

# Efficiency of Rice Farming Households in Vietnam: A DEA with Bootstrap and Stochastic Frontier Application\*

*Preliminary Draft (comments are mostly welcome)*

**Linh H. Vu**

*Department of Applied Economics, University of Minnesota, #332M, 1994 Buford Ave, St Paul, MN 55108.*

---

## *Abstract*

This study estimates technical efficiency obtained from both Data Envelopment Analysis (DEA) and stochastic frontier approach using household survey data for rice farming households in Vietnam. A bootstrap method is used to provide statistical precision of DEA estimator. Bootstrap methods have not commonly used in empirical analysis despite being an important statistical tool for improving the estimation precision. Technical efficiency is modeled as a function of household and production factors. 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 indicates that many farms in Vietnam are operating with less than optimal scale of operation, especially in the Center region.

*Keywords:* Data Envelopment Analysis (DEA), stochastic frontier, efficiency, rice, bootstrap, Vietnam.

---

\* Corresponding author: vuxx0090@umn.edu.

## I. Introduction

Agriculture in Vietnam is the most important sector as it contributes about 21.8 percent to gross domestic product (World Bank 2006) and support jobs for 67.3 percent of the population<sup>1</sup>. In agriculture, rice is the most important crop in Vietnam. It is planted on about 84 percent of agricultural area and contributes more than 85 percent of food grain output. It also provides about 85 percent of the total daily calorie intake for Vietnamese people (Nghiem and Coelli 2002) and Vietnam.

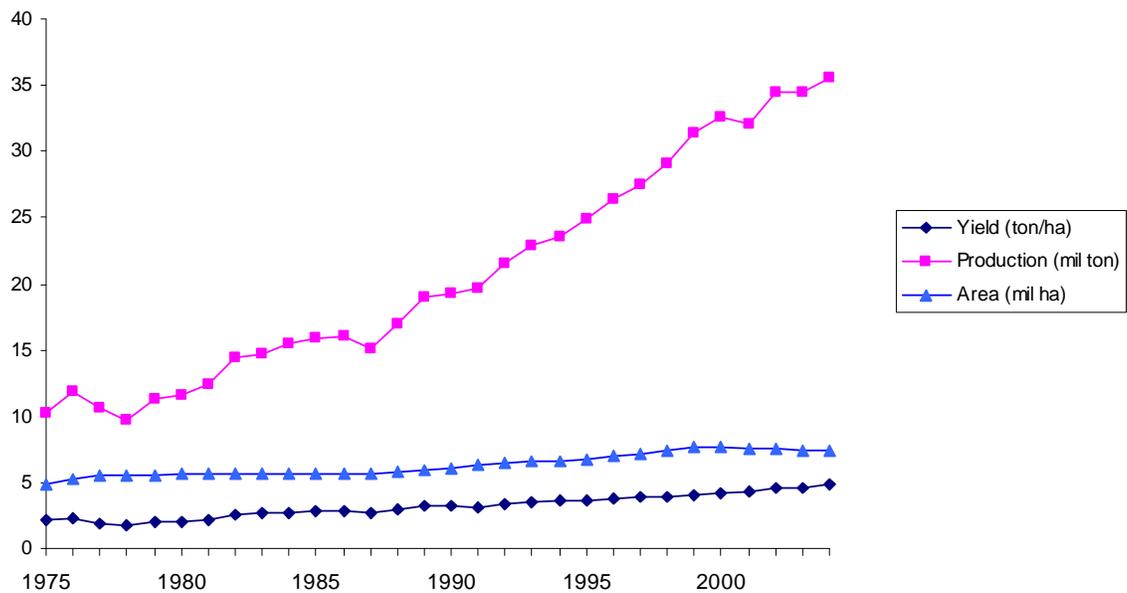
Since the reforming Doi Moi policy launched in December, 1986, the government has liberalized the rice market as well as the markets for agricultural inputs. The government has also promoted high-yielding variety cultivation. Since then, rice production and export has increased steadily. Rice production increased from 15.1 million tons in 1987 to 32.6 million tons in 2000, a growth of 6.1% 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% per year (IRRI 2006). Figure 1 indicates the growth of rice production, rice area and rice yield over the past 20 years. Since the launch of Doi Moi policy, rice production, rice area and rice yield have increased significantly although recently the growth of rice area has been slowed down and even slightly declined.

Vietnam has been a major rice exporter since 1989, currently the second largest rice exporter, exporting as much as 5.2 million tons in 2005 (Government Information Gateway 2006), equivalent to 18.2 percent of total world rice trade (FAO 2006). Recently, modern rice technology has been widely applied. The adoption rate of fertilizer-responsive, high-yielding modern rice varieties increased from 17% in 1980 to nearly 90% in 2000 (Tran and Kajisa 2006).

---

<sup>1</sup> Author calculated based on International Rice Research Institute (IRRI)'s World Rice Statistics data.

**Figure 1: Rice Production, Yield and Area in Vietnam 1975-2004**



Source: Author calculated from IRRI 2006

Given the importance of rice production in Vietnam, it is unclear that there have been very few studies on the efficiency of Vietnamese rice farms. This paper is the first attempt to estimate rice farm-level technical and scale efficiency and factors influencing technical efficiency in Vietnam by applying several novel methods. This paper would be useful for those interested in Vietnam's rice production as well as a contribution to the empirical work on efficiency, notably the application of bootstrap procedure to establish the statistical properties of DEA technical efficiency.

## II. Analytical Framework

Following seminal work by Farrell (1957) and others, economic efficiency is typically decomposed into three types: technical, allocative and scale efficiency<sup>2</sup>. Technical efficiency (TE) measures the firm's ability to use the best practices and 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

<sup>2</sup> Farrell (1957) used the term "price inefficiency" instead of "allocative efficiency"

allocation. Combining measures of technical and allocative efficiency yields a measure of economic efficiency. Scale efficiency measures the optimality of the firm's size.

Efficiency can be estimated by either parametric or nonparametric method. Parametric measurement includes specifying and estimating a stochastic production frontier or stochastic cost frontier. In this method, the output (or cost) is assumed as a function of inputs, inefficiency and random error. The main strength of the stochastic frontier function approach (SFAF) is its incorporation of stochastic error, and therefore permitting hypothetical testing. The disadvantage of this approach is the imposition of an explicit functional form and distribution assumption of the error term. Therefore, stochastic frontier method is sensitive to the parametric form chosen.

On the other hand, the non-parametric approach or the data envelopment analysis (DEA) has the advantage of no prior parametric restrictions on the technology, therefore less sensitive to model mis-specification. DEA method is also not subject to assumptions on the distribution of the error term and imposes minimal assumptions on production behavior. Furthermore, estimation of DEA method is based on a piecewise production frontier, making the estimated frontier close to real activity. 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. Furthermore, DEA is known to be sensitive to outliers.

There have been many studies on efficiency in agriculture in developing countries in which a majority is stochastic frontier studies. Thiam et al (2001) summarizes 51 observations of TE in developing countries from 32 studies published before 1999. They include 27 stochastic frontier, six deterministic frontier and two DEA studies. Rice is the most studied crop (in 17 studies). However, the application of DEA method has gradually increased. Recent application of DEA method on the estimation and explanation of agricultural efficiency in developing countries include Dhungana et al (2004) on Nepal rice farms, Krasachat (2003) on Thailand rice farms, Chavas et al (2005) on Gambia farms, Shafiq and Rehman (2000) on Pakistan farms. There are several studies that use both DEA

and stochastic methods such as Sharma et al (1999), Wadud (2003) and Wadud and White (2000).

In Vietnam, there are only a few papers that calculate efficiency and determine the factors affecting efficiency in Vietnam's agriculture. They include Tran et al (1993) on state rubber farms (using stochastic frontier method) and Rio and Shively (2005) on coffee farms (using DEA method). Past studies on efficiency of rice production in Vietnam only use simple measures such as yield per hectare, which is not a measure of efficiency but productivity. Some studies focus on total factor productivity (TFP) growth. Nghiem and Coelli (2002) use region-level panel data on rice production in Vietnam to investigate TFP growth in Vietnam since reunification in 1975 and find an average annual TFP growth of between 3.3 and 3.5 percent during the period. Kompas (2004) estimates TFP growth in the same period and finds the TFP growth rates are 0.6, 2.74, 4.43 and 4.46 percent during the period 1976-80, 1981-87, 1988-94 and 1995-99 respectively. Kompas (2004) is also the only attempt to calculate average technical efficiency for rice sector in Vietnam, using a stochastic production frontier based on a region-level panel data. In his study, the average technical efficiency for the whole country is 0.65 in 1999 and about 0.78 for the principal rice areas (Red River Delta and Mekong River Delta). Among other factors, farm size, tractor used proportion and a dummy variable for the principal rice areas are found to be positively associated with higher technical efficiency. However, this study uses regional data, and may not give useful information on the efficiency at farm level as well as the determinants affecting farm efficiency. There have been none studies on the efficiency of rice farms that use farm-specific data yet.

Given the advantages and disadvantages of both DEA and SFA method, it maybe helpful to use and compare them on the same data set. In addition, establishing the statistical properties of DEA estimator is useful for overcoming the disadvantage of the nonparametric method and improves the robustness of the results. Recent advances in DEA literature include using bootstrap to establish the confidence interval of technical efficiency (see 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.

Brummer(2001) uses it to establish the confidence interval of technical efficiency for

private farms in Slovenia; Latruffe et al (2003) apply the method for crop and livestock farms in Poland; Ortner et al (2006) for dairy farms in Austria; and Olson and Vu (forthcoming) for farms in Minnesota, USA.

The objectives of this paper are : (1) using DEA methods to estimate technical and scale efficiency of rice farming households in Vietnam (2) using a semi-parametric bootstrapping procedure based on Simar and Wilson (2000) to establish the statistical property of technical efficiency of rice farming households in Vietnam; (3) using stochastic frontier method to estimate technical efficiency and compare with the results from DEA methods; (4) using estimates from both DEA and SFA in the second stage to determine the factors influencing these estimates. To my knowledge, this is the first study that integrates parametric, nonparametric and semi-parametric methods to assess technical efficiency as well as determine the factor influencing it empirically in the literature. This is also the first study of both technical and scale efficiency at farm household level for rice farming in Vietnam.

### **III. Methods to Estimate the Efficiency**

#### ***1. Deterministic Nonparametric Method: Data Envelopment Analysis***

As a nonparametric approach, DEA (Charnes et al 1978, Färe et al 1994) is used to derive technical and scale efficiency. DEA method can be applied using either output-based or input-based approach depending on whether they use input distance function or output distance function. In input-based approach, the idea is to compare the farm's efficiencies with the minimum inputs given a bundle of outputs. In contrast, the output-based measure is to compare with the maximal outputs with given inputs. Unless under constant return to scale, these two measures are not equal (Färe et al 1994).

In this paper, we first used DEA method to estimate input-based technical and scale efficiency as well as the output-based technical efficiency. Estimates were made using linear programming in the software GAMS/OSL (Brooke et al 1998, Olesen and Peterson 1996, Kalvelagen 2004). However, the input-based technical efficiency under VRS is the focus of this study. Based on the 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 technical efficiency with VRS, using the package FEAR developed by Wilson (2005) in the R platform.

### ***Input-based Technical and Scale Efficiency***

For the  $j^{\text{th}}$  farm out of  $n$  farms, the input-based technical efficiency (TE) under constant return to scale (CRS) is obtained by solving the following problem

$$TE_j = \underset{\theta_j^{CRS}, \lambda}{\text{Min}} \theta_j^{CRS} \quad (1)$$

subject to

$$\begin{aligned} Y_j &\leq Y\lambda \\ \theta_j^{CRS} X_j &\geq X\lambda \\ \lambda &\geq 0 \end{aligned}$$

where  $X$  and  $Y$  are the input and output vector respectively,  $\theta_j^{CRS}$  is technical efficiency of farm  $j$  under CRS and  $\lambda$  is an  $n \times 1$  vector of weights. In general,  $0 \leq \theta_j^{CRS} \leq 1$ , 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 variable returns to scale, one can find technical efficiency  $\theta_j^{VRS}$  under variable return to scale (VRS) by adding the convexity constraint  $\sum_{j=1}^n \lambda_j = 1$  to (1). Because the variable return to scale is more flexible so the convex hull envelops the data more tightly than under CRS,  $\theta_j^{VRS}$  is always equal or greater than  $\theta_j^{CRS}$ <sup>3</sup>.

Scale efficiency (SE) is measured by dividing the two technical measures:

---

<sup>3</sup> Some authors call  $\theta_j^{VRS}$  “total” or “overall” technical efficiency and  $\theta_j^{CRS}$  “pure” technical efficiency.

$$SE_j = \frac{\theta_j^{CRS}}{\theta_j^{VRS}} \quad (2)$$

In general,  $0 \leq SE \leq 1$ , with  $SE=1$  representing efficient economy of scale.  $SE < 1$  implies that the inputs are not scale efficient, which can be either increasing returns to scale (IRS) or decreasing returns to scale (DRS). Among farms with scale inefficiency, one can decide which farms with IRS or DRS by running a DEA problem with non-increasing returns to scale (NIRS) imposed. This can be done by adding the constraint  $\sum \lambda_i \leq 1$  into equation (1). Let  $\theta_j^{NIRS}$  be the obtained technical efficiency measures under NIRS. If  $\theta_j^{NIRS} = \theta_j^{VRS}$  and  $SE_j < 1$ , the farm  $j$  is operating with decreasing return to scale. If  $\theta_j^{NIRS} < \theta_j^{VRS}$  the farm  $j$  is under increasing return to scale.

### ***Output-based Technical Efficiency***

For the  $j^{\text{th}}$  farm out of  $n$  farms, the output-based technical efficiency under constant return to scale (VRS) is obtained by solving the following problem

$$\text{Min}_{\theta_j^{VRS}, \lambda} \theta_j^{CRS} \quad (3)$$

subject to

$$Y_j / \theta_j^{VRS} \leq Y\lambda$$

$$X_j \geq X\lambda$$

$$\lambda \geq 0$$

$$\sum_{j=1}^n \lambda_j = 1$$

Under CRS, the output-based and the input-based technical efficiency are essentially the same.

## ***2. Bootstrapping the DEA estimator***

While DEA methods have been widely applied, most researchers largely ignored the statistical properties in the estimators and any deviation from the frontier is attributed to inefficiency. Ignoring the noise in the estimation can lead to biased DEA estimates and misleading result. Recently, some attempts have been made to establish theoretically and empirically the statistical properties of DEA estimators. Banker (1993) and Korostelev et al (1995) prove the consistency of DEA estimators. Kneip et al (1998) provides convergence results for the general multi-output, multi-input case. Simar and Wilson (1998, 2000) argue that bootstrap is the most currently feasible method to establish the statistical property for DEA estimators. While naïve bootstrapping has been used in several papers (Löthgren (1998), Löthgren and Tombour (1997), Ferrier and Hirshberg (1997)), it is criticized by Simar and Wilson (1999) as giving inconsistent results. To get consistent estimation, Simar and Wilson (1998, 2000) proposes a smoothed bootstrapping method. This paper applies Simar and Wilson (1998, 2000) smoothed bootstrap procedure to correct the bias in DEA estimators and establish their confidence interval.

Bootstrapping is based on the idea that if the by re-sampling the data with replacement, we can mimic the data-generating process (DGP) characterizing the true data generation. The smoothed bootstrap procedure of Simon and Wilson relies on the assumption that the distribution of efficient scores is independently distributed. In this method, the efficiency score are estimated from the original data to form pseudo data where the output vector is the same and the input vector is adjusted by the estimated efficiency scores. From the pseudo data, a new DEA efficiency score is calculated for each farm, taking the pseudo data as a reference. By replicating the re-sampling process, say B times, we can establish the empirical distribution of the efficiency measures from the B efficiency scores. The procedure is described in more details in the Appendix.

### **3. Stochastic Frontier Method**

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

$$\ln Y_i = f(X_i; \beta) + \varepsilon_i \quad (4)$$

with  $X_i$  is the input and  $Y_i$  the output vector for farm  $i$ ;  $f(X_i; \beta)$  is normally assumed either Cobb-Douglass production technology or translog technology. Both functional forms are used extensively in literature<sup>4</sup>. In this paper, we choose Cobb-Douglass functional form. We choose the Cobb-Douglass functional form for convenience because we have a relatively large number of inputs in the production frontier function- ten inputs. Furthermore, the Cobb-Douglass functional form is also more convenient in testing the return to scale hypothesis. Yet, we also calculated technical efficiency by the translog functional form. Using paired t-test, we can not reject the hypothesis that the efficiency scores estimated from both specifications are similar at 1% level, although we can reject that hypothesis at 5% level. Nevertheless, in estimating the determinants of technical efficiency, we find similar results for the TE scores estimated from both functional forms

The Cobb-Douglass production function under VRS is:

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

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

$$\varepsilon_i = v_i - u_i$$

where  $v_i$ s are assumed to be independently and identically  $N(0, \sigma_v^2)$  representing the random errors. The term  $u_i$  represents technical inefficiency of farm  $i$  but it unlike  $v_i$ , it is only a one-sided variable taking non-negative values. There is no consensus about the distribution of  $u_i$  with its distribution assumed half-normal, exponential and truncated normal-distribution and gamma distribution in previous studies. In this paper, we assume  $u_i$  to be half-normal distribution, mentioned by Greene (1997) as “the most useful formulation”. In other words,

$$u_i = |U|, \text{ where } U \sim N(0, \sigma_u^2).$$

---

<sup>4</sup> In Thiam et al (2001)’s meta-analysis, among 33 studies used stochastic frontier methods, 19 used the Cobb-Douglass functional form against 14 that used a translog functional form.

And the technical efficiency of farm  $i$  is  $TE_i = \exp(-u_i | \varepsilon_i)$ , which is greater than zero and lesser than 1. The estimation of stochastic frontier model is done by maximum likelihood method in STATA version 9.0 software.

#### **IV. Data**

The data is taken from Vietnam Household Living Standard Survey 2003-2004 (VHLSS 2004). The survey is implemented by General Statistics Office of Vietnam with technical support from World Bank. In the VHLSS 2004 survey, there are 8813 households living in both rural and urban areas surveyed, including about 4300 households producing rice. From that sample, we chose randomly a sub-sample of 600 households. After calculating the efficiency, we dropped 5 extreme observations as it is known that DEA method is sensitive to outliers. The final sample includes 595 farms<sup>5</sup>.

#### **Rice output and inputs.**

The measure for output is the harvested rice quantity during last year. The inputs include ten categories: Fertilizers, pesticides, seed, equipment, 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, household 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 expenditures used for these inputs measured in current money value. Because quantity and price information for most inputs are not available, the measure of technical efficiency in this study encompasses certain input-based allocative efficiency. In our sample, on average, rice occupies for 46% of agricultural household outputs. This number is quite close to the percentage of rice production value in total agricultural production

---

<sup>5</sup> We choose a sample of 595 farms instead of 4300 farms to enable calculating the efficiencies by the DEA and bootstrap method. For example, to run a bootstrap of 2000 replications for a sample of 4300 farms will require simultaneously solving  $4300 \times (2 \times 2000 + 1) \approx 17.2$  millions linear equations, an overburdened task for an average PC (see Simar and Wilson 2000 for the calculation of number of equations). In our analysis, the number of linear equations to be solved in the bootstrap step is  $595 \times (2 \times 2000 + 1) \approx 2.4$  millions linear equations. It takes about 3.5 hours for doing it in a Pentium IV, 2.8 GHz computer with 704 Megabytes of RAM.

value for the country- 41.5% in 2001 (Vietnam Business Forum 2003). Summary statistics for these households are listed in Table 1.

**Table 1. Summary statistics for rice farming farms**

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
<b>Input and output vectors</b>				
Rice quantity (kgs)	7560	11125	100	100640
Rice value *	6562	8428	200	100048
Seed expenditures *	291	530	0	9900
Fertilizer expenditures *	976	1353	0	13800
Pesticide expenditures *	308	706	0	6540
Family hours for farming *	2184	1766	64	9432
Percent of rice (percent)	46	25	0.7	100
Estimated family hours for rice production (hours)	904	871	7.5	5333
Rice land area (square meters)	6991	8770	250	74000
Fixed asset and equipment value *	6414	12976	0	164500
Hired-in labor expenditure *	262	674	0	8750
Asset hire and maintenance *	529	964	0	6540
Small tool and energy *	98	255	0	8750
Other expenditure *	242	419	0	8312
<i>Household characteristics</i>				
Total number of household member	4.71	1.79	1	13
Number of adults/ total household members	0.65	0.21	0.2	1
Age of household head	49	14	21	87
Years of schooling of household head	6.59	3.51	0	12
<i>Production characteristics</i>				
Total farm output *	17391	20784	350	212773
Capital/Labor ratio	5.21	18.35	0	250
Land/Labor ratio	4.21	5.54	0.22	81.6
Non-farm income/Total household income	0.32	0.26	0	0.96
Number of extension visits	1.70	3.11	0	27

\* in thousand VND.

Table 2 indicates the percentage of expenditure for rice farming. Fertilizers and asset hire and maintenance are the highest expenditure items. Approximately 58% of variable input

expenditures are used for buying seeds (10.8%), fertilizers (36.1%) and pesticides (11.4%). About 29% are used for hiring outside labor (9.7%), hiring cattle, fixed asset and asset maintenance (19.5%). The rest, about 13% are used for buying or repairing small tools, buying energy (3.6%) and tax, irrigation fee and other expenses (8.9%).

**Table 2: Percentage of expenditure for rice farming**

Expenditure	Percent
Seeds	10.8
Fertilizers	36.1
Pesticides	11.4
Hired-in labor	9.7
Asset hire and maintenance	19.5
Small tool and energy	3.6
Other expenditure	8.9

## V. Empirical results

### Technical and Scale Efficiency

The estimated DEA and SFA efficiencies are in Table 3. The average technical efficiency estimated by DEA method is higher than by SFAA method. Similar results have been reached in Johansson (2005) for Swedish dairy farms, Kalaitzandonakes and Dunn (1995) for corn farms in Guatemala and Wadud and White (2000) for rice farmers in Bangladesh. Thiam et al (2001) reviews 27 stochastic stochastic frontier, six deterministic frontier and two DEA studies and reports that the average technical efficiency of all studies is 0.68 while the average technical efficiency in two DEA studies are 0.95. On the other hand, in Jaforullah and Premachandra (2003), Sharma et al (1999), the average DEA estimate and SPF estimate are similar under VRS assumption and in Paul et al (2004) and Brummer (2000), the average DEA estimates are smaller than SPF estimate. The differences in empirical studies in comparing these two approaches can be due to differences in the data characteristics, input and output variables as well as specification and estimation procedures.

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

**Table 3: DEA and SFAA estimates**

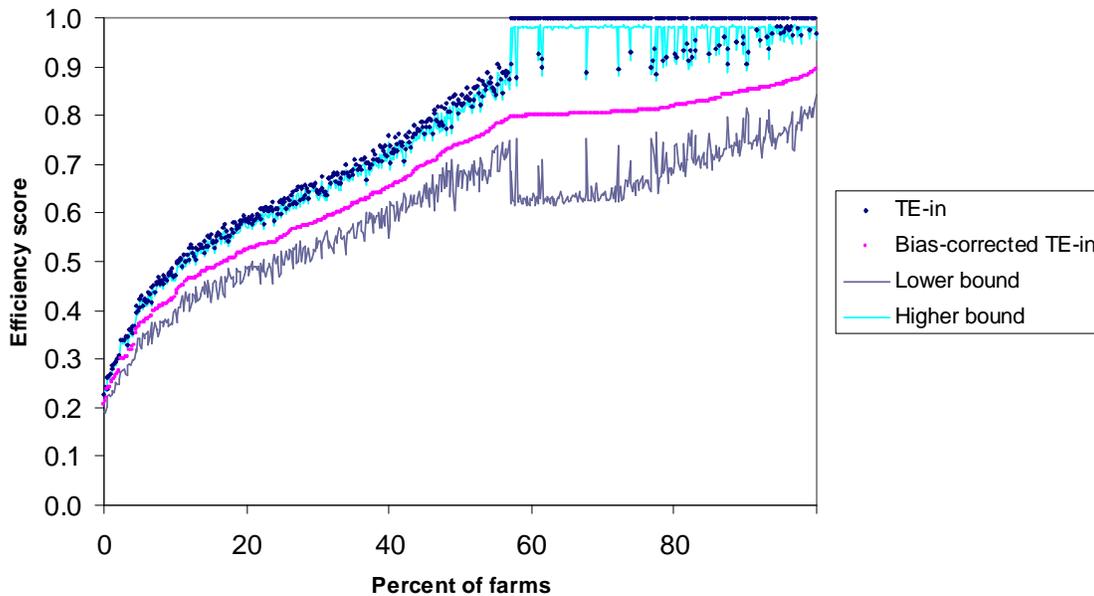
	DEA						Stochastic Frontier TE by SFAA	
	Output-based			Input-based (VRS)				
	CRS	(VRS)		Bias- corrected TE	Lower bound	Higher bound		SE
	$TE_{CRS}$	$TE_{VRS-OUT}$	$TE_{VRS-IN}$					
Average	0.704	0.765	0.785	0.678	0.593	0.771	0.890	0.634
Median	0.711	0.816	0.824	0.741	0.627	0.811	0.958	0.674
Std. Dev.	0.244	0.238	0.212	0.167	0.137	0.208	0.160	0.193
Min	0.090	0.174	0.228	0.205	0.190	0.224	0.09	0.109
Max	1	1.000	1	0.896	0.844	0.986	1	0.952

Estimates from 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 a “true” production frontier. Using bootstrap method as in Simar and Wilson (2000), the bias-corrected TE is significantly lower than the initial DEA estimates and closer with the stochastic frontier estimates.

Figure 2 shows the distribution of initial DEA estimates, bias-corrected DEA estimates and the 95-percent confidence interval for the input-based methods. The confidence interval is larger for the farms considered initially as technically efficient than other farms. If we only know the initial DEA estimates, it appears that on averages, rice farms in Vietnam can reduce their inputs by 27.4% and still can produce the same outputs. Yet, after correcting for the bias, they can reduce their inputs by 47.5% ( $1/0.678 - 1$ ) to produce the same level of outputs as before. An average can reduce their inputs in the range from 29.7% to 68.6% with 95% confidence interval.

By stochastic frontier method, the corresponding values 57.8% ( $=1/0.634 - 1$ ) for Cobb-Douglas 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 in Table 4. Based on paired t-test, on average, the technical efficiencies 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.

**Table 4: Paired t- tests and Spearman rank correlation tests**

Efficiency	Sample Mean		<i>t- ratio</i>	<i>Spearman rank correlation</i>
	<i>DEA</i>	<i>SFAA</i>		
Initial TE	0.785	0.634	19.16*	0.5284*
Bias-corrected TE	0.678	0.634	6.50*	0.5526*

\* significant at the 1% level

Figure 3 shows the percentage cumulative frequency distribution of technical efficiency from both DEA and SFAA method. Again, the correlation between bias-corrected technical efficiency and stochastic technical efficiency is also higher than between the uncorrected technical efficiency and stochastic technical efficiency.

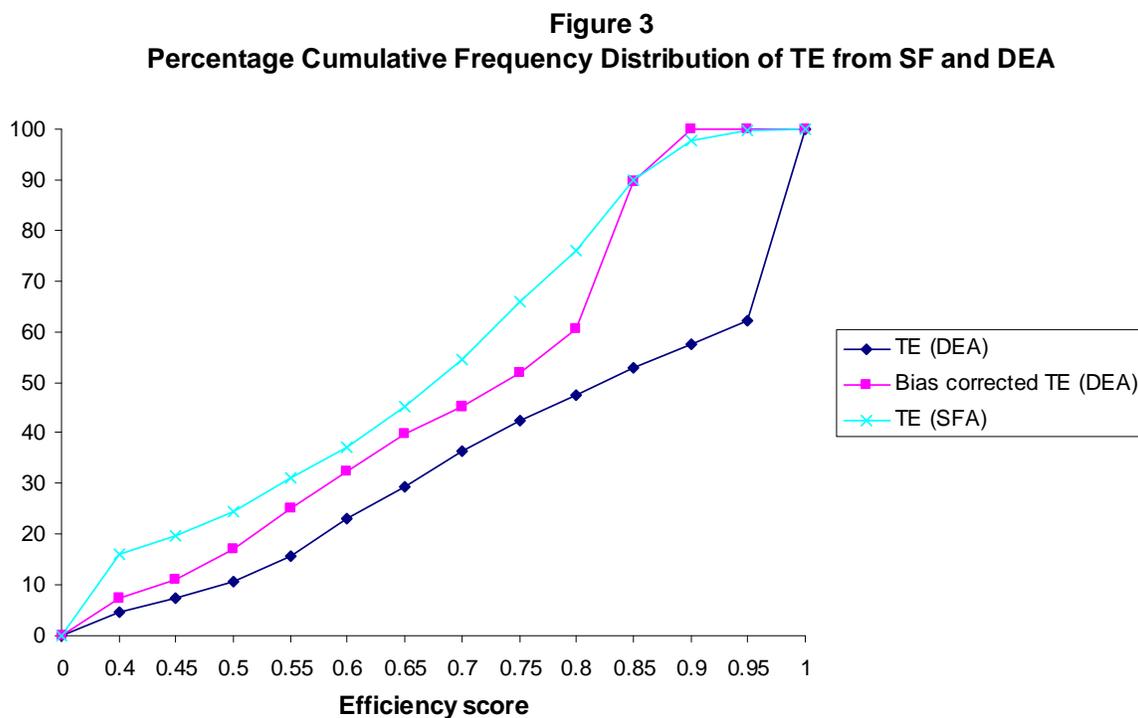
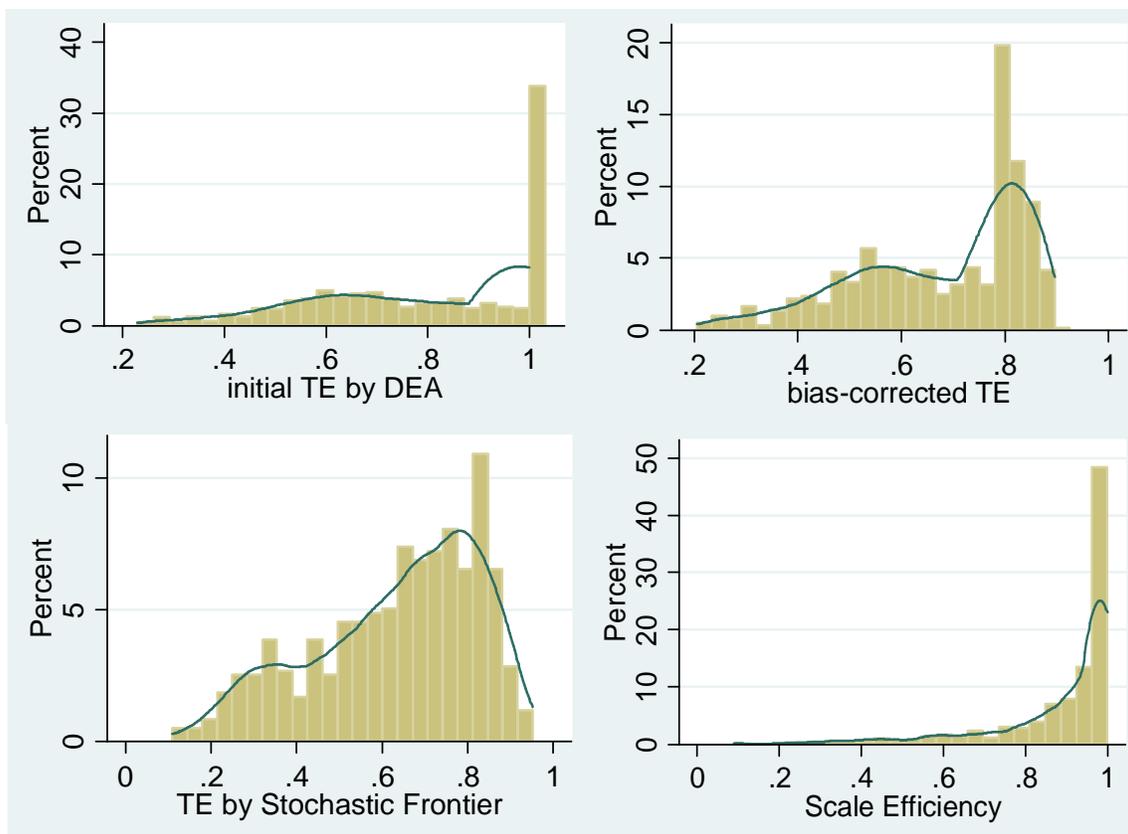


Figure 4 indicates the kernel density and histogram of estimates from deterministic DEA method, bias-corrected TE, TE by Stochastic Frontier and SE by DEA method.

Table 5 shows the distribution of technically efficient farm in the dataset according to DEA method. Farms in Southern Region- the main production region in Vietnam - are most effective. Farms in Center Region are least technically efficient. In addition, average technical efficiency and percentage of technical efficient farms are higher for large farms than for small farms and for diversified farms than for exclusive rice farms. Large farms are defined as farms with total farm output value higher than 15 million VND (about

\$1000)<sup>6</sup>. Mainly rice farms are farms with rice output equivalent more than 70% of total farm output value. About 70% of farms in our sample are mainly rice farms and 37% of farms are large farms.

**Figure 4: Kernel Density and Histogram of Initial TE by DEA, bias-corrected TE, TE by Stochastic Frontier and Scale Efficiency**



<sup>6</sup> That number is chosen arbitrarily based on the distribution of farm output value.

**Table 5: Distribution of average technical efficiency**

<b>Region</b>	<b>Average TE</b>	<b>Bias-corrected TE</b>	<b>Number of farms with TE=1</b>	<b>% of farms with TE=1</b>
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

### **Scale efficiency**

Similar to the distribution of technical efficiencies, 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% of total farms are working with optimal scale operation and a majority of farms (59%) are operating with increasing return to scale. That suggests that in general, a majority of rice farms can increase their operating scale to gain scale efficiency.

For the stochastic functional form, the sum of coefficients from the Cobb-Douglas production frontier is 1.098 implying increasing return to scale. We also reject the hypothesis of constant return to scale (sum of coefficient equal to one) at one-percent level.

**Table 6: Distribution of average scale efficiency**

	SE	Number of farms with			% with SE=1	Total farm output (mil. VND)
		SE=1	DRS	IRS		
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

### Factors associated with efficiency

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

Household characteristics variables include total number of household members (household size), adult ratio in the household, household head's age and household head's schooling. Household head's schooling is categorized in to four category: no formal education, with primary schooling (from 1 to 5 years), with secondary schooling (from 6 to 9 years) and with high schooling or higher (10 years and up). In our data, 32% of household heads have primary schooling, 45% have some secondary schooling, 14% have more than 9 years of schooling and only 7% never go to school.

Other variables 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), loans, modern irrigation, and regional binary variables which are categorized in two sets- one set include dummies for Center and South region with North being the reference region, and the other set include dummies for North East, North West, North Central Coast, South Central Coast, Central Highland, South East and Mekong River Delta with Red River Delta being the reference region.

Five models are developed to determine the factors affecting technical efficiency. Model 1 is the standard Tobit analysis in which the dependent variable is the original DEA estimate. This model is employed in most of papers using DEA method to estimate the effects on technical efficiency. However, this model has a disadvantage because it does not account for the bias and confidence interval in the DEA initial scores. Three models are developed with the information obtained from the bootstrap procedure. Model 2 is the weighted Tobit analysis. The dependent variable in Model 2 is the initial TEs calculated by DEA but with the weights equal to the reciprocals of the width between higher bounds and lower bounds. The idea is that, the higher the width is, the larger the measurement error can occur. Therefore, weighted Tobit analysis reduces the estimation errors by punishing the observations with larger width, or higher possibility of measurement error. Model 3 is the Tobit analysis with similar to Model 1 but the dependent variable being the bias-corrected

TE instead of initial TE. While this model is beneficial in correcting the bias from the deterministic model, it also has disadvantage because the variances of bias-corrected TEs are often considerably higher than the variances of initial TEs (Simar and Wilson 2000). Model 4 approaches the problem from the confidence interval, by method of maximum likelihood estimation for interval regressions, in which the interval points are lower and higher bounds from the confidence interval established by bootstrap procedure. Model 2, 3 and 4 can be called semi-parametric models because they utilize both the deterministic DEA method and the bootstrap method to derive statistical properties of the DEA estimator. Finally, model 5 is maximum likelihood estimation for stochastic frontier TEs.

The result in table 7 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), Llewelyn and Williams (1996) and Dhungana et al (2004).

Primary education of the household heads are positively related to the farmer technical efficiency in all models but the impact is more significant for the interval regression and for the stochastic frontier estimates. The impacts of secondary and higher education to technical efficiency are more ambiguous. While secondary and higher education are associated with higher technical efficiency indices as calculated by stochastic frontier, they are insignificant for those calculated by non-parametric (Model 1) and semi-parametric method (Model 2, 3, 4). In particular, they are negative but insignificant for Model 1 and 2, positive and insignificant for Model 4 and positive and significant for model 5. This indicates the role of primary education rather than secondary or higher education for improving farmers' efficiency.

One possibility is that farmers with higher education tend to shift to non-farm activities, therefore their education do not contribute for improving farm technical efficiency. Table 9 shows that is the case as average non-farm ratio increase with the household head's education level. A simple regression also indicates that non-farm ratio is positively associated with the household head's year of schooling at 5% 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 don't find any significant relationship between the household's highest education level and its technical efficiency. This finding suggests that the head's education may be a more important factor in deciding the household technical efficiency.

The land/labor ratio has a significant positive impact on technical efficiency for the nonparametric and semi-parametric models but not for stochastic frontier model. This means that increasing rice land is generally associated with better technical efficiency. Given the shortage and fragmentation of land in a populated economy as in Vietnam, this finding is expected. Based on World Rice Statistics of International Rice Research Institute (IRRI 2005), we estimated that nearly 90 percent of farms in Vietnam has 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 technical efficiency is insignificant in all models.

Farm size has significant positive effect on technical efficiency in all models except Model 5 where the effect is positive but insignificant. It indicates that farm operations in Vietnam are in general not optimal for technical efficiency. Modern irrigation also has positive effect but the effect is only strongly significant for stochastic frontier model.

Among the binary variables, land quality effect is positive in all models but only significant for Model 3, Model 4 and Model 5. The effect is strongly significant for model 4. Farms with better land quality generally are in better position to improve their technical efficiencies than those with poor land quality.

Farms with loans may have lower technical efficiencies than farms without loans although the effect is only significant for Model 1 and 2. This finding is as expected since farms with loan may be more constrained with the debt burden than those without loans. About 50% of the farms in our sample borrow money from various sources. The average loan value of a

borrowed household is 8.4 million VND, in which about 3.5 million VND is used for agricultural production, and about 4 million VND for consumption and the rest (nearly 1 million VND) is for non-agricultural production purpose. We also test if the purpose of borrowing affect technical efficiency and find that loans for agricultural production are associated with lower technical efficiency while there is no significant relation between consumption loans and farm technical efficiency.

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 South or in the Center. The impact of Center dummy is strongly significant at 1% level in all models while South dummy is insignificant in Model 1 and Model 3 and only significant at 1% level in Model 5. Yet, in Table 5, we see that average technical efficiency score is higher in the South than in the North. This fact can be explained by other factors, such as farm size since an average farm in the South is almost as big as 2.5 times an average farm in the North (see Table 6).

The effect of dummy variable for mainly rice farming is inconsistent: it is negative for the non-parametric and semi-parametric models (Model 1, 2 , 3 and 4) but positive for the parametric model (Model 5). Further investigations with separate subsets for exclusive rice farms and diversified may be necessary before we reach to a consistent conclusion about the difference in efficiency between these groups.

Other factors such as household size, household adult ratio, extension visits and rented land ratio are insignificant to technical efficiency in all models. The extension exposure in rural Vietnam is rather limited with only 46% of household in our sample have at least one visit by or to the extension service during the year and only 6% have more than five visits. The quality of extension service as rated by the farmers is average. On a scale from 1 to 4 in which '1' is the best and '4' the worst, the average helpfulness score of extension service is 1.88 for cropping decisions and 2.2 for livestock raising decisions. It is necessary to raise both the quantity and quality of extension services so that they can be of a more important role in improving technical efficiency of rice farming in Vietnam

**Table 7: Factors influencing technical efficiency (3 regions)**

Dependent variable	Standard	Weighted		Interval	Stochastic
	Tobit	Tobit	Bias-corrected	regression	frontier
	Model 1	Model 2	Model 3	Model 4	Model 5
	TE by DEA	TE by DEA	Bias-corrected TE	Lower and higher bounds	TE by SPF
Number of obs	595	595	595	595	595
LR chi2(18) =	101.1	92.5	62.3	108.8	62.3
Prob > chi2	0	0	0	0	0
Log likelihood	-211	-204.2	251.9	190	251.9
Adult ratio	0.011 (0.14)	-0.02 (-0.27)	-0.024 (-0.59)	-0.029 (-0.69)	0.02 (0.44)
Household size	-0.007 (-0.85)	-0.008 (-0.96)	-0.005 (-1.25)	-0.006 (-1.35)	0.005 (1.06)
Capital/Labor	0.050 (0.07)	-0.149(-0.15)	0.062 (0.16)	0 (0)	-0.282 (-0.68)
Land/Labor	0.021 (5.83) <sup>a</sup>	0.025 (5.66) <sup>a</sup>	0.005 (3.60) <sup>a</sup>	0.005 (3.6) <sup>a</sup>	0.001 (0.4)
Head's age	-0.003 (-2.75) <sup>a</sup>	-0.002 (-1.81) <sup>c</sup>	-0.001 (-1.26)	-0.001 (-1.07)	-0.001 (-0.38)
Primary	0.079 (1.64) <sup>c</sup>	0.082 (1.66) <sup>c</sup>	0.053 (2.03) <sup>b</sup>	0.057 (2.07) <sup>b</sup>	0.061 (2.1) <sup>b</sup>
Secondary	-0.031 (-0.63)	-0.002 (-0.05)	0.008 (0.3)	0.013 (0.45)	0.073 (2.49) <sup>b</sup>
High education	-0.038 (-0.68)	-0.041 (-0.73)	-0.002 (-0.08)	0.001 (0.04)	0.06 (1.76) <sup>c</sup>
Farm output	2.134 (2.67) <sup>a</sup>	2.085 (2.13) <sup>b</sup>	0.814 (2.14) <sup>b</sup>	0.878 (2.05) <sup>b</sup>	0.591 (1.4)
Land quality	0.013 (0.51)	0.037 (1.51)	0.027 (1.98) <sup>c</sup>	0.031 (2.2) <sup>b</sup>	0.083 (5.47) <sup>a</sup>
Non-farm ratio	0.002 (0.03)	-0.009 (-0.18)	0.012 (0.44)	0.013 (0.45)	-0.044 (-1.45)
Irrigation	0.002 (0.07)	0.003 (0.12)	0.014 (0.89)	0.015 (0.92)	0.068 (3.81) <sup>a</sup>
Extension visit	0.001 (0.29)	0.004 (0.9)	0.001 (0.53)	0.002 (0.69)	0.001 (0.49)
Rented land ratio	-0.024 (-0.27)	0.03 (0.31)	0.04 (0.81)	0.042 (0.81)	0.006 (0.12)
Borrow	-0.066 (-2.7) <sup>a</sup>	-0.043 (-1.74) <sup>c</sup>	-0.02 (-1.48)	-0.018 (-1.30)	-0.013 (-0.91)
Exclusive rice	-0.087 (-2.99) <sup>a</sup>	-0.109 (-3.58) <sup>a</sup>	-0.038 (-2.47) <sup>b</sup>	-0.04 (-2.47) <sup>b</sup>	0.053 (3.05) <sup>a</sup>
Center	-0.097 (-3.22) <sup>a</sup>	-0.119 (-4.18) <sup>a</sup>	-0.064 (-3.84) <sup>a</sup>	-0.067 (-3.89) <sup>a</sup>	-0.071 (-3.83) <sup>a</sup>
South	-0.044 (-1.22)	-0.085 (-2.25) <sup>b</sup>	-0.026 (-1.38)	-0.034 (-1.74) <sup>c</sup>	-0.112 (-5.42) <sup>a</sup>
Constant	1.02 (11.63) <sup>a</sup>	0.865 (9.33) <sup>a</sup>	0.733 (15.48) <sup>a</sup>	0.721 (14.49) <sup>a</sup>	0.494 (9.41) <sup>a</sup>

<sup>a b c</sup> significant at 1%, 5% and 10% respectively.

**Table 8: Factors influencing technical efficiency (8 regions)**

Dependent variable	Standard	Weighted		Interval	Stochastic
	Tobit	Tobit	Bias-corrected	regression	frontier
	Model 1	Model 2	Model 3	Model 4	Model 5
			Bias-corrected	Lower and	
	TE by DEA	TE by DEA	TE	higher bounds	TE by SPF
Number of obs	595.00	595.00	595.00	595.00	595.00
LR chi2(18) =	108.18	99.17	65.18	63.44	151.04
Prob > chi2	0.00	0.00	0.00	0.00	0.00
Log likelihood	-207.47	-200.89	253.38	-891.18	211.15
Adult ratio	0.033 (0.44)	0.010 (0.14)	-0.019 (-0.47)	-0.024 (-0.56)	0.020 (0.46)
Household size	-0.007 (-0.90)	-0.008 (-0.96)	-0.005 (-1.15)	-0.006 (-1.23)	0.008 (1.78) <sup>c</sup>
Capital/Labor	0.115 (0.80)	0.199 (1.10)	-0.012 (-0.26)	-0.012 (-0.24)	-0.161 (-3.32) <sup>a</sup>
Land/Labor	0.023 (6.13) <sup>a</sup>	0.025 (5.66) <sup>a</sup>	0.005 (3.65) <sup>a</sup>	0.005 (3.57) <sup>a</sup>	0.002 (1.16)
Head's age	-0.003 (-2.63) <sup>a</sup>	-0.002 (-1.70) <sup>c</sup>	-0.001 (-1.29)	-0.001 (-1.08)	-0.001 (-1.34)
Primary	0.086 (1.76) <sup>c</sup>	0.094 (1.88) <sup>c</sup>	0.053 (1.99) <sup>b</sup>	0.056 (2.01) <sup>b</sup>	0.037 (1.29)
Secondary	-0.024 (-0.47)	0.006 (0.11)	0.007 (0.24)	0.011 (0.37)	0.029 (0.97)
High education	-0.029 (-0.51)	-0.033 (-0.57)	-0.003 (-0.1)	0.000 (0.01)	0.015 (0.43)
Farm output	1.933 (2.18) <sup>b</sup>	1.824 (1.75) <sup>c</sup>	0.853 (2.09) <sup>b</sup>	0.917 (2.01) <sup>b</sup>	1.055 (2.41) <sup>b</sup>
Land quality	0.018 (0.68)	0.041 (1.55)	0.025 (1.73) <sup>c</sup>	0.028 (1.93)	0.062 (4.07) <sup>a</sup>
Non-farm ratio	-0.004 (-0.08)	-0.003 (-0.07)	0.015 (0.55)	0.016 (0.57)	-0.040 (-1.35)
Irrigation	0.023 (0.73)	0.013 (0.43)	0.012 (0.73)	0.013 (0.72)	0.049 (2.69) <sup>a</sup>
Extension visit	0.001 (0.38)	0.004 (1.05)	0.001 (0.62)	0.002 (0.77)	0.002 (0.71)
Rented land ratio	-0.020 (-0.22)	0.029 (0.30)	0.042 (0.85)	0.041 (0.81)	0.000 (0.01)
Borrow	-0.068 (-2.79) <sup>a</sup>	-0.044 (-1.79) <sup>c</sup>	-0.021 (-1.55)	-0.019 (-1.35)	-0.019 (-1.31)
Exclusive rice	-0.079 (-2.66) <sup>a</sup>	-0.104 (-3.42) <sup>a</sup>	-0.038 (-2.44) <sup>b</sup>	-0.041 (-2.47) <sup>b</sup>	0.053 (3.1) <sup>a</sup>
North East	0.019 (0.48)	0.001 (0.02)	-0.010 (-0.49)	-0.012 (-0.53)	-0.041 (-1.8) <sup>c</sup>
North West	0.043 (0.76)	0.039 (0.65)	-0.003 (-0.11)	-0.003 (-0.09)	-0.152 (-4.73) <sup>a</sup>
NCC	-0.088 (-2.16) <sup>a</sup>	-0.080 (-2.07) <sup>b</sup>	-0.067 (-2.97) <sup>a</sup>	-0.068 (-2.92) <sup>a</sup>	-0.099 (-4.09) <sup>a</sup>
SCC	-0.076 (-1.66) <sup>c</sup>	-0.156 (-3.66) <sup>a</sup>	-0.068 (-2.67) <sup>a</sup>	-0.075 (-2.86) <sup>a</sup>	-0.113 (-4.11) <sup>a</sup>
Central Highland	0.098 (1.42)	0.015 (0.19)	0.003 (0.08)	-0.006 (-0.15)	-0.086 (-2.21) <sup>b</sup>
South East	-0.032 (-0.47)	-0.075 (-1.03)	-0.068 (-1.9) <sup>c</sup>	-0.083 (-2.16) <sup>b</sup>	-0.246 (-6.4) <sup>a</sup>
Mekong Delta	-0.083 (-1.79) <sup>c</sup>	-0.111 (-2.27) <sup>b</sup>	-0.028 (-1.15)	-0.034 (-1.35)	-0.145 (-5.63) <sup>a</sup>
Constant	0.965 (9.87) <sup>a</sup>	0.813 (7.84) <sup>a</sup>	0.735 (13.92) <sup>a</sup>	0.723 (13) <sup>a</sup>	0.592 (10.46) <sup>a</sup>

<sup>a b c</sup> significant at 1%, 5% and 10% respectively.

NCC and SCC stand for North Central Coast and South Central Coast respectively

Table 8 is similar to Table 7 except that now there are eight regions. The main rice production in Vietnam is in Red River Delta and Mekong River Delta. About 52% of Vietnam's rice is produced in the Mekong River Delta and 18% in the Red River Delta (IRRI 2006). Table 10 indicates that, on average, a farm household in Mekong River Delta produces as much as four times than a farm household in Red River Delta.. Yet, it seems that other factors being equal, rice growing in Red River Delta is more efficient than Mekong River Delta. Rice farming households in North Central Coast and South Central Coast are least technically efficient among all the regions. If based on stochastic frontier method alone, Red River Delta is most favorable for technical efficiency. This result is consistent with the rice yield data. In 2002, the rice yield in the Mekong Delta was 4611 kg per hectare (300kg over the average in Vietnam), in the Red River Delta was 5641 kg/ha (1058kg over the average) (Vietnam Business Forum 2003).

**Table 9 : Non-farm ratio and farmer's schooling**

	Average non-farm ratio
No school	0.29
Primary school	0.31
Secondary school	0.32
High school or above	0.42

**Table 10: Rice value, quantity and percentage of an average farm distributed by regions**

	Rice percentage	Rice Value (million VND)	Rice Quantity (tons)
Red River Delta	47.5	4.60	5.3
North East	35.8	4.29	4.4
North West	45.0	4.46	4.0
North Central Coast	44.5	4.98	5.5
South Central Coast	45.4	4.04	4.2
Central Highlands	27.8	4.55	4.3
South East	42.0	7.66	8.5
Mekong River Delta	63.7	18.41	23.6

## Summary and Conclusion

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

The mean technical efficiency 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. Confidential interval is also established for the bias-corrected estimates. Stochastic frontier estimation yields a mean estimate of 0.634. The variance of estimates from DEA and SFA methods are similar. But the variances of bias-corrected TEs after bootstrapping are significantly lower than parametric approach. The Spearman correlation test confirms that our efficiency scores calculated from different approaches are positively and significantly correlated.

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. So, a farm can reduce its cost by 30-69% depending upon the method employed. A further 12 percent cost reduction can be obtained by operating with optimal scale. A majority of farms, particulaly in the Center region, are operating with increasing return to scale. 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 efficiency 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 technical efficiency. 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 technical efficient. The Mekong River Delta, which produces more than a half of the whole country rice production, has more potential for improving technical efficiency. The land in this region one of the best rice growing regions of the world and there is still capability for increasing rice area. Almost all arable land is under intensive cultivation in the northern while only 67% of the arable land is under cultivation in the Mekong Delta (UNEP 1998). On the other hand, factor such as non-farm ratio or extension support do not significantly affect farm household technical efficiency. For extension support, the reason may be due to limited exposure of farmers to extension personnels. Policies leading to improvement of farm education, land quality and land holding will be beneficial for improving farmers' technical efficiency. The distribution of technical efficiency and scale efficiency across regions also provides useful information for policy makers in raising efficiency for each region.

#### REFERENCES

- Aigner, D. J., C. A. K. Lovell, and P. Schmidt. (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6, 21-37.
- Balcombe, K., S. Davidova, and L., Latruffe. (2005). "Application of a Double Bootstrap to Investigation of Determinants of Technical Efficiency of Farms in Central Europe." *Working Paper*, INRA- Unité ESR Rennes, France.
- Banker, R.D. (1993). "Maximum Likelihood, Consistency and Data Envelopment Analysis: a Statistical Foundation." *Management Science* 39(10), 1265-1273.
- Battese, G.E and T.J. Coelli. (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India." *The Journal of Productivity Analysis* 3, 153-169.
- Brooke, A., D. Kendrick, A. Meeraus, and R. Raman. (1998). *GAMS: A User's Guide*. GAMS Development Corporation, Washington, D.C.
- Brümmer, B. (2001). "Estimating Confidence Intervals for Technical Efficiency: The Case of Private Farms in Slovenia." *European Review of Agricultural Economics* 28(3), 285-306.
- Bymes, P., R. Fare, S. Grosskopf, and S. Kraft. (1987). "Technical Efficiency and Size: The Case of Illinois Grain Farms." *European Review of Agricultural Economic* 14, 367-81.
- Charnes, A., W. Cooper, and E. Rhodes. (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research* 2, 429-444.

- Chavas, J-P, R. Petrie, and M. Roth. (2005). "Farm Household Production Efficiency: Evidence from The Gambia." *American Journal of Agricultural Economics* 87 (1), 160-179.
- Coelli, T. J. (1996). "A guide to DEAP version 2.1: A Data Envelopment Analysis (Computer) program." *CEPA Working Papers No.8/96*. ISSN 1327-435X. ISBN 1 86389 4969.
- Coelli, T.J. and G.E. Battese. (1996). 'Identification of factors which influence the technical inefficiency of Indian farmers.' *Australian Journal of Agricultural Economics* 40, 103-128.
- Dhungana, R. B., P. L. Nuthall, and G.V. Nartea. (2004). "Measuring the Economic Inefficiency of Nepalese Rice Farms Using Data Envelopment Analysis.", *The Australian Journal of Agricultural and Resource Economics*, 48 (2), 347-369.
- FAO (2006). *Foot Outlook- Global Market Analysis* No 1, 2006, online at <http://www.fao.org/docrep/009/J7927e/j7927e01.htm>.
- Färe, R., S. Grosskopf, and C.A.K. Lovell. (1994). *Productivity Frontiers*. Cambridge: Cambridge University Press.
- Farrell, M. J. (1957) "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society, Series A*, 120, 253-90.
- Ferrier, G.D., and J.G. Hirschberg. (1997), "Bootstrapping Confidence Intervals for Linear Programming Efficiency Scores: With an Illustration Using Italian Bank Data." *Journal of Productivity Analysis* 8, 19–33.
- Government of Vietnam Information Gateway (2006). "Export market in February 2006", online at <http://tintuc.egov.gov.vn/tintuc.nSFA/0/F9F697AB2B233F9A4725711A0047ED51?OpenDocument&fullmode>, accessed on Sept, 2006.
- Green, W.H. (1997). "Frontier Production Functions", in *Handbook of Applied Econometrics*, Blackwell Publishers Ltd., Oxford, UK.
- Grosskoff, S. (1996). "Statistical Inference and Nonparametric Efficiency: A Selective Survey." *The Journal of Productivity Analysis* 7, 161-176 (1996).
- Hollingsworth, B., (1999). "Data Envelopment Analysis and Productivity Analysis: a Review of the Options." *The Economic Journal*, 458-462.
- IRRI (2006). "Vietnam", online at <http://www.irri.org/science/cnyinfo/vietnam.asp>, accessed at August 2006.
- IRRI (2005). *World Rice Statistics*, online at <http://www.irri.org/science/ricestat/>, accessed at August 2006.
- Kalaitzandonakes, N.G., and E. G. Dunn. (1995). "Technical Efficiency, Managerial Ability and Farmer Education in Guatemala Corn Production: A Latent Variable Analysis." *Agricultural Resource Economics Review* 24, 36-46.

Kalvelagen, E. (2004). "Efficiently Solving DEA Models with GAMS", GAMS Corporation, online at <http://www.gams.com/~erwin/dea/dea.pdf> , assessed 04/15/2006.

Kneip, A, Park, B & Simar, L. (1998). "A Note on the Convergence of Nonparametric DEA Efficiency Measures." *Econometric Theory*, 14, 783-793.

Krasachat. (2003). "Technical Efficiencies of Rice Farms in Thailand: a Non-Parametric Approach." Paper presented to the 2003 Hawaii International Conference on Business, Honolulu, June 18-21, 2003.

Kompas, T. (2004). "Market reform, productivity and efficiency in rice production." *International and Development Economics Working Papers*, Asia Pacific School of Economics and Government, The Australian National University.

Korostelev, A., L. Simar, and A. Tsybakov. (1995). "Efficient Estimation of Monotone Boundaries." *The Annals of Statistics*, 23, 476-489.

Latruffe, L., K. Balcombe, S. Davidova, and K. Zawalinska. (2005). "Technical and Scale Efficiency of Crop and Livestock Farms in Poland: Does Specialization Matter?." *Agricultural Economics* 32, 281-296.

Llewelyn, R.V and J.R. Williams. (1996). "Non-Parametric Analysis of Technical, Pure Technical and Scale Efficiencies for Food Crop Production in East Java, Indonesia." *Journal of Agricultural Economics* 15, 113-126

Löthgren, M., and M. Tambour. (1996). "Testing Scale Efficiency in DEA Models: A Bootstrapping Approach." *Applied Economics* 31(10), 1231-37.

Löthgren, M. (1998). "How to Bootstrap DEA Estimators: A Monte Carlo Comparison." *Working Paper Series in Economics and Finance #223*, Department of Economic Statistics, Stockholm School of Economics, Sweden.

Nghiem, H. S. and T. Coelli. (2002). "The Effect of Incentive Reforms Upon Productivity: Evidence from the Vietnamese Rice Industry." *The Journal of Development Studies* 39 (1), 74-93.

Olesen, O.B and N.C. Peterson. (1996). "A Presentation of GAMS for DEA." *Computers and Operations Research* 23 (4), 323-339.

Ortner, K. M., J. Hambrusch, and L. Kirner. (2006). "The Efficiency of Dairy Farms in Austria: Do Natural Conditions Matter?." Federal Institute of Agricultural Economics, Vienna 2006, online at <http://www.fat.admin.ch/eaee96/abstracts/s88.pdf>, accessed on August 2008.

Olson, K., and L.H. Vu. (forthcoming). "Changes in Economic Efficiency and Factors Explaining Differences between Minnesota Farm Households." Monograph, Department of Applied Economics, University of Minnesota.

Paul, C.J.M., R. Nehring, D. Banker and A. Somwaru. (2004). "Scale Economics and Efficiency in U.S. Agriculture: Are Traditional Farms History?." *Journal of Productivity Analysis* 22, 185-205.

- Rio, A. R and G. Shively. (2005). "Farm Size and Nonparametric Efficiency Measurements for Coffee Farms in Vietnam." Selected Paper Prepared for Presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005.
- Sharma, K. R., P. Leung and H. M. Zaleski. (1999). "Technical, Allocative and Economic Efficiencies in Swine Production in Hawaii: A Comparison of Parametric and Nonparametric Approaches." *Agricultural Economics* 20, 23-35.
- Seyoum, E.T., G.E. Battese, and E.M. Fleming. (1998). "Technical Efficiency and Productivity of Maize Producers in Easter Ethiopia: A Study of Farmers Within and Outside the Sasakawa Global 2000 Project." *Agricultural Economics* 19, 341-348.
- Shafiq, M., and T. Rehman. (2000). "The Extent of Resource Use Inefficiencies in Cotton Production in Pakistan's Punjab: an Application of Data Envelopment Analysis." *Agricultural Economics* 22, 321-330.
- Simar, L., and P., Wilson. (1998). "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models." *Management Science* 44(1), 49-61.
- Simar, L., and P., Wilson. (1999). "Some Problems with the Ferrier/Hirschberg Bootstrap Idea." *Journal of Productivity Analysis* 11, 67-80.
- Simar, L., and P., Wilson. (2000). "A General Methodology for Bootstrapping in Non-Parametric Frontier Models." *Journal of Applied Statistics* 27(6), 779-802.
- Simar, L., and P., Wilson. (2004). "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes." *Working paper*, Institut de Statistique, Université Catholique de Louvain.
- Thiam, A., B.E. Bravo-Ureta, and T.E. Rivas. (2001). "Technical Efficiency in Developing Country Agriculture: A Meta-Analysis." *Agricultural Economics* 25, 235-243.
- Tran, T. U., and K. Kajisa. (2006). "The Impact of Green Revolution on Rice Production in Vietnam." *The Developing Economies* XLIV-2, 167-89.
- Tran, V. H. S., T. Coelli., and E. Fleming. (1993). "Analysis of the Technical Efficiency of State Rubber Farms in Vietnam." *Agricultural Economics* 9 (3), 183-201.
- UNEP. (1998). "Land Use/Land Cover Change In South-East Asia." Environment Assessment Technical Reports.
- Vietnam Business Forum. (2003). "Vietnam Rice Development Strategy." *Vietnam Business Forum* 4 (131), October 25, 2003.
- Wadud, A. (2003). "Technical, Allocative, and Economic Efficiency of Farms in Bangladesh: a Stochastic Frontier and DEA Approach." *The Journal of Developing Areas* 37 (1), 109-126.
- Wadud., A and B. White. (2000). "Farm Household Efficiency in Bangladesh: A Comparison of Stochastic Frontier and DEA Methods." *Applied Economics* 32, 1665-1673.

Wilson, P.W. (2005). "Frontier Efficiency Analysis with R. FEAR 0.913 User's Guide." Online at <http://www.eco.utexas.edu/faculty/Wilson/Software/FEAR/fear.html>, accessed at August 2006.

World Bank. (2006). "Vietnam- At a Glance." Online at [http://devdata.worldbank.org/AAG/vnm\\_aag.pdf](http://devdata.worldbank.org/AAG/vnm_aag.pdf), accessed at Sep, 2006

**Appendix A: Bootstrapping procedure for technical efficiency (CRS case) as in Simar and Wilson (1998, 2000)**

- i. Calculate the DEA efficiency scores under constant returns to scale (CRS) for each farm among N farms as in equation (1), denoted as  $\hat{\theta}_i$  for the  $i^{\text{th}}$  farm.
- ii. Let  $\beta_1^*, \dots, \beta_k^*$  be a simple bootstrap sample from  $\hat{\theta}_1, \dots, \hat{\theta}_k$ . Generate a random sample of size k for the random generator:

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

where h is the bandwidth of a standard normal kernel density and  $\varepsilon_i^*$  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

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

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

- iv. For  $i=1, \dots, N$ , a pseudo data set of  $(x_{i,b}^*, y_{i,b}^*)$  where  $x_{i,b}^* = (\hat{\theta}_i / \theta_i^*) x_i$  and  $y_{i,b}^* = y_i$  with  $x_i, y_i$  the original input and output vectors of the  $i^{\text{th}}$  farm, respectively.
- v. Calculate the new DEA score  $\hat{\theta}_i^*$  for each farm by taking the pseudo data as reference

vi. Repeat step (i) to (iv) for B times to yield B new DEA technical efficiency scores  $\hat{\theta}_i^*$  for  $i=1, \dots, N$ .

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

$$\widehat{bias}_B(\hat{\theta}_i) = B^{-1} \sum_{b=1}^B \hat{\theta}_i^* - \hat{\theta}_i .$$

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

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  $a_\alpha, b_\alpha$  such that  $\text{Prob}(-b_\alpha \leq \hat{\theta}_i^* - \hat{\theta}_i \leq -a_\alpha) = 1 - \alpha$ . Since we do not know the distribution of

$(\hat{\theta}_i^* - \hat{\theta}_i)$ , we can use the bootstrap values to find  $\hat{a}_\alpha, \hat{b}_\alpha$  such that

$\text{Prob}(-\hat{b}_\alpha \leq \hat{\theta}_i^* - \hat{\theta}_i \leq -\hat{a}_\alpha) = 1 - \alpha$ . It involves sorting the value of  $(\hat{\theta}_i^* - \hat{\theta}_i)$  for  $b=1, \dots, B$  in increasing order and deleting  $((\alpha/2) \times 100)$  percent of the elements at either end of this sorted array and setting  $-\hat{a}_\alpha$  and  $-\hat{b}_\alpha$  at the two endpoints, with  $\hat{a}_\alpha \leq \hat{b}_\alpha$ .

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.