Efficiency of rice farming households in Vietnam

Vu Hoang Linh
Department of Development Economics, University of Economics and Business, Hanoi, Vietnam and
Indochina Research and Consulting, Hanoi, Vietnam

Abstract

Purpose – The purpose of this paper is to estimate technical efficiency obtained from both data envelopment analysis (DEA) and stochastic frontier approaches using household survey data for rice farming households in Vietnam.

Design/methodology/approach – A bootstrap method is used to provide statistical precision of the DEA estimator. Technical efficiency is modeled as a function of household and production factors.

Findings – The results from the deterministic, semi-parametric and parametric approaches indicate that among other things, technical efficiency is significantly influenced by primary education and regional factors. In addition, scale efficiency analysis shows that many farms in Vietnam are operating with less than optimal scale of operation.

Originality/value – The study is among the first that employ a bootstrap method and compare estimates from both Data Envelopment Analysis (DEA) and stochastic frontier approaches.

Keywords Vietnam, Farms, Rice, Data envelopment analysis, Stochastic frontier, Efficiency, Bootstrap

Paper type Research paper

1. Introduction

Agriculture in Vietnam is the most important sector as it contributes about 21.8 percent to gross domestic product (World Bank, 2006) and supports jobs for 67.3 percent of the population (IRRI, 2005). In agriculture, rice is the most important crop in Vietnam. It is planted on 84 percent of cultivated area and constitutes more than 85 percent of Vietnam's total grain output. It also provides about 85 percent of the total daily calorie intake for the Vietnamese people (Nghiem and Coelli, 2002).

Since the reforming the *Doi Moi* policy launched in December, 1986, the government has liberalized the rice market as well as the markets for agricultural inputs. The government has also promoted the cultivation of high-yielding rice varieties. As a result, rice production and exports have increased steadily. Rice production increased from 15.1 million tons in 1987 to 32.6 million tons in 2000, a growth of 6.1 percent per year, while rice yields increased from 2.70 tons/ha in 1987 to 4.25 tons/ha in 2000, a growth of 3.3 percent per year (IRRI, 2006). Since the launch of the *Doi Moi* policy, rice production, rice area and rice yield have increased significantly although recently, the growth of rice area has slowed down and even become slightly negative.

Vietnam has been a major rice exporter since 1989, and is currently the second largest rice exporter, exporting 5.2 million tons in 2005 which is equivalent to 18.2 percent of total world rice trade (FAO, 2006). Recently, modern rice technology

The author would like to thank Kent Olson, Paul Glewwe, Philip Pardey and Terrance Hurley for their helpful comments; however all errors and views expressed are the author's.

International Journal of Development Issues
Vol. 11 No. 1, 2012
pp. 60-73

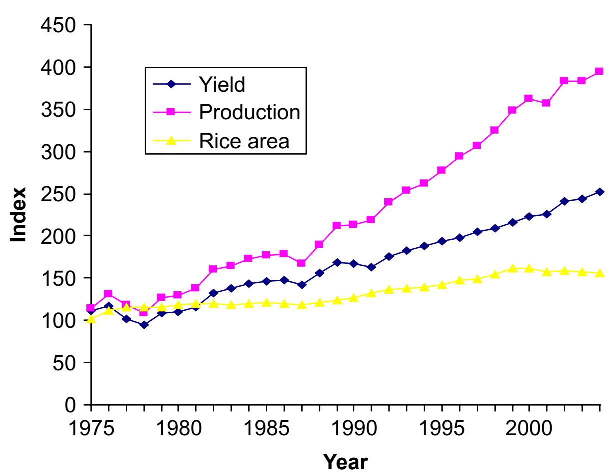
1446-8956
DOI 10.1108/14468951211213868

has been widely applied. The adoption rate of fertilizer-responsive, high-yielding modern rice varieties increased from 17 percent in 1980 to nearly 90 percent in 2000 (Tran and Kajisa, 2006) (Figure 1).

Despite the importance of rice production in Vietnam as well as in the world market, there have been very few studies on the efficiency of Vietnamese rice farms. This paper is the first attempt to estimate farm-level technical and scale efficiency (SE), and determine the factors influencing technical efficiency (TE) for rice production in Vietnam. This paper would be useful for those interested in Vietnam's rice production as it is one of the first studies on efficiency of rice farming in Vietnam. It is also a contribution to the empirical work on efficiency, notably the application of a bootstrap procedure to establish the statistical properties of data envelopment analysis (DEA) TE.

Efficiency can be estimated by either parametric or nonparametric methods. Parametric measurement includes specifying and estimating a stochastic production frontier or stochastic cost frontier. In this method, the output (or cost) is assumed to be function of inputs, inefficiency and random error. The main strength of the stochastic frontier function approach (SFA) is its incorporation of stochastic error, and therefore permitting hypothesis testing. The disadvantage of this approach is the imposition of an explicit functional form and distribution assumption on the error term. On the other hand, the non-parametric approach or the DEA has the advantage of no prior parametric restrictions on the technology, hence less sensitive to model misspecification. However, because DEA is a deterministic approach, all deviations from the frontier are considered as inefficiencies, making it sensitive to measurement errors and data noises.

There have been many studies on efficiency in agriculture in developing countries, most of which apply SFA. Thiam *et al.* (2001) summarize 51 observations of TE in developing countries from 32 studies published before 1999. In Vietnam, there are only a few papers that calculate efficiency and determine the factors affecting efficiency of Vietnam's agriculture. Past studies on efficiency of rice production in Vietnam only use simple partial measures of productivity such as yield per hectare. To our knowledge, Kompas (2004) is the only attempt to calculate average TE for rice sector in Vietnam,



Source: Author calculated from IRRI (2006)

Figure 1. Rice production, yield and area in Vietnam 1975-2004 (1975 = 100)

using a stochastic production frontier based on a region-level panel data. In his study, average TE for the whole country is 0.65 in 1999 and 0.78 for the principal rice areas (Red River delta and Mekong River delta). However, Kompas (2004) uses aggregate regional data, which may not give useful information on the efficiency at farm level.

Given the advantages and disadvantages of both the DEA and SFA methods, it may be helpful to use both methods and compare them using the same data set. In addition, establishing the statistical properties of the DEA estimator is useful for overcoming the disadvantage of the nonparametric method and improving the results' robustness. Recent advances in the DEA literature include using bootstrap methods to establish the confidence interval of TE (Simar and Wilson, 2000). The bootstrap method in Simar and Wilson (2000) has been applied empirically in several studies of farm efficiency in developed countries (Brümmer, 2001; Latruffe *et al.*, 2005; Ortner *et al.*, 2006; Olson and Vu, 2007).

The objectives of this paper are twofold. First, it uses both the bootstrapped DEA method to estimate technical and SE of rice farming households in Vietnam. Second, it uses estimates from both DEA and SFA approaches in the second stage to determine the factors influencing these estimates. This paper contributes to the efficiency literature by using weighted Tobit regressions to estimate the effects of factors on farm TE. While most of the studies on efficiency are limited to point estimates, this paper adds to the few papers (Brümmer, 2001; Fraser *et al.*, 2006) that cover both point estimates and confidence intervals by DEA and SFA methods. It is also the first paper studying rice farming efficiency in Vietnam using household data. The results would be of interests to the researchers as well as to policy-makers as they provide information on the causes and disparities of farm efficiency in Vietnam.

2. Efficiency measurement

Following the seminal work by Farrell (1957) and others, economic efficiency is typically decomposed into three types: technical, allocative and SE. TE measures the firm's ability to use the available technology in the most effective way. Allocative efficiency (AE) is dependent on prices and measures the firm's ability to make optimal decisions on product mix and resource allocation. Combining measures of technical and AE yields a measure of economic efficiency. SE measures the optimality of the firm's size.

2.1 Data envelopment analysis

As a nonparametric approach, DEA (Charnes *et al.*, 1978; Färe *et al.*, 1994) is used to derive technical and SE. DEA method can be applied using either output-based or input-based approaches depending on whether they use an input distance function or an output distance function. In this paper, we use the DEA method to estimate an input-based technical and SE as well as output-based TE. Estimates were made using linear programming in the software GAMS/OSL. The input-based TE under variable returns to scale (VRS) is the focus of our study. Based on a smoothed bootstrap procedure for DEA estimators proposed by Simar and Wilson (1998, 2000), the paper estimates the bias and the confidence interval of the input-based TE with VRS, using the package FEAR developed by Wilson (2005).

Technical and scale efficiency. For the jth farm out of n farms, the input-based TE under constant return to scale (CRS) is obtained by solving the following problem:

$$TE_j = \underset{\theta_j^{CRS},\,\lambda}{Min\ \theta_j^{CRS}} \quad \text{subject to} \;\; Y_j \le Y\lambda;\; \theta_j^{CRS} X_j \ge X\lambda; \lambda \ge 0 \tag{1}$$

where X and Y are the input and output vector, respectively, θ_j^{CRS} is TE of farm j under CRS and λ is an $n \times 1$ vector of weights. In general, $0 \le \theta_j^{CRS} \le 1$, where $\theta_j^{CRS} = 1$ if the farm is producing on the production frontier and hence, technically efficient. When $\theta_j^{CRS} < 1$, the farm is technically inefficient. In the case of VRS, one can find TE θ_j^{VRS} under VRS by adding the convexity constraint $\sum_{j=1}^{n} \lambda_j = 1$ to equation (1). Because the VRS is more flexible so the convex hull envelops the data more tightly than under CRS, θ_j^{VRS} is always equal or greater than θ_j^{CRS}.

SE is measured by the formula:

$$SE_j = \frac{\theta_j^{CRS}}{\theta_j^{VRS}} \tag{2}$$

In general, $0 \le \text{SE} \le 1$, with $\text{SE} = 1$ representing efficient economy of scale. $\text{SE} < 1$ implies that the inputs are not scale efficient, which can be either increasing returns to scale (IRS) or decreasing returns to scale (DRS).

2.2 Bootstrapping the DEA estimator

While DEA methods have been widely applied, most researchers have largely ignored the statistical properties in the estimators. Ignoring the statistical noise in the estimation can lead to biased DEA estimates and misleading result because all the deviations from the frontier are considered as inefficiency. Simar and Wilson (1998, 2000) argue that bootstrap is the most currently feasible method to establish the statistical property for DEA estimators. This paper applies the Simar and Wilson (1998, 2000) smoothed bootstrap procedure to correct the bias in DEA estimators and establish their confidence interval. The procedure for this paper is described in more details in the Appendix.

2.3 Stochastic frontier method

The production function under VRS is specified as (Aigner *et al.*, 1977; Battese and Coelli, 1992):

$$\text{In} Y_j = f(X_i; \beta) + \varepsilon_i \tag{3}$$

with X_i as the input and Y_i as the output vector for farm *i*; *f*(X_i;*β*) is normally assumed either Cobb-Douglass production technology or translog technology. Both functional forms are used extensively in literature. For example, in Thiam *et al.* (2001)'s meta-analysis, of 33 studies applying stochastic frontier methods, 19 used the Cobb-Douglass functional form while 14 used a translog functional form. In this paper, we choose the Cobb-Douglass functional form for convenience because we have a relatively large number of inputs in the production frontier function. Furthermore, the Cobb-Douglass functional form is also more convenient in testing the return to scale hypothesis. The drawback of using Cobb-Douglas functional form lies in its relative restrictiveness of coefficients. Yet, our tests using both functional forms showed quite similar results.

The Cobb-Douglass production function under VRS is:

$$\text{In}\, Y_i = \beta_0 + \sum_{k=1}^{T} \beta_k \,\text{In}\, X_k + \varepsilon_i$$

The error term in equation (3) is composed of two components (Aigner *et al.*, 1977):

$$\varepsilon_i = v_i - u_i$$

where v_is are assumed to be independently and identically $\mathrm{N}(0, \sigma_v^2)$ representing the random errors. The term u_i represents technical inefficiency of farm i but unlike v_i, it is only a one-sided variable taking non-negative values. In this paper, we assume u_i to be half-normal distribution, stated by Greene (1997) as "the most useful formulation". In other words, $u_i = |U|$ where $U \sim N(0, \sigma_u^2)$. The TE of farm i is $TE_i = exp(-u_i)$, which is greater than zero and less than 1. The estimation of stochastic frontier model is done by maximum likelihood methods in STATA version 9.0 software. The confidence intervals of TE in this paper are established following Horrace and Schmidt (1996).

3. Data

The data is taken from Vietnam Household Living Standard Survey 2003-2004 (VHLSS, 2004). The survey is implemented by the General Statistics Office of Vietnam with technical support from World Bank. In the VHLSS 2004 survey, there are 8,813 households living in both rural and urban areas surveyed, including about 4,300 households producing rice. From that sample, we chose randomly a sub-sample of 600 households. After calculating the efficiency, we dropped five extreme observations to reduce the possibility of DEA's sensitivity to outliers. Efficiency scores are recalculated using the final sample of 595 farm households.

The measure for output is the harvested rice quantity during the last year. We chose rice quantity as output instead of rice value because we wanted to exclude the price effects from calculating TE. The inputs include nine categories: fertilizers, pesticides, seed, family labor, hired labor, owned fixed asset and equipment value, asset hire (including cattle hire) and maintenance, small tool and energy, and other farming expenditure and rice land. Since, beside rice growing, households are also engaged in other activities, family labor is measured by the total family hours allocated in farming adjusted by the percentage of rice production over total farm production. Rice land is measured by the land area allocated for rice production. Other inputs are measured by the expenditures in current money value. In our sample, on average, rice occupies for 46 percent of agricultural household outputs. This number is close to the macro percentage of 41.5 percent in 2001, which is the percentage of rice production value in total agricultural production value for the whole country. Summary statistics for these households are listed in Table I.

4. Empirical results

4.1 Technical efficiency

The estimated DEA and SFA efficiencies are presented in Table II. The average TE estimated by DEA method is higher than that estimated by SFA method. Similar results have been reached in Kalaitzandonakes and Dunn (1995) for corn farms in Guatemala and Wadud and White (2000) for rice farmers in Bangladesh. In our estimation,

Variable	Mean	SD	Min.	Max.
Input and output vectors				
Rice quantity (kgs)	7,560	11,125	100	100,640
Rice value[a]	6,562	8,428	200	100,048
Seed expenditures[a]	291	530	0	9,900
Fertilizer expenditures[a]	976	1,353	0	13,800
Pesticide expenditures[a]	308	706	0	6,540
Family hours for farming[a]	2,184	1,766	64	9,432
Percent of rice (percent)	46	25	0.7	100
Estimated family hours for rice production (hours)	904	871	7.5	5,333
Rice land area (square meters)	6,991	8,770	250	74,000
Fixed asset and equipment value[a]	6,414	12,976	0	164,500
Hired-in labor expenditure[a]	262	674	0	8,750
Asset hire and maintenance[a]	529	964	0	6,540
Small tool and energy[a]	98	255	0	8,750
Other expenditure[a]	242	419	0	8,312

Note: [a]In thousand VND at current value

Table I.
Summary statistics for rice farming farms

	TE_{CRS}	$TE_{VRS\text{-}OUT}$	$TE_{VRS\text{-}IN}$	Bias-corrected TE	Lower bound	Higher bound	TE by SFA	Lower bound	Higher bound
Average	0.704	0.765	0.785	0.678	0.593	0.771	0.634	0.449	0.825
Median	0.711	0.816	0.824	0.741	0.627	0.811	0.674	0.462	0.927
SD	0.244	0.238	0.212	0.167	0.137	0.208	0.193	0.152	0.208
Min	0.09	0.174	0.228	0.205	0.19	0.224	0.109	0.074	0.155
Max	1	1	1	0.896	0.844	0.986	0.952	0.839	0.999

Table II.
DEA and SFA estimates

the input-based TE is 0.785, slightly higher than the output-based TE of 0.765. It means that with a given bundle of inputs, an average household can increase its output by 30.7 percent ($=1/TE_{VRS\text{-}OUT} - 1$). On the other hand, that household can reduce its inputs by 27.4 percent ($=1/TE_{VRS\text{-}IN} - 1$) without changing the level of its output.

Estimates from the deterministic DEA model have downward biases in efficiency scores because in the model, the "true" production frontier is unknown, and the points on the observed production frontier may be inefficient in the presence of the "true" production frontier. Using bootstrap method as in Simar and Wilson (2000), we estimate bias-corrected TE scores and find them significantly lower than the initial TE scores.

Figure 2 shows the distribution of initial DEA estimates, bias-corrected DEA estimates and the 95-percent confidence interval for the input-based methods. If we only know the initial DEA estimates, it appears that on averages, rice farms in Vietnam can reduce their inputs by 27.4 percent and still can produce the same outputs. Yet, after correcting for the bias, the amount of input saving is 47.5 percent ($=1/0.678 - 1$). In the same way, an average farm can reduce their inputs in the range from 29.7 to 68.6 percent with 95 percent confidence interval. By stochastic frontier method, the corresponding value is 57.8 percent ($=1/0.634 -$) for Cobb-Douglass specification. It is clear that the amount of input saving is considerable.

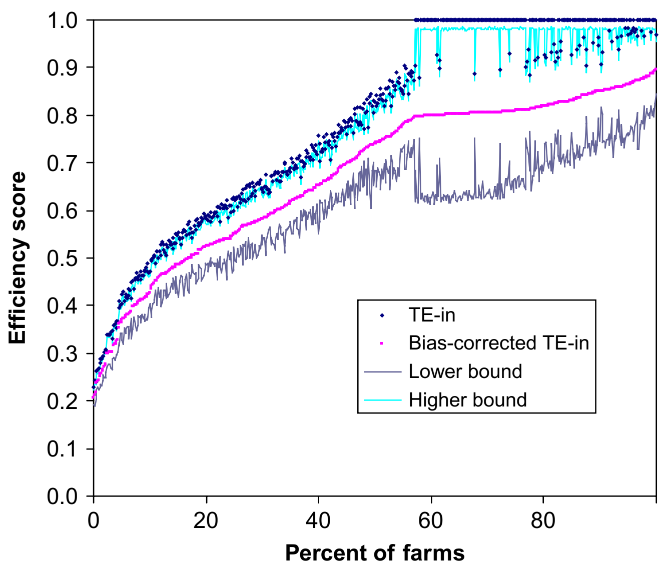


Figure 2.
Initial and bias-corrected input-based technical efficiency under VRS

To compare the estimates from nonparametric and parametric approaches, we use the paired *t*-tests and Spearman rank correlation. The results are presented in Table III. Based on paired *t*-test, on average, the TE scores in nonparametric, both before and after correcting for bias, are higher than in parametric method although the difference is smaller for bias-corrected estimates. The Spearman correlation coefficients between the efficiency rankings of the sample farms are positive and significant, implying that the efficiency scores calculated in both methods are not independent. In other words, the efficiency rankings of farms in Vietnam are consistent in both methods.

Table IV shows the distribution of technically efficient farm in the dataset according to DEA method. Farms in the southern region-the main production region in Vietnam – are most efficient. Farms in the central region are least technically efficient. In addition, average TE and percentage of technical efficient farms are higher for large farms than for small farms and the same for diversified farms than for mainly rice farms. Large farms are defined as farms with total farm output value higher than 15 million VND (about $1,000). Mainly rice farms are farms with rice output equivalent more than 70 percent of total farm output value. About 70 percent of farms in our sample are mainly rice farms and 37 percent of farms are large farms.

Table III.
Paired *t*-tests and spearman rank correlation tests

Efficiency	Sample mean DEA	SFA	*t*-ratio	Spearman rank correlation
Initial TE	0.785	0.634	19.16*	0.5284*
Bias-corrected TE	0.678	0.634	6.50*	0.5526*

Note: Significant at: *1 percent level

Region	Average TE	Bias-corrected TE	Number of farms with TE = 1	Percentage of farms with TE = 1
All farms	0.785	0.678	201	33.8
Red River delta	0.801	0.698	49	28.3
North east	0.786	0.678	37	34.9
North west	0.806	0.688	23	42.6
North central coast	0.704	0.619	14	18.9
South central coast	0.715	0.622	15	27.8
Central highlands	0.867	0.723	15	57.7
South east	0.785	0.652	14	53.8
Mekong River delta	0.831	0.710	34	41.5
North	0.797	0.690	109	32.7
Center	0.709	0.621	29	22.7
South	0.829	0.701	63	47.0
Large farm	0.812	0.697	81	36.7
Small farm	0.770	0.667	120	32.1
Diversified farm	0.816	0.701	70	39.8
Mainly rice farm	0.772	0.668	131	32.0

Table IV.
Distribution of average technical efficiency

4.2 Scale efficiency

Farm household SE scores are presented in Table V. The farm households in the south are more scale efficient than farms in the north, and the center and large farms are more scale efficient than small farms. However, mainly rice farms are more scale efficient than diversified farms. About 23.4 percent of total farms are working with optimal scale

	SE	Number of farms with SE = 1	DRS	IRS	Percentage with SE = 1	Total farm output (mil. VND)
All farms	0.890	139	104	352	23.4	17.4
Red River delta	0.900	35	22	116	20.2	14.6
North east	0.893	26	15	65	24.5	13.4
North west	0.881	17	3	34	31.5	10.9
North central coast	0.881	3	14	57	4.1	11.4
South central coast	0.823	7	9	38	13.0	12.3
Central highlands	0.879	14	3	9	53.8	34.7
South east	0.907	9	5	12	34.6	32.8
Mekong River delta	0.923	28	33	21	34.1	31.1
North	0.895	78	40	215	23.4	13.6
Center	0.857	10	23	95	7.8	11.8
South	0.911	51	41	42	38.1	32.1
Large farm	0.924	62	68	91	28.1	33.1
Small farm	0.871	77	36	261	20.6	8.1
Diversified farm	0.831	37	14	125	21.0	26.1
Mainly rice farm	0.915	102	90	227	24.3	13.7

Table V.
Distribution of average scale efficiency

operation and a majority of farms (59 percent) are operating with IRS. This suggests that a large number of Vietnamese rice farms should increase their scale of operations to gain SE.

For the stochastic functional form, the sum of coefficients from the Cobb-Douglass production frontier is 1.098 implying IRS. We reject the hypothesis of CRS (sum of coefficient equal to one) at one-percent level of significance.

4.3 Factors associated with efficiency

A relevant question is what factors can influence the farm TE. The factors included for close examination in this study include household characteristics, production structure, land characteristics and regional variables.

Household characteristics variables include household size (i.e. total number of household members), adult ratio in the household, household head's age and household head's schooling. Household head's schooling is divided into four categories: no formal education, with primary schooling (from one to five years), with secondary schooling (from six to nine years) and with high schooling or higher (ten years and up). In our data, 32 percent of household heads have primary schooling, 45 percent have some secondary schooling, 14 percent have more than nine years of schooling and only 7 percent never went to school.

Other variables that might affect farm TE include farm size (representing by total farm output value), capital to labor ratio (million VND/hour), land to labor ratio (square meter/hour), non-farm income ratio and number of extension visits. Total farm output value includes both rice and other crop/livestock income. Capital is measured as total fixed asset value.

Binary variables include dummies for land characteristics (rented land, high quality land), education level (primary, secondary, high school), borrow money, modern irrigation, and regional binary variables which are grouped into two sets – one set include dummies for center and south region with north being the reference region.

Most of the literature on measuring the effects of factors affecting efficiency use Tobit analysis for DEA estimates. This model is employed in most of papers using the DEA method to estimate the factors associated with TE. However, the standard Tobit model has a disadvantage because it does not account for the bias and confidence interval in the DEA initial scores. We develop a weighted Tobit model with the information obtained from the bootstrap procedure to overcome this limitation. The dependent variable in this model is the initial TE calculated by DEA but with the weights equal to the reciprocals of the width between higher bounds and lower bounds for the bias-corrected TE. The idea is that, the higher the width is, the larger the measurement error that could occur. Therefore, weighted Tobit analysis reduces estimation error by punishing the observations with larger width or higher possibility of measurement error. Finally, Model 5 is the maximum likelihood estimation for stochastic frontier TEs.

The result in Table VI shows that farmer's age has a negative effect to TE although the effect is only significant for Model 1 and Model 2. This is consistent with the findings of Coelli and Battese (1996), Seyoum *et al.* (1998), and Dhungana *et al.* (2004).

Primary education of the household heads is positively related to the farmer TE in all models but the impact is more significant for the stochastic frontier estimates. The impacts of secondary and higher education to TE are more ambiguous.

Dependent variable	Standard tobit Model 1 TE by DEA	Weighted tobit Model 2 TE by DEA	Stochastic frontier Model 5 TE by SPF
Number of obs.	595	595	595
LR $\chi^2(18) =$	101.1	92.5	62.3
Prob > χ^2	0	0	0
Log likelihood	−211	−204.2	251.9
Adult ratio	0.011 (0.14)	−0.02 (−0.27)	0.02 (0.44)
Household size	−0.007 (−0.85)	−0.008 (−0.96)	0.005 (1.06)
Capital/Labor	0.050 (0.07)	−0.149 (−0.15)	−0.282 (−0.68)
Land/Labor	0.021 (5.83)**	0.025 (5.66)**	0.001 (0.4)
Head's age	−0.003 (−2.75)**	−0.002 (−1.81)*	−0.001 (−0.38)
Primary	0.079 (1.64)*	0.082 (1.66)*	0.061 (2.1)**
Secondary	−0.031 (−0.63)	−0.002 (−0.05)	0.073 (2.49)**
High education	−0.038 (−0.68)	−0.041 (−0.73)	0.06 (1.76)*
Farm output	2.134 (2.67)**	2.085 (2.13)**	0.591 (1.4)
Land quality	0.013 (0.51)	0.037 (1.51)	0.083 (5.47)**
Non-farm ratio	0.002 (0.03)	−0.009 (−0.18)	−0.044 (−1.45)

Notes: Significant at: *10 percent and **5 percent; *t*-statistics in parentheses

Table VI. Factors influencing technical efficiency

While secondary and higher education are associated with higher TE indices as calculated by stochastic frontier, they are insignificant for those calculated by the standard and weighted Tobit. This might indicate a more consistent role of primary education rather than secondary or higher education for improving farmers' efficiency.

One justification for the possible limited effects of higher education to TE is that the farmers with higher education tend to shift to non-farm activities, and therefore their education does not contribute to improving farm TE. A simple OLS regression indicates that the non-farm ratio is positively associated with the household head's year of schooling at 5 percent significant level.

To test the hypothesis that household decisions are collective and influenced by the household member with highest education level rather than the household head's education, we also use the maximal education level of the households as a regressor instead of head's education level. We do not find any significant relationship between the household's highest education level and its TE. The finding suggests that the head's education may be a more important factor in deciding the household TE.

The land/labor ratio has a significant positive impact on TE for both DEA models but not for the SFA model. This means that increasing rice land is generally associated with better TE. Given the shortage and fragmentation of land in a populated economy as in Vietnam, this finding is reasonable. Based on World Rice Statistics of International Rice Research Institute (IRRI, 2005), we estimated that nearly 90 percent of farms in Vietnam have farm area less than 1 ha in 1994 while the corresponding ratio for Philippines in 1991, Pakistan in 1990 and Thailand in 1988 are 37, 36 and 11 percent, respectively. On the other hand, the capital/labor ratio effect on TE is insignificant in all models.

Farm size has a significant positive effect on TE in DEA models but not in SFA model, where the effect is positive but insignificant. It indicates that farm operations in Vietnam are in general not optimal for TE. Modern irrigation also has positive effect but the effect is only strongly significant for the stochastic frontier model.

Among the binary variables, land quality effect is positive in all models but only significant for the SFA model. Farms with loans seem have lower TE scores than farms without loans although the effect is only significant for DEA models. This finding is as expected since farms with loans may be more constrained with the debt burden than those without loans.

Regional dummies show that both the center and the south dummies are negative, indicating that other thing being equal, a farm in the north is more technically effective than in the southern or in the central region. The impact of center dummy is strongly significant at 1 percent level in all models while south dummy is only insignificant in the standard Tobit model. Yet, in Table IV, we see that the average TE score is higher in the south than in the north. This higher efficiency scores can be explained by the influences of other factors, such as farm size: an average farm in the south is almost 2.5 times as large as an average farm in the north (Table V).

Other factors such as household size, household adult ratio, extension visits and rented land ratio are insignificant in explaining TE in all models.

5. Summary and conclusion

This paper analyses TE for a sample of rice producers in Vietnam using the parametric, non-parametric and semi-parametric frontier approaches and then compares the efficiency estimates obtained from these approaches and discusses the effects influencing TE estimates.

The mean TE is 0.704 under CRS, 0.765 under VRS for output-oriented DEA and 0.785 under VRS for input-oriented DEA. A bootstrap procedure correcting for the bias, yields a mean estimate of 0.678 for input-oriented DEA. A confidence interval is also established for the bias-corrected estimates. Stochastic frontier estimation yields a mean estimate of 0.634. The variances of estimates from DEA and SFA methods are similar but the variances of bias-corrected TEs after bootstrapping are significantly lower than the parametric approach, which is a further advantage of the bootstrap method for DEA over the parametric approach. The Spearman correlation test confirms that our efficiency scores calculated from different approaches are positively and significantly correlated. Thus, it implies a consistency of both methods. The results reveal substantial production inefficiency for sample rice farmers in Vietnam and hence significant potential for farmers to reduce their costs by increasing efficiency. On average, a farm can reduce its cost by 30-69 percent depending upon the method employed. A further 12 percent cost reduction can be obtained by operating with optimal scale. A majority of farms, particularly in the central region, are operating with IRS. Given the importance of rice production for income, food security, employment and export in Vietnam, the benefits from increasing farm efficiency are very substantial.

Results from stochastic, non-parametric as well as new semi-parametric approaches suggest that TE in production is influenced by education, especially primary education. The impacts of secondary and higher education are less robust to model specification. Secondary schooling is highly positive for stochastic model but not for the other models. The analysis also indicates that increasing land holding and farm size has substantial benefits for efficiency improvement. Besides, regional factors are important in influencing TE. The Red River delta, which is very densely populated and has very small landholdings, highly lowland irrigated and highly labor intensive rice cultivation methods, is most technically efficient. The Mekong River delta, which produces more than a half of the country's rice production, has more potential for improving TE. The land in this region is one of the best rice growing regions of the world and there is still capability for increasing rice area. While almost all arable land is under intensive cultivation in the north, only 67 percent of the arable land is under-cultivation in the Mekong Delta. On the other hand, factors such as non-farm ratio or extension support do not significantly affect farm household TE. For extension support, the reason may be due to limited access of farmers to extension service. Policies leading to improvement of farm education, land quality and land holding will be beneficial for improving farmers' TE. The distribution of TE and SE across regions also provides useful information for policy makers in enhancing the farm efficiency for each region.

References

Aigner, D.J., Lovell, C.A.K. and Schmidt, P. (1977), "Formulation and estimation of stochastic frontier production function models", *Journal of Econometrics*, Vol. 6 No. 1, pp. 21-37.

Battese, G.E. and Coelli, T. (1992), "Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India", *The Journal of Productivity Analysis*, Vol. 3, pp. 153-69.

Brümmer, B. (2001), "Estimating confidence intervals for technical efficiency: the case of private farms in Slovenia", *European Review of Agricultural Economics*, Vol. 28 No. 3, pp. 285-306.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research*, Vol. 2, pp. 429-44.

Coelli, T. and Battese, G.E. (1996), "Identification of factors which influence the technical inefficiency of Indian farmers", *Australian Journal of Agricultural Economics*, Vol. 40, pp. 103-28.

Dhungana, B.R., Nuthall, P.L. and Nartea, G.V. (2004), "Measuring the economic inefficiency of Nepalese rice farms using data envelopment analysis", *The Australian Journal of Agricultural and Resource Economics*, Vol. 48 No. 2, pp. 347-69.

FAO (2006), *Foot Outlook – Global Market Analysis No. 1*, Food and Agriculture Organization of the United Nations, Rome, available at: www.fao.org/docrep/009/J7927e/j7927e01.htm (accessed August 15).

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

Farrell, M.J. (1957), "The measurement of productive efficiency", *Journal of the Royal Statistical Society*, Vol. 120, pp. 253-90 (Series A).

Fraser, I., Balcombe, K. and Kim, P. (2006), "Estimating technical efficiency of Australian dairy farms using alternative frontier methodologies?", *Applied Economics*, Vol. 38, pp. 2221-36.

Greene, W.H. (1997), "Frontier production functions", in Pesaran, M.H. and Schmidt, P. (Eds), *Handbook of Applied Econometrics*, Blackwell, Oxford.

Horrace, W.C. and Schmidt, P. (1996), "Confidence statement for efficiency estimates from stochastic frontier models", *Journal of Productivity Analysis*, Vol. 7, pp. 257-82.

IRRI (2005), *World Rice Statistic*, International Rice Research Institute, Manila, available at: www.irri.org/science/ricestat/ (accessed August 15, 2006).

IRRI (2006), *Vietnam*, International Rice Research Institute, Manila, available at: www.irri.org/science/cnyinfo/vietnam.asp (accessed August 15).

Kalaitzandonakes, N.G. and Dunn, E.G. (1995), "Technical efficiency, managerial ability and farmer education in Guatemala corn production: a latent variable analysis", *Agricultural Resource Economics Review*, p. 24.

Kompas, T. (2004), "Market reform, productivity and efficiency in rice production", International and Development Economics working papers. Asia Pacific School of Economics and Government, Australian National University, Australia.

Latruffe, L., Balcombe, K., Davidova, S. and Zawalinska, K. (2005), "Technical and scale efficiency of crop and livestock farms in Poland: does specialization matter?", *Agricultural Economics*, Vol. 32, pp. 281-96.

Nghiem, H.S. and Coelli, T. (2002), "The effect of incentive reforms upon productivity: evidence from the Vietnamese rice industry", *The Journal of Development Studies*, Vol. 39 No. 1, pp. 74-93.

Olson, K. and Vu, L.H. (2007), *Changes in Economic Efficiency and Factors Explaining Differences Between Minnesota Farm Households*, Department of Applied Economics, University of Minnesota, Washington, DC, mimeo.

Ortner, K.M., Hambrusch, J. and Kirner, L. (2006), *The Efficiency of Dairy Farms in Austria: Do Natural Conditions Matter?*, Federal Institute of Agricultural Economics, Vienna, available at: www.fat.admin.ch/eaae96/abstracts/s88.pdf (accessed August 10, 2006).

Simar, L. and Wilson, P. (1998), "Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models", *Management Science*, Vol. 44 No. 1, pp. 49-61.

Simar, L. and Wilson, P. (2000), "A general methodology for bootstrapping in non-parametric frontier models", *Journal of Applied Statistics*, Vol. 27 No. 6, pp. 779-802.

Thiam, A., Bravo-Ureta, B.E. and Rivas, T.E. (2001), "Technical efficiency in developing country agriculture: a meta-analysis", *Agricultural Economics*, Vol. 25, pp. 235-43.

Tran, T.U. and Kajisa, K. (2006), "The impact of green revolution on rice production in Vietnam", *The Developing Economies*, Vol. XLIV No. 2, pp. 167-89.

Wadud, A. and White, B. (2000), "Farm household efficiency in Bangladesh: a comparison of stochastic frontier and DEA methods", *Applied Economics*, Vol. 32, pp. 1665-73.

Wilson, P.W. (2005), "Frontier efficiency analysis with R. FEAR 0.913 user's guide", available at: www.eco.utexas.edu/faculty/Wilson/Software/FEAR/fear.html (accessed August 3, 2006).

World Bank (2006), *Vietnam – At a Glance*, World Bank, Hanoi, available at: http://devdata.worldbank.org/AAG/vnm_aag.pdf (accessed August 12).

Further reading

Chavas, J.-P., Petrie, R. and Roth, M. (2005), "Farm household production efficiency: evidence from the Gambia", *American Journal of Agricultural Economics*, Vol. 87 No. 1, pp. 160-79.

Paul, C.M., Nehring, R., Banker, D. and Somwaru, A. (2004), "Scale economics and efficiency in USA agriculture: are traditional farms history?", *Journal of Productivity Analysis*, Vol. 22, pp. 185-205.

Sharma, K., Leung, P.S. and Zaleski, H.M. (1999), "Technical, allocative and economic efficiencies in swine production in Hawaii: a comparison of parametric and nonparametric approaches", *Agricultural Economics*, Vol. 20, pp. 23-35.

Appendix. Bootstrapping procedure for technical efficiency (CRS case) as in Simar and Wilson (2000)

(i) Calculate the DEA efficiency scores under CRS for each farm among N farms as in equation (1), denoted as $\hat{\theta}_i$ for the ith farm.

(ii) Let $\beta_1^*, \ldots, \beta_k^*$ be a simple bootstrap sample from $\hat{\theta}_1, \ldots, \hat{\theta}_k$. Generate a random sample of size k for the random generator:

$$\tilde{\theta}_i^* = \begin{cases} \beta_i^* + h\varepsilon_i^* & \textit{if} \ \ \beta_i^* + h\varepsilon_i^* \leq 1 \\ 2 - \beta_i^* - h\varepsilon_i^* & \textit{otherwise} \end{cases}$$

where h is the bandwidth of a standard normal kernel density and ε_i^* is a random deviation from the standard normal.

(iii) To correct the variance of the generated bootstrap sequence when kernel estimators are used, construct another sequence:

$$\theta_i^* = \bar{\beta}^* + \frac{1}{\sqrt{1 + h^2/\hat{\sigma}_\theta^2}} \left(\tilde{\theta}_i^* - \bar{\beta}^* \right) \text{ where } \bar{\beta}^* = (1/\text{n}) \sum_{i=1}^{N} \beta_i^*.$$

Thus, the sequence θ_i^* is obtained by the smoothed bootstrap. It has better properties than the simple bootstrap sequence in the sense that the variance of θ_i^* is asymptotically correct.

(iv) For i = 1, …, N, a pseudo data set of $\left(x_{i,b}^*, y_{i,b}^*\right)$ where $x_{i,b}^* = (\hat{\theta}_i/\theta_i^*)\, x_i$ and $y_{i,b}^* = v_i$ with x_i, y_i the original input and output vectors of the ith farm, respectively.

(v) Calculate the new DEA score $\hat{\theta}_i^*$ for each farm by taking the pseudo data as reference.

(vi) Repeat Steps (i)-(iv) for B times to yield B new DEA TE scores $\hat{\theta}_i^*$ for i = 1, …, N.

(vii) Calculate the bootstrap bias estimate for the original estimator $\hat{\theta}_i$ as:

$$b\frown ias_B(\hat{\theta}_i) = B^{-1} \sum_{b=1}^{B} \hat{\theta}_i^* - \hat{\theta}_i.$$

The bias-corrected estimator of $\hat{\theta}_i$ can be computed as $\hat{\hat{\theta}}_i = \hat{\theta}_i - \widehat{b}ias_B(\hat{\theta}_i)$.

(viii) The percentile method is involved in constructing confidence interval. The confidence interval for the true value of $\hat{\theta}_i$ can be established by finding value a_a, b_a such that Prob $(-b_\alpha \leq \hat{\theta}_i^* - \hat{\theta}_i \leq -a_a) = 1 - \alpha$. Since we do not know the distribution of $(\hat{\theta}_i^* - \hat{\theta}_i)$, we can use the bootstrap values to find $\hat{a}_a, \hat{b}_a$ such that Prob $(-\hat{b}_\alpha \leq \hat{\theta}_i^* - \hat{\theta}_i \leq -\hat{a}_a) = 1 - \alpha$. It involves sorting the value of $(\hat{\theta}_i^* - \hat{\theta}_i)$ for b = 1, …, B in increasing order and deleting $((\alpha/2) \times 100)$ percent of the elements at either end of this sorted array and setting $-\hat{a}_a$ and $-\hat{b}_a$ at the two endpoints, with $\hat{a}_a \leq \hat{b}_a$. In our empirical work, we set B = 2000 to ensure the low variability of the bootstrap confidence intervals. The value of bandwidth of the density estimate h is found by Simar and Wilson (2000)'s method of minimizing an approximation to the mean weighted integrated square error.

Corresponding author
Vu Hoang Linh can be contacted at: vhlinh@vnu.edu.vn