ELSEVIER
Journal of Development Economics 69 (2002) 85–101
JOURNAL OF Development ECONOMICS
www.elsevier.com/locate/econbase

Smallholder technical efficiency controlling for environmental production conditions

Shane M. Sherlund [a], Christopher B. Barrett [b,*], Akinwumi A. Adesina [c]

[a] *Department of Economics, 1180 Observatory Drive, University of Wisconsin-Madison, Madison, WI 53706-1393, USA*
[b] *Department of Applied Economics and Management, 315 Warren Hall, Cornell University, Ithaca, NY 14853-7801, USA*
[c] *Rockefeller Foundation, PO Box MP167, Mt. Pleasant, Harare, Zimbabwe, USA*

Received 1 October 1999; accepted 1 October 2001

Abstract

Smallholder agricultural production depends heavily on environmental production conditions that are largely exogenously determined. Yet, few data sets collect necessary, detailed information on environmental production conditions. This oversight raises the spectre of likely omitted variables bias because farmers' input choices typically respond in part to environmental conditions. Moreover, because environmental production conditions are rarely symmetrically distributed, the omission also generally leads to upward bias in estimated technical inefficiency and to biased estimates of the correlates of estimated technical inefficiency as well. Using panel data from 464 traditional rice plots in Côte d'Ivoire, we show that controlling for heterogeneous environmental production conditions significantly changes inferences, perhaps especially with respect to smallholder rice farmers' estimated technical inefficiency.

JEL classification: O12; Q12; D2
Keywords: Production frontiers; Agricultural productivity; Rice; Africa (Sub-Saharan)

1. Introduction

In his classic "poor but efficient" hypothesis, Schultz (1964) argued that traditional farmers, given a long enough period of time to learn their production processes, will

* Corresponding author. Tel.: +1-607-255-4489; fax: +1-607-255-9984.
E-mail address: cbb2@cornell.edu (C.B. Barrett).
URL: http://aem.cornell.edu/faculty_sites/cbb2/.

0304-3878/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0304-3878(02)00054-8

identify their respective optimal input and output bundles. Based on this strong belief in the managerial efficiency of smallholders, Schultz advocated an agricultural development policy based on expanding 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.

Yet, 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, rapid advances in biotechnology have led to major increases in potential crop yields and crop tolerance and resistance to drought, pests, and disease—problems which reduce the productivity potential for major crops in developing countries (Conway, 1997). With rapid globalization of major seed and agrochemical industries into developing countries, it is becoming increasingly important for farmers to become more efficient in their ability to access and use available technologies. However, public sector investments in increasing the productivity of farmers in these countries require accurate assessment of the efficiency of farmers and identification of the sources of inefficiencies in order to develop policy and institutional innovations to minimize extant inefficiencies.

A significant subset of the production frontier estimation literature focuses on smallholder agriculture (Greene, 1997; Coelli, 1995; Bravo-Ureta and Pinheiro, 1993; Battese, 1992; Ali and Byerlee, 1991; all provide surveys of this literature). Due mainly to data limitations, however, few such studies control for interfarm heterogeneity in environmental production conditions. Given the extraordinary dependence of smallholder farmers on the underlying agroecology, we conjecture that this omission could partially explain the inconsistency between the mass of empirical results that find considerable smallholder technical inefficiency and the elegant logic of Schultz's longstanding claim.[1]

This paper uses a rich panel data set of rice farmers in the West African nation of Côte d'Ivoire to reconsider inference with respect to technical inefficiency when one controls carefully for environmental production conditions. In particular, we show that the neglect of interfarm heterogeneity in environmental conditions such as pest and weed infestation, plant disease, and rainfall leads not only to obvious omitted variables bias in the estimated parameters of the production frontier, but also to significantly inflated estimates of plot-specific technical inefficiency and to bias in estimates of the correlates of technical inefficiency.

The remainder of the paper is organized as follows. Section 2 briefly reviews the inferential impacts of omitting environmental conditions when estimating a production frontier and conducting inference about technical inefficiency. Sections 3 and 4 describe the data and econometric methodology we use, respectively. Section 5 contains our empirical results and Section 6 concludes.

[1] Our concern harkens back to Hall and Winsten's (1959) response to Farrell's (1957) seminal work on efficiency measurement, wherein they questioned whether comparisons of output across different production conditions really provide useful information on managerial performance.

2. Environmental production conditions

There are industries, such as banking and semiconductors, in which firms have considerable, or even complete, control over their physical production environment. This is not the case, however, in traditional smallholder agriculture, which relies strongly upon environmental conditions that vary markedly over time and space. The environment conditions the results of farmers' production decisions. Otherwise identical producers—same technologies, same abilities—will produce different quantities of grain if faced with different rainfall, plant disease, pest or weed infestation, or other environmental production conditions.[2] Moreover, farmers will adjust commonly measured inputs, such as labor, land, and mineral fertilizer, in response to such environmental conditions. These fundamental features of smallholder agriculture should inform the estimation of production frontiers. In practice, however, few farm production data sets contain detailed, farm- or plot-specific information on the environmental conditions facing producers. Lack of data forces analysts to omit potentially relevant environmental variables, with at least three consequences.

Suppose farmer i produces output, Y_i, using the productive inputs, X_i, in the presence of environmental conditions, W_i, adjusted for the farmer's technical inefficiency, $u_i \geq 0$. Output is assumed to be strictly monotonically increasing in both productive inputs and environmental conditions.[3] This relationship may be estimated as either a nonstochastic production frontier, $Y_i = f(X_i, W_i) - u_i$, or, given mean zero, symmetric sampling and measurement error, v_i, as a stochastic production frontier, $Y_i = f(X_i, W_i) - u_i + v_i$. The production unit achieves technical efficiency if and only if $u_i = 0$. However, the relationship typically estimated in the literature, $Y_i = g(X_i, W_i^*) - u_i^* + v_i^*$, where $W_i^* \subseteq W_i$, omits some or all of the elements of W_i; call these omitted elements $\tilde{W}_i$. As any undergraduate econometrics text explains, this will lead to biased and inconsistent estimates of the parameters of $f(\cdot)$ if $\tilde{W}_i$ is correlated with both X and Y. This omitted relevant variables bias is the first of the three problems that concern us.

The second problem arises from the potentially asymmetric distribution of $\tilde{W}_i$ on the technical inefficiency parameter estimate, u_i^*. If there is variation in sample in $\tilde{W}_i$, then a nonstochastic production frontier estimated without controlling for $\tilde{W}_i$ will necessarily generate $u_i^* \geq u_i$ (recall that $u_i \geq 0$) for any $i = 1,\ldots, n$ for which $\tilde{W}_i - \max_i\{\tilde{W}_i\} < 0$. This problem exists even when estimating a stochastic production frontier, albeit under slightly less general conditions. Suppose that $\tilde{W}_i$ has distribution $\Psi_i \equiv \Psi(\tilde{W}_i)$ and that v_i has distribution $\Phi_i \equiv \Phi(v_i)$. The effects of $\tilde{W}_i$ on Y_i that are uncorrelated with X_i, and thereby picked up as omitted relevant variables bias, will then be absorbed in the composite error term, $v_i^* - u_i^*$. A necessary, but not sufficient, condition for the totality of the effect of $\tilde{W}_i$

[2] It is also likely that environmental 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 would be to relax this assumption.

[3] In this section, we assume that W represents states of nature ordered from worst to best, hence the monotonicity of Y in W. In practice, however, such an ordering may require a nonmonotonic transform of the raw, underlying data since moderate measures may be optimal. In the empirical section to follow, we work with polynomial functions of the raw data.

to be captured by the statistical residual v_i^*, and not to affect the technical inefficiency parameter u_i^*, is that the Ψ_i and Φ_i distributions differ from one another only by location and scale parameters. In general, Ψ_i can be represented as a mixture of two distributions, Γ_i and Φ_i, the latter possibly transformed by a location parameter, α, and a scale parameter, β, or both, such that $\Psi(\tilde{W}_i) = \lambda\Gamma(\tilde{W}_i)+(1-\lambda)\Phi(\alpha+\beta\tilde{W}_i)$, where $\lambda\in[0,1]$. If $\lambda>0$, then u_i^* will capture part of the effect of the omitted environmental variables because of the deviation from location and scale differences introduced through Γ_i. If Ψ_i is asymmetric and Φ_i is symmetric (Φ_i is usually assumed to be normal), then it must be the case that $\lambda>0$ since there is more than a location-scale difference between the two distributions. Therefore, under the standard assumptions of symmetrically distributed statistical error and asymmetrically distributed technical inefficiency, u_i^* is an upwardly biased and inconsistent estimator of u_i because $E[u_i^*]>u_i$ for some i.

The third problem arises with respect to identifying the correlates of technical inefficiency. While knowing the extent of technical inefficiency prevailing in a sector is useful, it is helpful to know also the potential sources of technical inefficiency so as to target interventions appropriately and thereby potentially reduce inefficiency. Suppose that technical inefficiency is related to managerial variables, Z_i, up to statistical error, ξ_i, via the relationship $u_i=p(Z_i)+\xi_i$. However, if the technical inefficiency estimates generated by a model omitting $\tilde{W}_i$ are biased estimates of u_i, then regressing those biased u_i^* on Z_i will likewise yield biased and inconsistent estimates of the parameters of the relationship of interest, $p(\cdot)$.

Therefore, omission of environmental production conditions that intuitively affect both output and inputs subject to farmer control leads to biased estimates of the parameters describing the production frontier, overstatement of technical inefficiency, and biased estimates of the correlates of true technical inefficiency. In the remainder of this paper, we demonstrate these problems in one particular case: rice production in the West African nation of Côte d'Ivoire during 1993–1995. We estimate the production frontier directly "rather than the dual cost or profit frontiers" for several reasons. First, using observed market prices in estimating the production behavior of smallholders (most of whose labor, land, and animal allocation decisions do not involve market transactions) creates many inferential problems (Barrett, 1997). Second, Zellner et al. (1966) point out that a production function may be estimated consistently if the farm manager chooses his or her inputs and output to maximize expected, not actual, profits.

3. Data

The data we use come from the farm management and household survey (FMHS) fielded by the West Africa Rice Development Association (WARDA). The WARDA-FMHS tracked 120 randomly selected rice-producing households in Côte d'Ivoire during 1993–1995. Twenty-two survey instruments were administered (at least) annually and are described in detail in WARDA (1997). Because our interest is in traditional smallholder rice production of the sort Schultz hypothesized about, we exclude data from a few atypical mechanized plots in specialized development projects, as well as observations

Table 1
Descriptive statistics

Variable	Mean	Median	Std. Dev.	Skewness	Minimum	Maximum
Rice production (kg)	1676.45	1333.08	1399.14	5.9E + 9 *	46.62	10 094.02
Land area (ares)	94.10	74.50	80.84	1.5E + 6 *	4.13	710.00
Adult family labor (hours)	470.46	356.50	400.33	1.0E + 8	0.00	2545.50
Adult hired labor (hours)	298.46	232.00	262.14	3.4E + 7 *	0.00	1984.00
Child labor (hours)	408.24	125.00	640.06	6.2E + 8 *	0.00	3662.00
Chemical fertilizers (kg)	17.52	0.00	51.96	5.0E + 5 *	0.00	350.00
Soil erosivity (0 = N, 1 = Y)	0.39	0.00	0.49	0.05	0.00	1.00
Soil fertility (indexed[a])	1.77	2.00	0.65	0.07	1.00	3.00
Soil aptitude (indexed[a])	1.47	1.00	0.61	0.21	1.00	3.00
Plot slope (%)	4.19	3.00	4.71	238.87 *	0.00	27.00
Pest infestation (indexed[b])	2.45	2.00	1.16	4.17 *	1.00	7.00
Weed density (indexed[c])	3.10	3.00	0.84	1.18 *	2.00	5.00
Weed height (indexed[d])	2.89	3.00	0.97	2.25 *	1.00	5.00
Plant disease (indexed[b])	3.44	3.00	2.24	26.98 *	1.00	9.00
Uplands (0 = N, 1 = Y)	0.69	1.00	0.46	– 0.08	0.00	1.00
Hydromorphic fringe (0 = N, 1 = Y)	0.03	0.00	0.17	0.03 *	0.00	1.00
Lowlands (0 = N, 1 = Y)	0.29	0.00	0.45	0.09	0.00	1.00
Rainy days (days)	93.15	80.00	26.45	7767.67	67.00	132.00
Rainfall (cm)	134.45	132.86	15.15	– 432.97	108.83	158.35
Year 1993 (0 = N, 1 = Y)	0.25	0.00	0.43	0.09	0.00	1.00
Year 1994 (0 = N, 1 = Y)	0.38	0.00	0.49	0.06	0.00	1.00
Year 1995 (0 = N, 1 = Y)	0.37	0.00	0.48	0.06	0.00	1.00
Guinean savannah (0 = N, 1 = Y)	0.25	0.00	0.43	0.09	0.00	1.00
Transition zone (0 = N, 1 = Y)	0.36	0.00	0.48	0.06	0.00	1.00
Equatorial forest (0 = N, 1 = Y)	0.39	0.00	0.49	0.05	0.00	1.00
Rice variety (% modern)	50.64	100.00	49.87	– 3012.47	0.00	100.00
Experience (years)	6.04	5.00	3.68	41.98	0.00	22.00
Gender (0 = M, 1 = F)	0.19	0.00	0.39	0.10	0.00	1.00
Age (years)	47.50	48.00	12.35	237.59	20.00	87.00
No education (0 = N, 1 = Y)	0.78	1.00	0.41	– 0.10	0.00	1.00
Elementary education (0 = N, 1 = Y)	0.07	0.00	0.25	0.05 *	0.00	1.00
Secondary education (0 = N, 1 = Y)	0.08	0.00	0.27	0.06 *	0.00	1.00
College or higher (0 = N, 1 = Y)	0.07	0.00	0.26	0.06 *	0.00	1.00
Rice plots (plots)	1.66	1.00	0.83	0.67	1.00	4.00
Total crops (unique crops)	2.72	3.00	1.43	1.06	1.00	6.00

[a] 1 = Good, 2 = average, 3 = poor.
[b] 1 = 10–20%, 2 = 21–30%,. . ., 9 = 91–100% of crop.
[c] 2 = 5–20%, 3 = 21–40%, 4 = 41–60%, 5 = 61–100%.
[d] 1 = < 50%, 2 = 50–99%, 3 = 100%, 4 = 101–125%, 5=>125% of rice plant height.
* Statistically significant at the 95% confidence level (one-sided *z*-test).

with (nonsystematically) missing data, leaving us with an unbalanced panel of 464 rice plots for estimation.[4]

The comparative advantage of the WARDA-FMHS data is its inclusion of plot-level measurements of environmental production conditions, such as pest and weed infestation,

[4] A total of 88 observations are lost due to missing data, 37 observations due to mechanization.

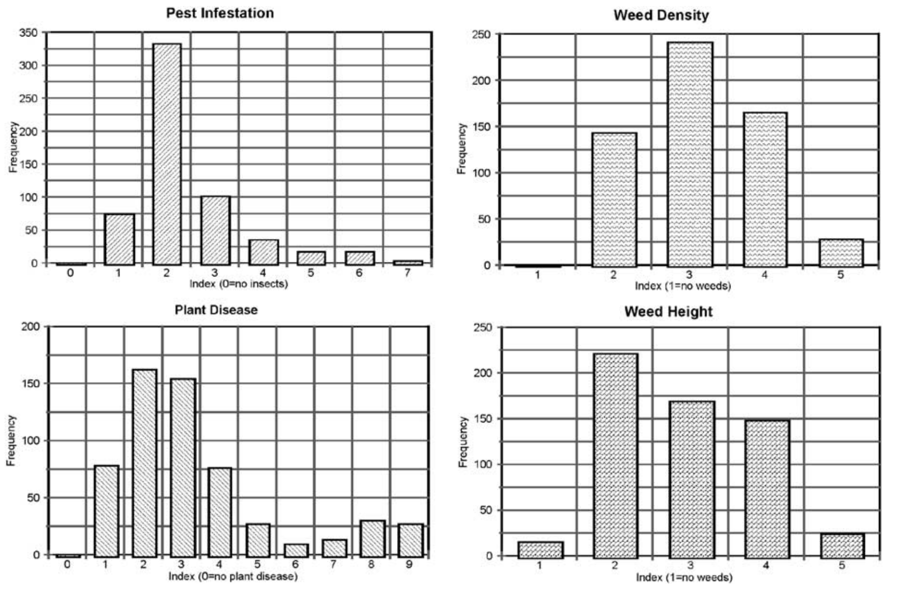

Fig. 1. Empirical distributions of environmental production conditions.

plant disease, plot slope, and soil quality.[5] Sample descriptive statistics are presented in Table 1 for the relevant variables (land, adult family labor, adult hired labor, child labor, chemical fertilizer usage, soil erosivity, soil fertility, soil aptitude for rice cropping, pest infestation, weed density, weed height, plant disease, topographic location, plot slope, rainy days, rainfall, rice variety, rice cropping experience, gender, age, education, rice plots, total crops, region, and year). As is evident, there is relatively little use of chemical fertilizer usage—this is prototypical smallholder, traditional rice cropping—and considerable variation in land and labor use patterns, as well as in environmental production conditions. Improving land and labor productivity is central to the task of agricultural development among these farmers.

The previous section pointed out that if environmental production conditions are not symmetrically distributed, then their omission will lead to upward-biased estimates of plot-specific technical inefficiency. As reflected in Table 1 and shown graphically for a few variables in Fig. 1, these environmental variables are asymmetrically distributed, with statistically significant positive skewness. Therefore, the problems identified in the previous section appear to be relevant in this data set, affording us an opportunity to

[5] Data on pests, weeds, and disease were collected during four different stages of crop growth (vegetative, reproductive, flowering, and mature) in highly disaggregated form. Crop vulnerability to different stresses varies significantly across growth stages and specific stresses. In order to conserve degrees of freedom, we use the maximal value across stages and specific stresses for each category. This therefore captures maximal threat exposure, although that need not equal maximal impact since different types of weeds, pests, or disease have different effects. This distinction should be kept in mind. We cannot isolate the effects of specific types of weeds, pests, disease or the impact of biotic stresses at particular stages of plant growth.

Table 2
Correlation matrix relating X_{it} and W_{it} variables

Variable	Land	Family labor	Hired labor	Child labor	Fertilizers
Soil erosivity	0.190	0.064	0.037	0.235	− 0.209
Soil fertility	0.051	0.254	− 0.169	0.281	0.232
Soil aptitude	0.193	0.265	0.012	0.227	0.034
Soil slope	0.329	0.445	0.090	0.427	− 0.118
Pest infestation	− 0.041	0.007	− 0.180	0.310	0.532
Weed density	0.088	− 0.076	− 0.035	0.189	− 0.018
Weed height	0.220	0.205	− 0.052	0.490	0.224
Plant disease	0.267	0.305	− 0.067	0.508	− 0.062
Rainy days	− 0.117	− 0.026	0.114	− 0.452	− 0.303
Rainfall	− 0.127	− 0.191	0.049	− 0.390	− 0.175

examine the consequences of the omission of measurable environmental production conditions.

One might justifiably challenge the exogeneity of what we label environmental production conditions. Smallholders can influence some of these variables, either within the season through the application of labor to combat pests, weeds, or disease, or of fertilizers to ameliorate poor soil quality, or across years, through land improvements to improve soil quality or reduce plot slope or soil erosivity. Our W_i variables thus encompass not only truly exogenous variables (e.g., rainfall), but also quasi-fixed characteristics (e.g., plot slope and soil erosivity) and outcomes that are jointly the product of exogenous shocks and managerial response (e.g., pest and weed infestation). As the correlation coefficients reported in Table 2 show, the relationship between the X_i and W_i variables is generally weak, but nonzero.[6] Exploring this relationship is a topic worthy of separate investigation. For present purposes, however, the nonzero correlation between X_i and W_i merely serves to underscore the potential omitted variables bias plaguing studies that omit W_i.

4. Estimation procedures

The empirical production frontier literature generally follows one of two methods. The preponderance of the published literature follows the stochastic production frontier approach independently pioneered by Aigner et al. (1977) and Meeusen and van den Broeck (1977). In this approach, one specifies a priori a functional form for the production frontier (e.g., Cobb–Douglas or translog) and probability density functions for the asymmetric technical inefficiency parameter (usually the half-normal or truncated normal) and the symmetric statistical error parameter (usually the normal). Then, the (log-) likelihood function may be written out and maximum likelihood used to estimate the parameters of interest. The second approach uses nonparametric, data envelopment

[6] With a sample size of 464, one can reject the null hypothesis of $r = 0$ at the 5% confidence level for any $|r| \geq 0.091$.

analysis (DEA) methods to impose monotonicity and concavity properties on the estimated frontier and otherwise make no functional form or distributional assumptions about the shape of the production frontier. This flexibility comes at the considerable cost, however, of an assumed absence of measurement or sampling error (Färe et al., 1994).[7] We focus on the former, stochastic parametric approach, but briefly present DEA results as well simply to demonstrate that our findings are robust to the estimation method employed.

Jondrow et al. (1982) show how to estimate the conditional expectation of the plot-specific technical efficiency parameter (conditional upon the composed error term, $v_i - u_i$) in stochastic parametric frontier estimation. One problem with this approach, however, is that the technical inefficiency parameter is assumed to be independently and identically distributed. This clearly is not the case if we suspect (and find) that the technical inefficiency parameter is related to variables such as managerial characteristics and practices that vary across firms.

To combat this potential problem, some studies suggest estimating the production frontier and the relationship between technical inefficiency and the sources of inefficiency jointly, rather than in a two-step procedure. Kumbhakar et al. (1991) generalize the stochastic production frontier model of Aigner et al. by specifying that the distribution of the technical inefficiency parameter be the positive truncation of a normal distribution with variable mean $Z_i\delta$, i.e., $u_i \sim N^+(Z_i\delta,\ \sigma_u^2)$, where δ is a vector of parameters to be estimated. Reifschneider and Stevenson (1991) instead propose that $u_i = Z_i\delta + \xi_i \geq 0$, where $\xi_i \sim N^+(0,\ \sigma_\xi^2)$. However, the latter method does not guarantee that $u_i \geq 0$. Huang and Liu (1994) take a slightly different approach. They specify $\xi_i \sim N(0,\ \sigma_\xi^2)$ and truncate this from below at the variable truncation point, $-Z_i\delta$. Huang and Liu also allow for interactions between the productive inputs and the managerial variables in the technical inefficiency relationship.

Other studies have concentrated on the panel data aspects of production frontier estimation. Pitt and Lee (1981) implement a random effects treatment to estimate a stochastic production frontier. Cornwell et al. (1990), Kumbhakar (1990, 1991), Battese and Coelli (1992), and Lee and Schmidt (1993) allow the technical efficiency parameter to vary across time via time-specific dummy variables or according to a specified functional form. However, in these models, the technical inefficiency parameter is assumed to follow the same pattern over time for all firms. Battese and Coelli (1995) generalize the model of Huang and Liu to allow for panel data, though not explicitly allowing for interactions between the inputs and managerial variables in the technical inefficiency relationship. This model allows the technical inefficiency parameter, and hence technical efficiency, to vary across time in a potentially different, but predictable, manner across firms.

We implement Battese and Coelli's technical inefficiency effects model for panel data. In this model, the technical inefficiency parameter is related to a vector of farmer-specific managerial variables subject to statistical error, so that $u_{it} = Z_{it}\delta + \xi_{it} \geq 0$, where

[7] The output-oriented, variable returns to scale, strong disposability DEA model may be written as: $\theta^*(X_i,Y_i|\mathrm{VRS,SD}) = \mathrm{Max}_{\theta,z}\ \theta$, subject to $\theta Y_i \leq zY$, $zX \leq X_i$, $\Sigma_i z_i = 1$, and $z \in \mathbf{R}_+^N$, where $i = 1,\ldots, N$ and z is the activity vector indicating to which plots the ith 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 ith plot been fully efficient.

$\xi_{it} \sim N(0, \sigma_\xi^2)$, i indexes firms, and t indexes time. However, since $u_{it} \geq 0$, $\xi_{it} \geq -Z_{it}\delta$ so that the distribution of ξ_{it} is truncated from below at the variable truncation point, $-Z_{it}\delta$. The statistical error of the production frontier is assumed to be mean zero, normally distributed with variance $\sigma_v{}^2$. Then the log-likelihood function for the ith firm at time t takes the form:

$$\begin{aligned}\ln L_{it} = & -\frac{1}{2}\left[\ln(2\pi) + \ln(\sigma^2)\right] - \frac{1}{2\sigma^2}\left[y_{it} - f(X_{it}, W_{it}; \beta, \theta) + Z_{it}\delta\right]^2 \\ & - \ln[\Phi(d_{it})] + \ln[\Phi(d_{it}^*)],\end{aligned} \tag{1}$$

where $f(X_{it}, W_{it}; \beta, \theta)$ is the production frontier, β, θ, δ, γ, and σ^2 are the parameters to be estimated, $d_{it} = Z_{it}\delta/(\gamma\sigma^2)^{1/2}$, $d_{it}^* = \{(1-\gamma)Z_{it}\delta - \gamma[y_{it} - f(X_{it}, W_{it}; \beta, \theta)]\}/[\gamma(1-\gamma)\sigma^2]^{1/2}$, and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Note that, under this parameterization, $\gamma = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$. We use the translog specification for the production frontier.[8]

In order to examine the consequences of omitting environmental production conditions, we estimate the production frontier with and without the environmental variables. The traditional, or ‘‘short’’ specification, which omits the W_{it} variables (i.e., $Y_{it} = g(X_{it}, \emptyset) - u_{it}^* + v_{it}^*$), may be written as:

$$\begin{aligned}\ln(Y_{it}) = & \beta_0^* + \sum_{k=1}^{K} \beta_k^* \ln(X_{ikt}) + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{K} \gamma_{jk}^* \ln(X_{ijt})\ln(X_{ikt}) \\ & - u_{it}^* + v_{it}^*, \qquad u_{it}^* = Z_{it}\delta^*/ + \xi_{it}^*,\end{aligned} \tag{2}$$

where β_0^*, β_k^*, and γ_{jk}^* ($j, k = 1,\ldots, K$) and δ^* are parameters to be estimated. Output (Y_{it}) is rice production. The productive inputs (X_{it}) are land, adult family labor, adult hired labor, child labor, and chemical fertilizer usage. Managerial variables (Z_{it}) include the proportion of area planted in modern rice varieties, rice cropping experience, gender, age, education dummies, number of rice plots cultivated, number of total crops cultivated, as well as region- and year-specific dummies. Units of measure and descriptive statistics for each variable are provided in Table 1.

[8] The presence of many zero-valued observations is troublesome. The convention in much literature is to set ln(0) = 0. However, due to observations taking on values in the range (0,1), setting ln(0) = 0 implicitly reorders observations with respect to that subspace. Therefore, we instead set ln(0) = ln (ζ/10), where ζ is the smallest strictly positive observation in the sample. We tried to address the problem of zero-valued observations instead through the use of other flexible functional forms, e.g., generalized Leontief, CES-CT-GL, and symmetric generalized McFadden, but all failed diagnostic tests for satisfaction of regularity conditions (due to Waldman, 1982).

As stated previously, we assume the separability of X_{it}, W_{it}, and Z_{it} in order to conserve degrees of freedom in estimating the "full" specification, $Y_{it}=f(X_{it}, W_{it}) - u_{it} + v_{it}$, or, written as above:

$$\ln(Y_{it}) = \beta_0 + \sum_{k=1}^{K} \beta_k \ln(X_{ikt}) + \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{K} \gamma_{jk} \ln(X_{ijt}) \ln(X_{ikt}) + W_{it}\theta - u_{it} + v_{it}, \qquad u_{it} = Z_{it}\delta + \xi_{it}, \tag{3}$$

where θ is a vector of estimable parameters, conformable with W_{it}, including the categorical variables soil erosivity, soil fertility, soil aptitude, pest infestation, weed density, weed height, plant disease, topographic location dummies, and region- and year-specific dummies, as well as the continuous variables plot slope, rainy days, and rainfall. Because both rainfall measures are common to all plots in a region, they also capture some year-and region-specific unobserved heterogeneity and should, therefore, be interpreted with care.

5. Results

Parameter estimates for both the short and full specifications are reported in Table 3 (standard errors in parentheses). The statistical superiority of the full specification is apparent in a likelihood ratio test statistic of 217, which has a p-value of 0 against the $\chi^2(23)$ distribution.[9] The parameter estimates under the full specification indicate that output is statistically significantly correlated with land, hired adult labor, and chemical fertilizer usage. This is consistent with existing observations of these rice systems, as land and labor are both important determinants of output.[10] Adult-hired labor and chemical fertilizer usage are substitute inputs, as evidenced by their statistically significant negative second-order effect.

Environmental conditions clearly affect output, which is statistically significantly higher in the lowlands and the hydromorphic fringe, where soils are heavier and proximity to ground water improves moisture availability (relative to the uplands) during dry periods. Output is also favorably influenced by rain spread over many days (around 90 rainy days per year are ideal) and moderate amounts of rainfall (around 1120 mm of rain per year appears "optimal"). Rice output decreases significantly with above-average weed density, high plant disease rates, and low soil fertility measures. Farmers located in the equatorial forest regions of Côte d'Ivoire produce significantly more rice than do farmers of the other two regions (Guinean savanna and a transition zone).

Returning to the three concerns expressed in Section 2 as to how omission of environmental production conditions might affect inference, we first see a significant

[9] The Cobb–Douglas functional form is rejected in favor of the translog under both the full and short specifications, with Wald test statistics of 27 and 87 and corresponding p-values of 0.028 and 0, respectively, against the appropriate χ^2 distributions. Constant returns to scale is also rejected under both specifications, with a Wald test statistic of 71 and 111 for the full and short specifications, respectively, each having a p-value of 0.

[10] There appears to be considerable variation in estimated elasticities of labor and land in these systems, depending upon the data set and specification used (Adesina and Djato, 1997a,b; Sherlund, 1998; Dalton, 1999). Much is ecosystem-dependent in this setting.

Table 3
Stochastic production frontier estimates

Parameter	Without environmental variables estimate (std. error)	With environmental variables estimate (std. error)
Constant	3.959 (0.646)****	– 19.16 (1.043)****
Land	0.190 (0.178)	0.799 (0.150)****
Family labor (FLabor)	0.034 (0.145)	0.030 (0.111)
Hired labor (HLabor)	0.353 (0.095)****	0.072 (0.076)
Child labor (CLabor)	0.100 (0.039)**	0.025 (0.033)
Chemical fertilizers (Chem)	0.197 (0.074)***	0.167 (0.056)***
$\frac{1}{2}$ Land^2	0.068 (0.049)	0.041 (0.046)
$\frac{1}{2}$ FLabor^2	0.033 (0.024)	0.014 (0.018)
$\frac{1}{2}$ HLabor^2	0.027 (0.013)**	0.015 (0.010)
$\frac{1}{2}$ CLabor^2	0.016 (0.005)***	0.003 (0.004)
$\frac{1}{2}$ Chem^2	0.043 (0.024)*	0.049 (0.022)**
Land × FLabor	0.057 (0.031)*	0.001 (0.029)
Land × HLabor	– 0.002 (0.022)	– 0.011 (0.018)
Land × CLabor	– 0.008 (0.009)	0.001 (0.007)
Land × Chem	– 0.027 (0.017)	– 0.027 (0.014)*
FLabor × HLabor	– 0.059 (0.017)****	– 0.008 (0.015)
FLabor × CLabor	– 0.012 (0.006)*	– 0.007 (0.005)
Flabor × Chem	0.006 (0.010)	0.0002 (0.008)
HLabor × CLabor	– 0.003 (0.006)	0.0007 (0.004)
HLabor × Chem	– 0.015 (0.004)****	– 0.009 (0.004)***
CLabor × Chem	– 0.002 (0.004)	– 0.001 (0.003)
Soil erosivity	–	– 0.031 (0.048)
Soil fertility	–	– 0.090 (0.027)****
Soil aptitude	–	0.006 (0.026)
Plot slope	–	– 0.018 (0.010)*
Slope^2	–	0.0005 (0.0004)
Pest infestation	–	– 0.083 (0.053)
Pests^2	–	0.010 (0.007)
Weed density	–	0.459 (0.123)****
Density^2	–	– 0.074 (0.019)****
Weed height	–	– 0.107 (0.082)
Height^2	–	0.012 (0.013)
Plant disease	–	– 0.018 (0.032)
Disease^2	–	– 0.002 (0.003)
Hydromorphic fringe	–	0.187 (0.070)***
Lowlands	–	0.101 (0.043)**
Rainy days	–	0.293 (0.047)****
Days^2	–	– 0.002 (0.0002)****
Rainfall	–	0.186 (0.040)****
Rainfall^2	–	– 0.0008 (0.0002)****
Year 1994	–	0.034 (0.114)
Year 1995	–	– 0.156 (0.062)***
Transition zone	–	– 0.672 (0.354)*
Equatorial forest zone	–	1.422 (0.255)****
$\sigma^2=\sigma_u^2+\sigma_v^2$	0.103 (0.007)	0.251 (0.026)
$\gamma=\sigma_u^2/(\sigma_u^2+\sigma_v^2)$	0.520 (0.046)	0.881 (0.020)
Log-likelihood	– 129.79	– 21.31

* Statistically significant at the 90% confidence level.
** Statistically significant at the 95% confidence level.
*** Statistically significant at the 99% confidence level.
**** Statistically significant at the 99.9% confidence level.

effect on the estimates of the parameters of the production frontier itself. This is somewhat apparent in Table 3's raw parameter estimates, but appears more readily in Table 4, which reports descriptive statistics of mean plot-specific output elasticity estimates for the five productive inputs and some of the environmental variables. Some key qualitative results are similar for both specifications. Output is most responsive to land under cultivation and chemical fertilizer usage. Adult-hired and family labor have essentially indistinguishable output elasticities, while output is only about one-third as responsive to added child labor as it is to adult labor. The major difference is that under the full specification, the elasticity of rice output with respect to land is 20% higher than it is under the short specification, whereas the elasticities with respect to adult family labor, adult-hired labor, and child labor each fall by more than 70%. Once one controls for environmental production conditions, output appears less responsive to variation in labor allocation and far more responsive to changes in cultivated area. This almost surely reflects two intuitive patterns. First, output expansion tends to be at the extensive margin, onto less fertile soils, so when one fails to control for environmental conditions negatively correlated with land quality, one understates the true output responsiveness to land ceteris paribus. Second, labor application rates increase when environmental production conditions are more favorable. Therefore, the marginal physical product of labor in the traditional, short regression picks up part of the environmental effects with which labor is positively correlated.

The next concern related to estimates of technical inefficiency. We use the Battese and Coelli (1993) method to estimate the conditional expectation of plot-specific technical inefficiency (conditional upon the composed error term, $\varepsilon_{it} = v_{it} - u_{it}$). This conditional expectation is calculated as:

$$E[\exp\{-u_{it}\} \mid \varepsilon_{it}] = \exp\left\{-\mu_* + \frac{1}{2}\sigma_*^2\right\}\frac{\Phi(\mu^*/\sigma^* - \sigma^*)}{\Phi(\mu^*/\sigma^*)}, \tag{4}$$

where $\mu_* = (1-\gamma)Z_{it}\delta - \gamma\varepsilon_{it}$ and $\sigma_*^2 = \gamma(1-\gamma)\sigma^2$. Fig. 2 plots the empirical cumulative distribution functions of the estimated technical efficiency scores for both specifications

Table 4
Output elasticity estimates[a]

Variable	Without environmental variables	With environmental variables
Land	0.7902 (0.1177)	0.9481 (0.0665)
Adult family labor	0.1543 (0.1019)	0.0315 (0.0237)
Adult hired labor	0.1563 (0.0797)	0.0384 (0.0231)
Child labor	0.0514 (0.0176)	0.0138 (0.0104)
Chemical fertilizers	0.2056 (0.0654)	0.1877 (0.0467)
Soil slope	–	– 0.0357 (0.0365)
Pest infestation	–	– 0.0616 (0.0691)
Weed density	–	– 0.0991 (0.4212)
Weed height	–	– 0.0802 (0.0483)
Plant disease	–	– 0.1364 (0.1380)
Rainy days	–	– 3.2073 (9.3051)
Rainfall	–	– 5.4056 (3.9300)

[a] Mean (standard deviation) of plot-specific elasticity estimates.

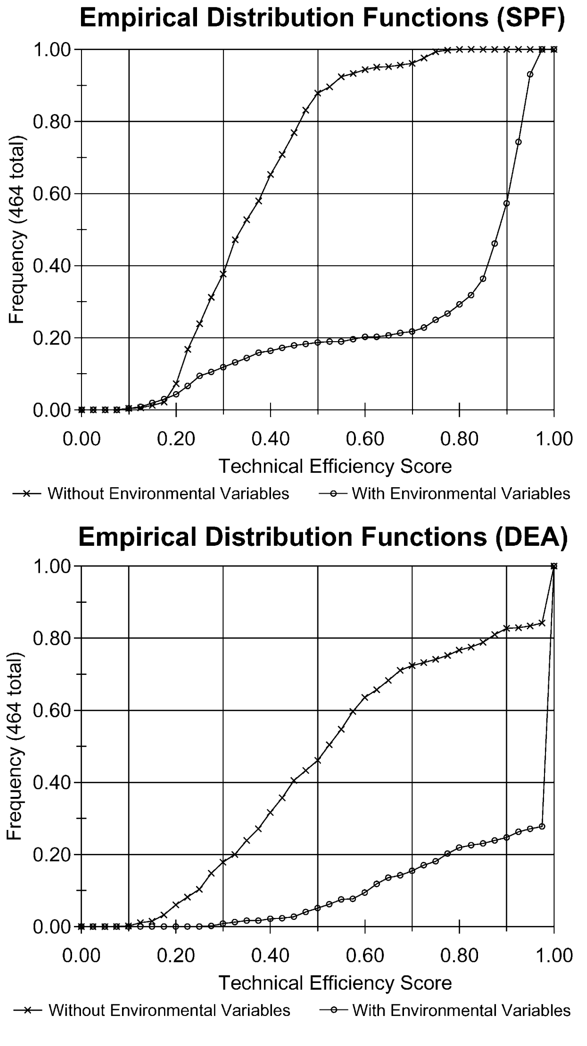

Fig. 2. Distribution functions for estimated plot-specific technical efficiencies.

under both the stochastic parametric and the DEA methods. Descriptive statistics for these estimates are reported in Table 5. The unambiguous reduction in estimated technical inefficiency appears as a near first-degree stochastic dominance of the full over the short specification's empirical cumulative distribution function under both methods, as well as a large increase in the mean and median of estimated technical efficiency.[11] This provides

[11] This is more than just a location shift in the distribution functions. The Pearson product-moment rank correlation between the estimated technical efficiency scores generated under the short and full specifications is only 0.43 under the stochastic parametric method and 0.35 under DEA. Hence, plots are also being re-ranked between the two specifications.

Table 5
Technical efficiency summary statistics

	Stochastic production frontier		DEA frontier	
	Without environmental variables	With environmental variables	Without environmental variables	With environmental variables
Mean	0.3595	0.7656	0.5618	0.9059
Median	0.3388	0.8871	0.5207	1.0000
Minimum	0.0798	0.0900	0.0897	0.2524
Maximum	0.7751	0.9649	1.0000	1.0000

empirical support for our claim that the omission of environmental production conditions leads to substantial upward bias in estimates of technical inefficiency when the W_{it} are asymmetrically distributed.

The third concern we raised in Section 2 concerned the effect of biased technical inefficiency estimates on the estimated parameters of the relationship between inefficiency and managerial characteristics. We find that omission of environmental production conditions indeed significantly affects the estimates of the relationship between estimated technical inefficiency and managerial characteristics and choice variables. The parameter estimates for this relationship for both the full and short specifications are reported in

Table 6
Sources of technical inefficiency

Parameter	Without environmental variables estimate (std. error)	With environmental variables estimate (std. error)
Constant	0.735 (0.423) *	– 2.134 (0.765)* * *
Rice variety	– 0.0009 (0.0005) * *	0.003 (0.001)* * *
Experience	– 0.0001 (0.013)	– 0.023 (0.032)
Experience^2	0.000009 (0.0008)	0.00007 (0.002)
Gender	– 0.025 (0.058)	0.247 (0.190)
Age	0.004 (0.008)	0.004 (0.026)
Age^2	– 0.00003 (0.00008)	0.00003 (0.0003)
Elementary education	0.066 (0.063)	0.671 (0.207)* * *
Secondary education	0.029 (0.067)	0.628 (0.274) * *
College or higher	0.114 (0.067) *	1.080 (0.223)** * *
Rice plots	0.083 (0.105)	1.012 (0.379)* * *
Plots^2	– 0.036 (0.023)	– 0.298 (0.096)* * *
Total crops	0.132 (0.063) * *	0.303 (0.144) * *
Crops^2	– 0.021 (0.010) * *	– 0.034 (0.020) *
Year 1994	0.302 (0.044)** * *	2.016 (0.283)** * *
Year 1995	0.195 (0.046)** * *	1.598 (0.257)** * *
Transition zone	0.317 (0.091)** * *	– 2.026 (0.325)** * *
Equatorial forest zone	– 0.425 (0.068)** * *	– 3.074 (0.273)** * *
Managerial variables	255.11** * *	351.98** * *
Less year and region	25.29 * *	52.82** * *

* Statistically significant at the 90% confidence level.
** Statistically significant at the 95% confidence level.
*** Statistically significant at the 99% confidence level.
**** Statistically significant at the 99.9% confidence level.

Table 6.[12,13] The estimated relationships between the technical inefficiency parameter, u_{it}, and the correlates are broadly similar across both regressions, both somewhat more intuitive and considerably more precise when this parameter is generated by the full specification. For example, year and region dummy variables are highly significant in both regressions. Yet, in the full specification, technical efficiency appears greatest in the wetter equatorial zone and least in the drier savannah zone, while in the short specification, efficiency appears significantly greater in the drier savannah zone than in the transition zone of intermediate rainfall. This is counterintuitive and inconsistent with the mass of previous studies of Ivorien rice production. When the technical inefficiency parameter, u_{it}, is generated from the full specification controlling for environmental production conditions, inefficiency is lower for unschooled farm managers,[14] for those who cultivate three or more rice plots, and for those who specialize in rice production. By contrast, technical inefficiency appears to be increasing in the proportion of area planted in modern rice varieties—perhaps signaling smallholder unfamiliarity with how best to grow these varieties—when farm managers diversify risk by planting multiple crops, and during 1994 and 1995, following the massive January 1994 devaluation of the local currency. When the annual and regional dummy variables are excluded, the managerial variables are overwhelmingly statistically significant when one has previously controlled for environmental production conditions, with a Wald test statistic of nearly 53 and a p-value of 0 against the $\chi^2(13)$ distribution. Without having controlled for environmental production conditions in the first stage estimation, the corresponding Wald statistic falls to 25, which has a p-value of 0.021 against the same $\chi^2(13)$ distribution. More careful control for environmental production conditions not only affects the estimates of the production frontier parameters and sharply reduces estimated technical inefficiency, it also (relatedly) improves the precision with which one can explain apparent technical inefficiency.

6. Conclusions

This paper is motivated by a concern that the empirical work on the technical efficiency of smallholder farmers does not always adequately control for environmental production conditions. This may cause analysts to draw false inferences, with undesirable consequences for the design and effects of the policies informed by such inferences.

We first explain why prevailing empirical methods, using either deterministic or stochastic production frontier methods, likely yield biased and inconsistent estimates of the parameters of production frontiers, plot-specific technical inefficiency, and the parameters relating estimated technical inefficiency to managerial characteristics. We demonstrate

[12] The parameter estimates of Tables 3 and 6 are estimated jointly by full information maximum likelihood. We report them separately for ease of presentation.

[13] While for the sake of brevity we do not report them here, analogous results emerge when one uses DEA-generated technical inefficiency estimates in the second stage regression.

[14] This likely reflects that farming is a secondary occupation for those with formal schooling, who focus primarily on the superior income distribution available through nonfarm employment based on education and skills (Barrett et al., 2001). Farming gets relatively less of their attention and thus exhibits greater technical inefficiency.

our claim in the case of smallholder rice production in the West African nation of Côte d'Ivoire during 1993–1995. Using detailed plot-level panel data, we show that controlling for measurable environmental production conditions yields significantly lower estimates of technical inefficiency, different output elasticity estimates, and more intuitive and precise estimates of the sources of technical inefficiency. Using the technical inefficiency effects model for panel data suggested by Battese and Coelli (1995), the median Ivorien rice plot exhibits relatively little technical inefficiency once appropriate measures are taken to control for environmental production conditions. However, when these controls are absent, technical inefficiency estimates become contaminated in a predictable way: they rise sharply. In an area where upland smallholders have been cultivating rice with similar technologies for millennia, the technical efficiency we find once proper control is made for environmental production conditions seems far more plausible than the widespread inefficiency common to efficiency studies without such controls.

These results have significant policy implications since the extent of estimated technical inefficiency prevailing in an agricultural economy matters to the determination of whether scarce agricultural development funds are best spent to develop improved technologies or to teach farmers how to better use existing technologies. T.W. Schultz appears to be right when one compares Ivorien rice farmers against the estimated production frontier they face, conditional upon their idiosyncratic realization of environmental production conditions, rather than against the best-realized production frontier, which implicitly pits them against colleagues enjoying considerably more favorable realized environmental shocks to production.

Acknowledgements

We thank the West Africa Rice Development Association for making the data available, and David Aadland, Tim Dalton, Moustapha Gayé, Doug Gollin, Dayri Mylène Kamara, Loren Tauer, Quinn Weninger, Alex Winter-Nelson, an anonymous referee, and seminar audiences at Utah State University and the 1998 annual meeting of the American Agricultural Economics Association for helpful discussions. The usual disclaimer applies.

References

Adesina, A.A., Djato, K.K., 1997a. Farm size, relative efficiency and agrarian policy in Côte d'Ivoire: profit function analysis. Agricultural Economics 14, 93–102.

Adesina, A.A., Djato, K.K., 1997b. Relative efficiency of women as farm managers: profit function analysis in Côte d'Ivoire. Agricultural Economics 16, 47–53.

Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic production function models. Journal of Econometrics 6, 21–37.

Ali, M., Byerlee, D., 1991. Economic efficiency of small farmers in a changing world: a survey of recent evidence. Journal of International Development 3, 1–27.

Barrett, C.B., 1997. How credible are estimates of peasant allocative, scale or scope efficiency? A commentary. Journal of International Development 9, 221–229.

Barrett, C.B., Bezuneh, M., Clay, D.C., Reardon, T., 2001. Heterogeneous Constraints, Incentives and Income Diversification Strategies in Rural Africa. Mimeo.

Battese, G.E., 1992. Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. Agricultural Economics 7, 185–208.
Battese, G.E., Coelli, T.J., 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis 3 (1/2), 153–169.
Battese, G.E., Coelli, T.J., 1993. A Stochastic Frontier Production Function Incorporating a Model for Technical Inefficiency Effects, Working Paper No. 69, Department of Econometrics, University of New England.
Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, 325–332.
Bravo-Ureta, B.E., Pinheiro, A.E., 1993. Efficiency analysis of developing country agriculture: a review of the frontier function literature. Agricultural and Resource Economics Review 22, 88–101.
Coelli, T.J., 1995. Recent developments in frontier modelling and efficiency measurement. Australian Journal of Agricultural Economics 39, 219–245.
Conway, G., 1997. The Doubly Green Revolution: Food For All in the 21st Century. Cornell University Press, Ithaca, NY, USA.
Cornwell, C., Schmidt, P., Sickles, R.C., 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46 (1/2), 185–200.
Dalton, T.J., 1999. Technological Change and Virtual Pricing in Upland Rice Improvement, Manuscript, West Africa Rice Development Association.
Färe, R., Grosskopf, S., Lovell, C.A.K., 1994. Production frontiers. Cambridge University Press, Cambridge, UK, New York, USA.
Farrell, M.J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A. General 120 (Part 3), 253–281.
Greene, W.H., 1997. Frontier production functions. In: Pesaran, M.H., Schmidt, P. (Eds.), Handbook of Applied Econometrics, Microeconomics, vol. II. Blackwell, Oxford (UK), Cambridge, MA (USA), pp. 81–166.
Hall, M., Winsten, C., 1959. The ambiguous notion of efficiency. Economic Journal 69, 71–86.
Huang, C.J., Liu, J.-T., 1994. Estimation of a non-neutral stochastic frontier production function. Journal of Productivity Analysis 5, 171–180.
Jondrow, J., Lovell, C.A.K., Materov, I.S., 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., 1990. Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics 46 (1/2), 201–212.
Kumbhakar, S.C., 1991. Estimation of technical inefficiency in panel data models with firm- and time-specific effects. Economics Letters 36, 43–48.
Kumbhakar, S.C., Ghosh, S., McGuckin, J.T., 1991. A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. Journal of Business and Economic Statistics 9, 279–286.
Lee, Y.H., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical inefficiency. In: Fried, H.O., Lovell, C.A.K., Schmidt, S.S. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications. Oxford Univ. Press, New York, NY, USA, pp. 237–255.
Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb–Douglas production functions with composed error. International Economic Review 18, 435–444.
Pitt, M., Lee, L., 1981. The measurement and sources of technical inefficiency in the Indonesian weaving industry. Journal of Development Economics 9, 43–64.
Reifschneider, D., Stevenson, R., 1991. Systematic departures from the frontier: a framework for the analysis of firm inefficiency. International Economic Review 32, 715–723.
Schultz, T.W., 1964. Transforming Traditional Agriculture. Yale Univ. Press, New Haven, CT, USA.
Sherlund, S.M., 1998. Exogenous Shocks and Technical Efficiency Among Ivorien Rice Farmers, MS thesis, Utah State University.
Waldman, D., 1982. A stationary point for the stochastic frontier likelihood. Journal of Econometrics 18, 275–279.
WARDA, 1997. WARDA/ADRAO Farm Management and Household Survey Data Guide West Africa Rice Development Association, Bouaké, Côte d'Ivoire.
Zellner, A., Kmenta, J., Dreze, J., 1966. Specification and estimation of Cobb–Douglas production functions. Econometrica 34, 784–795.