we check for monotonicity across the data, we find it is commonly violated in the short specification with respect to child labor in more than 80 percent of the sample, and with respect to animal traction in one-third (Table 5). Under the full specification, monotonicity holds almost universally, failing to hold only with respect to animal traction in 2.5 percent of the sample. In both specifications, however, the principal leading minor determinants of the Hessian of g(·) routinely fail to satisfy the conditions of negative semi-definiteness, implying violation of the quasi concavity assumption.[6] While concavity is satisfied with trivially greater frequency in the full specification, this inconsistency with the postulates of producer theory motivates us to repeat the production frontier estimation exercise using an alternative method.

5. Nonstochastic, Nonparametric Production Frontier and Technical Efficiency Estimation

The previous section showed that the omission of relevant measurable exogenous production conditions biases upward estimates of technical inefficiency and adversely affects the estimated shape of a production frontier estimated using stochastic, parametric methods. A second major branch of the production frontier estimation literature employs the nonparametric, data envelopment analysis (DEA) technique. DEA requires no *a priori* assumptions regarding either the functional form of the production frontier or the probability density functions of the technical inefficiency term, merely imposing basic weak monotonicity and concavity properties, thereby obviating problems of inconsistency between estimation results and theory, as the last section showed can occur with parametric, stochastic frontier estimation (Färe, Grosskopf, and Lovell, 1994). The main drawback of DEA is its assumed lack of measurement or sampling error, i.e., it is a nonstochastic method.

The output-oriented, variable returns to scale, strong disposability DEA model may be written:

[6] Many observations come close to satisfying the conditions for quasi-concavity, e.g. having one determinant of the wrong sign by $< 1 \times E - 8$. Indeed, at a tolerance level of 0.0001 (0.01) quasi-concavity is universally satisfied in

$$\theta^*(X_i, Y_i \mid VRS, SD) = \text{Max}_{\theta,z}\ \theta, \tag{3a}$$

$$\text{Subject to:} \quad \theta \cdot Y_i \leq z \cdot Y, \tag{3b}$$

$$z \cdot X \leq X_i, \tag{3c}$$

$$\Sigma_i\, z_i = 1, \tag{3d}$$

$$z \in R_{+}^{N}, \text{ where } i = 1,\ldots,N, \tag{3e}$$

and z is the activity vector indicating to which plots the i^{th} plot is being compared. The resulting output measure of technical efficiency is bounded from below at one, $\theta_i^* \geq 1$, and represents the multiple by which output may be expanded, holding the input bundle constant, had the i^{th} plot been fully efficient. Excluding constraint (3d) yields the analogous output-oriented, constant returns to scale, strong disposability DEA model. However, applying Banker's (1996) test we reject the null hypothesis of constant returns to scale (at the 99.977 percent confidence level) in favor of the variable returns to scale specification (Sherlund 1998).

We estimate plot-specific technical efficiency scores using the linear program of equations (3a)–(3e) under both the short and full specifications introduced in the previous section, now relaxing the functional form assumption but imposing weak monotonicity and concavity. The effect of controlling for exogenous production conditions becomes more apparent in these data when using DEA than when using the parametric method of the previous section. As shown in Figure 2b and the two rightmost columns of Table 3, inclusion of exogenous production conditions substantially shifts the empirical distribution of estimated plot-specific technical efficiency scores, lowering the mean θ_i^* estimate from 2.3278 to 1.1798. Of perhaps greater practical importance, once measurable exogenous conditions are controlled for, the nonparametric DEA method finds 71 percent of plots exhibit perfect technical efficiency, $\theta_i^* = 1$. This stands in stark contrast to the implication of the short specification DEA results, which suggest the median plot could increase output

---

the full (short) specification.

92 percent, given current inputs, by improving productive efficiency. The downward bias in technical efficiency estimation when one fails to control for exogenous production conditions can thus have a dramatic effect, as clearly demonstrated by the DEA results in these data. The implications for agricultural research policy could be quite significant.

6. Correlates of Technical Inefficiency

Our concern is not just that the incidence and degree of smallholder technical inefficiency is prone to overstatement by the omission of exogenous production conditions from frontier production function estimation. We are also concerned about biased subsequent inference about the correlates of true technical inefficiency. In order to design effective agricultural development policy to ameliorate apparent smallholder inefficiency, policy makers must know the sources of peasant farmer inefficiency. This is commonly done by estimating a second-stage relationship between the technical inefficiency estimates, $\hat{\mathbf{u}}$, and suspected sources of technical inefficiency, **Z**. Statistically significant correlates of technical inefficiency are then used to target policy interventions intended to improve efficiency levels. Nevertheless, because omission of relevant measurable exogenous variables leads to an upward bias in technical inefficiency estimates it may also generate spurious estimated relationships between true technical inefficiency, **u**, and **Z**, and therefore potentially misleading policy recommendations based on the estimated but incorrectly specified second-stage relation.

We demonstrate this by regressing the natural logarithm of output-oriented technical efficiency scores on a vector, $\mathbf{Z}_i$, of plot-specific managerial characteristics (e.g., age, gender, education, and experience) and plot-specific managerial practices (e.g., type of seed used and the number of crops and plots cultivated). For sample descriptive statistics, refer to Table 1. The regression model may be written:

$$\ln(\theta_i^*) = \psi + \mathbf{Z}_i\alpha + \mathbf{W}_i\rho + \boldsymbol{\tau}_i, \tag{4}$$

where **W** is the earlier-used vector of measurable exogenous shocks, τ_i is a Gaussian white noise error term, and ψ, α, and ρ are estimable parameters. Note that because $\theta_i^* \geq 1$, $\ln(\theta_i^*) \geq 0$, so we estimate this as a Tobit model.[7]

We first run the short second-stage specification prevailing in the literature, using the θ_i^* generated by the short first stage regression (i.e., omitting exogenous production conditions) and restricting ρ to be equal to zero. Then we estimate the full specification, (4), using the full first stage regression (i.e., including exogenous production conditions) to ascertain the effects of both using the biased technical efficiency parameter as the regresand and omitting the W variables from the regressors of the second stage regression. The results are reported in Table 6.

The likelihood dominance criterion non-nested specification test (Pollak and Wales 1991) strongly favor the full specification models over the short specification models for both the parametric and nonparametric cases, confirming in these data our claim that the omission of exogenous production conditions is problematic for inference on the correlates of smallholder technical inefficiency. But the estimation results do not yield results that are robust to choice of modeling method. With the exception of like findings that the (few) holders of professional degrees are economically (and at the ten percent level, statistically) significantly more inefficient than other smallholder rice producers, the DEA and stochastic, parameter model results vary from one another. The instability of these parameter estimates across techniques and the relatively low magnitude of the estimated technical inefficiency of most production units, once one controls properly for exogenous production conditions, underlines the challenge of appropriate targeting of extension and farmer education efforts to improve sectoral productivity in traditional, smallholder agriculture.

---

[7] Because the stochastic production frontier approach does not yield technical efficiency scores of one (perfect efficiency), the Tobit model using those θ estimates collapses to the OLS model.

Moreover, a striking finding of the DEA results reported in the rightmost column of Table 6 is that no managerial characteristic or managerial practice variable is statistically significantly related to estimated technical inefficiency at the 95 percent confidence level or higher. A likelihood ratio test of the null hypothesis that managerial characteristics and practices, **Z**, jointly have no statistically significant relation to estimated technical efficiency once proper control is made in the second stage regression for exogenous production conditions yields a test statistic of 21.01, with a p-value of 0.178 against the $\chi^2(16)$ distribution. We cannot reject the null hypothesis that managerial characteristics and practices are unrelated to technical efficiency estimated under the full, unbiased specification. This result, like those of the preceding two sections, suggests that these Ivorien rice farmers are largely managerially efficient. The results reported in Tables 3 and 6 thus suggest a need to focus on research to expand smallholder rice production frontiers if productivity is to increase appreciably in this particular setting.

6. Conclusions

This paper is motivated by a concern that the empirical literature on technical efficiency estimation of smallholder agriculture largely ignores that production decisions are made in and data are drawn from highly stochastic production environments largely beyond the producer’s control. We first explain why prevailing empirical methods, using either econometric or programming techniques, may yield biased and inconsistent estimates of technical efficiency, production frontiers parameters, and the relationship between estimated technical efficiency and managerial characteristics and practices. This may have serious implications for policy makers relying on statistical inference from such models to guide resource allocation in agricultural development. We then demonstrate the relevance of our concern to the case of smallholder rice production in the west African nation of Côte d’Ivoire. Using plot-level panel data, we show that controlling for observable exogenous production conditions yields significantly lower estimates of technical inefficiency, more intuitive (e.g., positive) output elasticity estimates, and the finding that managerial characteristics and practices are effectively unrelated to estimated technical inefficiency. Using nonparametric methods, the median Ivorien rice plot appears perfectly technically efficient. This is quite a different story than that either prevailing in most of the existing literature or which one obtains by using conventional methods omitting exogenous production conditions from both the first stage frontier estimation and second stage estimation of the correlates of technical inefficiency.

These results have significant policy implications. Conventional methods of estimating production frontiers, technical inefficiency in production and the correlates of technical inefficiency suggest that the traditional Ivorien rice farmers we study are highly inefficient, leaving open the question of whether scarce agricultural development funds are best spent to develop improved technologies or to teach farmers how to use existing technologies better. By controlling for variation in observed exogenous production conditions, however, we find instead that there is relatively little technical inefficiency at the level of rice plots, and the inefficiency that does seem to exist is not strongly correlated with targetable farmer characteristics or practices. Schultz appears to be right when one compares Ivorien rice producers against the estimated stochastic production frontier they actually face, given their idiosyncratic realization of the environmental conditions vector, $\mathbf{W}_i$, rather than against the state-unconditional best-practice frontier, which implicitly pits them against colleagues enjoying considerably more favorable realized exogenous shocks to production.

References

Aigner, D.J., Lovell, C.A.K., and Schmidt, P. (1977). ‘Formulation and estimation of stochastic frontier production function models’, *Journal of Econometrics*, **6**, 21-37.

Ali, M. And Byerlee, D. (1991). ‘Economic efficiency of small farmers in a changing world: a survey of recent evidence’, *Journal of International Development*, **3**, 1-27.

Banker, R.D. (1996). ‘Hypothesis tests using data envelopment analysis’, *Journal of Productivity Analysis*, **7**, 139-159.

Barrett, C.B. (1997). ‘How Credible Are Estimates of Peasant Allocative, Scale or Scope Efficiency? A Commentary,’ *Journal of International Development*, **9**, 221-229.

Färe, R., Grosskopf, S., and Lovell, C.A.K. (1994). *Production Frontiers*. Cambridge, MA: Cambridge University Press.

Jondrow, J., Lovell, C.A.K., Materov, I.S., and Schmidt, P. (1982). ‘On the estimation of technical inefficiency in the stochastic frontier production function model’, *Journal of Econometrics*, **19**, 233-238.

Kumbhakar, S.C. (1987). ‘The specification of technical and allocative inefficiency in stochastic production and profit functions’, *Journal of Econometrics*, **34**, 335-348.

Meeusen, W. and van den Broeck, J. (1977). ‘Efficiency estimation from Cobb-Douglas production functions with composed error’, *International Economic Review*, **18**, 435-444.

Pollak, R.A. and T. J.Wales (1991). "The Likelihood Dominance Criterion.” *Journal of Econometrics* **47**, 227-242.

Schultz, T.W. (1964). *Transforming Traditional Agriculture*. New Haven, CT: Yale University Press.

Sherlund, S.M. (1998). ‘Exogenous shocks and technical efficiency among Ivorien rice farmers’, M.S. thesis, Logan, UT: Utah State University.

WARDA. (1997). ‘WARDA/ADRAO farm management and household survey data guide’, Bouaké, Côte d’Ivoire: West Africa Rice Development Association.

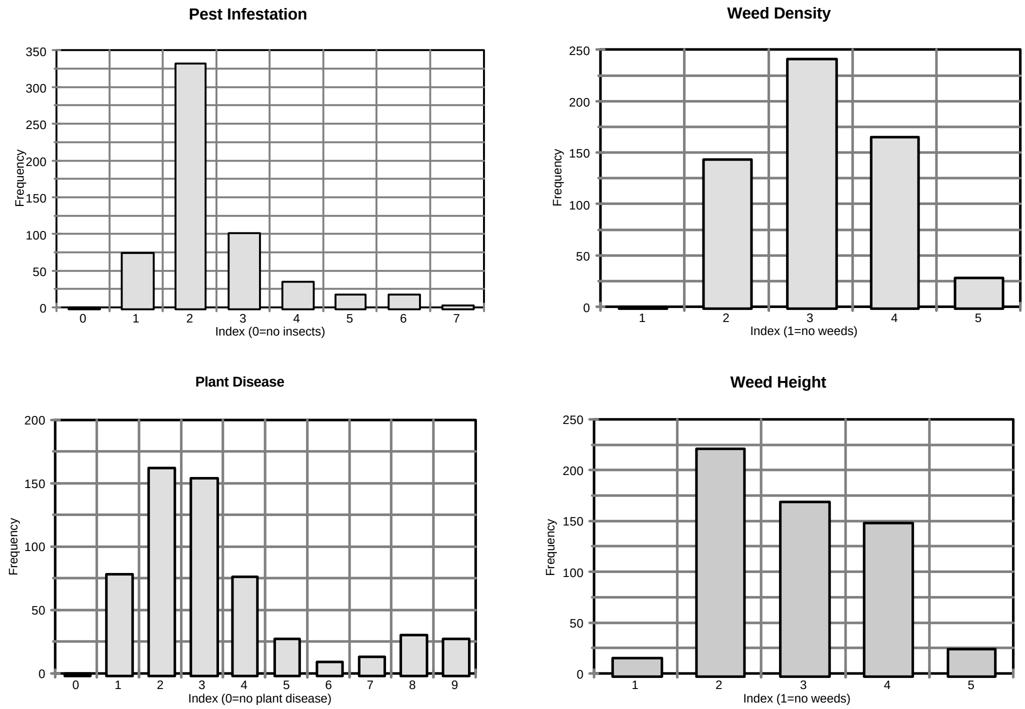
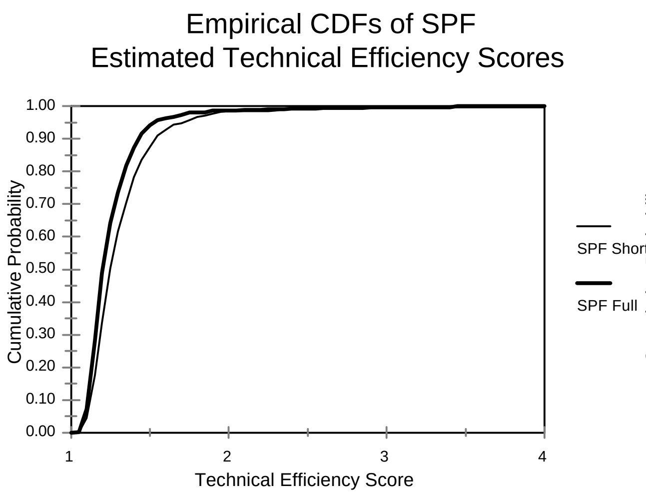


Figure 1. Distributions of asymmetric exogenous shocks



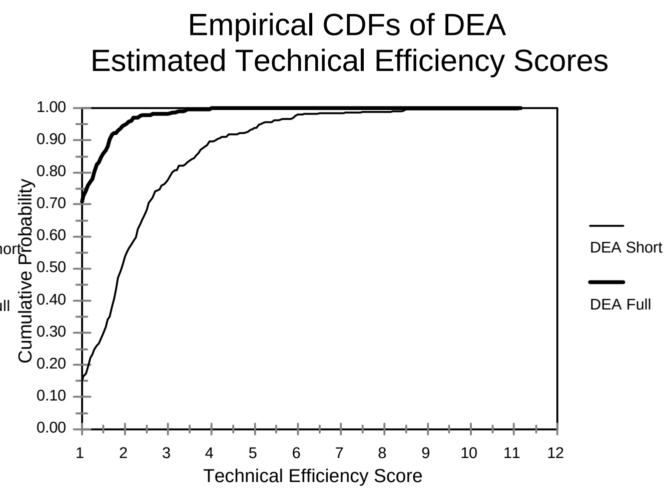


Figure 2a. Empirical distributions of plot-technical efficiency parameters, estimated by a stochastic parametric frontier

Figure 2b. Empirical distributions of plot-specific specific technical efficiency parameters, estimated by data envelopment analysis

Table 1
Descriptive Statistics

	Units of Measure **Mean**	**Standard Deviation** **Skewness**	**Minimum** **Maximum**
Production	kilograms 1676.4541	1399.1367 2.1546	46.6200 10094.0175
Land Area	ares (100 m^2) 94.1020	80.8435 2.7984	4.1300 710.0000
Familial Labor	hours 470.4610	400.3348 1.6097	0.0000 2545.5000
Hired Labor	hours 298.4638	262.1401 1.9130	0.0000 1984.0000
Child Labor	hours 408.2386	640.0552 2.3759	0.0000 3662.0000
Animal Traction	hours 10.9253	28.7920 3.9397	0.0000 213.0000
Fertilizer Use	kilograms 17.5175	51.9609 3.5552	0.0000 350.0000
Erosion problems	0=no, 1=yes 0.3879	0.4873 *** 0.4615	0.0000 1.0000
Fertility of plot soil	1=very, 2-okay, 3=poor 1.7694	0.6472 *** 0.2643	1.0000 3.0000
Aptitude of plot soil	1=very, 2-okay, 3=poor 1.4741	0.6118 *** 0.9203	1.0000 3.0000
Slope of plot	percent 4.1929	4.7081 2.2963	0.0000 27.0000
Pest infestation rate	1=10-20% ... 7=71-80% 2.4526	1.1587 *** 1.6834	1.0000 7.0000
Weed density rate	2=5-20% ... 5=60-100% 3.1034	0.8370 *** 0.2235	2.0000 5.0000
Weed height	1=<50% ... 5=>125% 2.8944	0.9669 *** 0.3420	1.0000 5.0000
Plant disease rate	1=10-20%...9=91-100% 3.4353	2.2419 1.2603	1.0000 9.0000
Hydromorphic fringe	0=no, 1=yes 0.0280	0.1650 *** 5.7388	0.0000 1.0000

Lowland topography	0=no, 1=yes 0.2866	0.4522 *** 0.9467	0.0000 1.0000
Irrigated plots	0=no, 1=yes 0.0022	0.0464 *** 21.5407	0.0000 1.0000
Rainfall days	days 93.1509	26.4520 0.4210	67.0000 132.0000
Rainfall quantity	centimeters 134.4516	15.1513 -0.1249	108.8300 158.3500
Modern rice variety	percent of seed planted 50.6430	49.8700 -0.0244	0.0000 100.0000
Experience	years 6.0366	3.6790 0.8458	0.0000 22.0000
Gender	0=male, 1=female 0.1853	0.3886 *** 1.6248	0.0000 1.0000
Age	years 47.4957	12.3479 0.1266	20.0000 87.0000
Elementary education	0=no, 1=yes 0.0668	0.2497 *** 3.4810	0.0000 1.0000
Secondary education	0=no, 1=yes 0.0776	0.2675 *** 3.1683	0.0000 1.0000
Some college education	0=no, 1=yes 0.0539	0.2258 *** 3.9647	0.0000 1.0000
Completed college	0=no, 1=yes 0.0065	0.0801 *** 12.3555	0.0000 1.0000
Professional education	0=no, 1=yes 0.0129	0.1130 *** 8.6504	0.0000 1.0000
Unique rice plots	number 1.6616	0.8277 *** 1.1807	1.0000 4.0000
Unique crops	number 2.7220	1.4332 0.3598	1.0000 6.0000

*** = statistically significant at the 99% confidence level.

Table 2
Stochastic Production Frontier Estimates

(t-ratios in parentheses)	Short Specification	Full Specification
Constant	0.0059 (0.004)	-3.6753 (-0.314)
Land (L)	** 1.5557 (2.407)	*** 1.7685 (3.379)
Familial Labor (FL)	0.3518 (1.222)	0.3631 (1.468)
Hired Labor (HL)	0.3294 (1.524)	0.1789 (0.948)
Child Labor (CL)	0.1724 (1.111)	** 0.2628 (1.982)
Animal Traction (AT)	-0.1711 (-0.509)	0.1192 (0.386)
Fertilizers (F)	0.1144 (0.247)	0.4190 (1.259)
L·L	-0.0227 (-0.107)	-0.0039 (-0.022)
L·FL	0.0705 (1.063)	0.0494 (0.793)
L·HL	0.0296 (0.415)	0.0251 (0.381)
L·CL	-0.0052 (-0.147)	-0.0102 (-0.284)
L·AT	-0.0214 (-0.208)	-0.0304 (-0.353)
L·F	0.0462 (0.604)	0.0798 (1.204)
FL·FL	-0.0307 (-0.985)	-0.0038 (-0.108)
FL·HL	-0.0139 (-0.597)	-0.0212 (-0.922)
FL·CL	-0.0124 (-1.002)	** -0.0261 (-2.080)
FL·AT	** -0.0727 (-2.189)	-0.0344 (-1.237)
FL·F	-0.0305 (-1.000)	-0.0320 (-1.134)
HL·HL	-0.0058 (-0.219)	0.0210 (0.804)
HL·CL	-0.0110 (-0.906)	-0.0091 (-0.739)
HL·AT	-0.0178 (-0.519)	-0.0131 (-0.430)
HL·F	** -0.0610 (-2.102)	-0.0293 (-1.200)
CL·CL	* -0.0230 (-1.810)	0.0047 (0.366)

CL·AT	*** 0.0544 (2.636)	** 0.0377 (2.126)
CL·F	0.0264 (1.172)	-0.0068 (-0.355)
AT·AT	0.1929 (1.500)	0.0581 (0.482)
AT·F	0.0299 (0.555)	-0.0709 (-1.622)
F·F	0.1614 (1.178)	0.0389 (0.368)
Erosion		-0.1896 (-0.862)
Fertility		-0.1841 (-1.519)
Aptitude		0.0307 (0.250)
Slope		-0.0638 (-1.279)
$Slope^2$		0.0024 (1.099)
Pests		*** -0.8476 (-3.162)
$Pests^2$		** 0.1029 (2.507)
Weed Density		*** 1.9615 (3.466)
Weed $Density^2$		*** -0.3126 (-3.751)
Weed Height		0.1768 (0.391)
Weed $Height^2$		-0.0135 (-0.183)
Plant Disease		-0.0338 (-0.214)
Plant $Disease^2$		-0.0160 (-1.044)
Hydromorphic Dummy		0.6643 (0.887)
Lowland Dummy		* 0.3983 (1.903)
Irrigated Dummy		-2.7088 (-0.001)
Rain Days		*** 0.4078 (3.209)
Rain $Days^2$		*** -0.0020 (-3.142)
Rainfall		** -0.2506 (-2.334)
$Rainfall^2$		** 0.0008 (2.006)
σ	1.6311	1.3005

λ	***0.9523 (3.713)	*** 1.0601 (4.205)
ln(L)	-801.26	-684.29

***, **, * = statistically significant at the 99, 95, and 90 percent confidence levels, respectively.

Table 3
Production Frontier Technical Efficiency Summary Statistics

Estimation Method	Stochastic, Parametric		DEA	
Specification	Short	Full	Short	Full
Mean	1.3140	1.2552	2.3278	1.1798
Median	1.2496	1.2013	1.9206	1.0000
Standard Dev.	0.2699	0.2104	1.4209	0.4146
Skewness	6.0010	4.8336	2.0618	3.3328
Relative Kurtosis	61.6568	36.2383	6.2162	13.2858
Minimum	1.0358	1.0221	1.0000	1.0000
Maximum	4.7498	3.4067	11.1443	3.9617

Table 4
*Output Elasticity Estimates**

	Short Specification	Full Specification
Land	0.801	0.802
Familial Labor	0.190	0.146
Hired Labor	0.125	0.070
Child Labor	-0.004	0.010
Animal Traction	-0.011	0.0002
Fertilizer	0.025	0.021

*--estimated at the sample means

Table 5

Tests of Concavity and Monotonicity

(frequency with which assumption holds in sample)

	Short Specification	Full Specification
Monotonicity:		
Land	1.000	1.000
Familial Labor	1.000	1.000
Hired Labor	1.000	1.000
Child Labor	0.184	1.000
Animal Traction	0.655	0.949
Fertilizer	1.000	1.000
Concavity		

Table 6

Second-Stage Estimates of Correlates of Technical Inefficiency

	Short Specifications		Full Specifications	
	Stochastic, Parametric	**Nonstochastic, Nonparametric**	**Stochastic, Parametric**	**Nonstochastic, Nonparametric**
Constant	*** 0.3208 (3.48)	-0.1537 (-0.52)	-1.1839 (-0.98)	* 13.4708 (1.70)
Modern	-0.0020 (-0.46)	-0.0172 (-0.98)	-0.0012 (-0.32)	-0.1091 (-0.03)
$Modern^2$	1.9×10^{-5} (0.44)	1.7×10^{-4} (0.95)	1.4×10^{-5} (0.39)	0.0011 (0.03)
Experience	-0.0080 (-1.38)	0.0141 (0.77)	-0.0061 (-1.18)	0.0095 (0.29)
$Experience^2$	3.9×10^{-4} (1.11)	-0.0011 (-1.04)	1.1×10^{-4} (0.36)	-9.8×10^{-4} (-0.48)
Gender	*** 0.0766 (3.69)	0.0582 (0.89)	*** 0.0639 (3.11)	-0.1442 (-1.17)
Age	9.7×10^{-4} (0.27)	-0.0083 (-0.74)	-0.0013 (-0.42)	0.0129 (0.74)
Age^2	-8.8×10^{-6} (-0.25)	8.8×10^{-5} (0.78)	1.5×10^{-5} (0.47)	-7.4×10^{-5} (-0.45)
Elem. Edu.	0.0202 (0.71)	0.0066 (0.08)	** 0.0581 (2.35)	0.0229 (0.17)
Sec. Edu.	-0.0056 (-0.19)	0.0371 (0.39)	0.0159 (0.62)	0.1007 (0.68)
Some Coll.	0.0301 (0.96)	-0.1650 (-1.55)	0.0276 (1.00)	0.2371 (1.63)
Comp. Coll.	-0.0695 (-0.80)	-2.3462 (-0.03)	-0.0841 (-1.12)	-1.7765 (-0.03)
Prof. Deg.	*** 0.2190 (3.47)	0.1516 (0.79)	*** 0.1783 (3.22)	* 0.5074 (1.82)
Plots	-0.0546 (-1.32)	0.2073 (1.53)	0.0048 (0.12)	0.2180 (0.91)
$Plots^2$	0.0101 (1.12)	** -0.0664 (-2.19)	-0.0039 (-0.45)	0.0661 (-1.18)
Crops	-0.0110 (-0.43)	** 0.1808 (2.20)	* 0.0403 (1.72)	0.1360 (0.95)
$Crops^2$	0.0015 (0.37)	* -0.0223 (-1.74)	-0.0047 (-1.30)	-0.0165 (-0.75)
Erosion			0.0139 (0.646)	* 0.2192 (1.65)
Fertility			0.0064 (0.50)	** -0.1559 (-2.09)
Aptitude			-0.0042 (-0.32)	* -0.1563 (-1.84)
Slope			0.0016 (0.31)	-0.0377 (-0.74)
$Slope^2$			-7.7×10^{-5} (-0.39)	-0.0016 (-0.33)

Pests			0.0293 (1.20)	0.0988 (0.69)
$Pests^2$			-0.0044 (-1.31)	-0.0066 (-0.34)
Weed Density			0.0086 (0.15)	-0.5470 (-1.53)
Weed $Density^2$			-4.3×10^{-4} (-0.05)	0.0561 (0.98)
Weed Height			0.0295 (0.70)	0.1926 (0.75)
Weed $Height^2$			-0.0054 (-0.81)	-0.0492 (-1.09)
Plant Disease			0.0169 (1.14)	5.1×10^{-4} (0.01)
Plant $Disease^2$			-6.5×10^{-4} (-0.45)	2.6×10^{-4} (0.03)
Hydromorph			-0.0271 (-0.70)	** -0.4925 (-2.18)
Lowland			*** 0.0578 (3.09)	-0.1174 (-1.01)
Irrigated			0.0122 (0.09)	-1.4348 (-0.01)
Rain Days			0.0087 (0.69)	-0.1294 (-1.60)
Rain $Days^2$			-4.3×10^{-5} (-0.70)	6.5×10^{-4} (1.62)
Rainfall			0.0127 (1.12)	-0.1165 (-1.54)
$Rainfall^2$			-5.0×10^{-5} (-1.12)	4.5×10^{-4} (1.54)
ln(L)	-312.5	-294.9	-231.7	-105.9

t-ratios in parentheses

***, **, * = statistically significant at the 99, 95, and 90 percent confidence levels, respectively.