STOCHASTIC PRODUCTION FRONTIER AND PANEL DATA: MEASURING PRODUCTION EFFICIENCY OF WETLAND RICE FARMS IN WEST JAVA

Erwidodo*)

ABSTRAk

Parameter fungsi produksi yang diestimasi secara statistik, yang umumnya dilakukan dengan metoda estimasi Least Square (LS), merupakan parameter dari fungsi produksi **rataan (average)**. Dengan cara ini, tingkat (in) efiseinsi teknis, sebagaimana disebutkan dalam teori ekonomi produksi, sulit untuk dihitung. Konsep estimasi fungsi produksi **frontier**, yang belakangan ini mulai populer, memungkinkan kita untuk mengestimasi tingkat inefisiensi produksi secara lebih tepat dan konsisten dengan teori ekonomi produksi. Fungsi produksi frontier ini dapat diduga dengan menggunakan data **cross-section** maupun dengan data **panel**. Ketersediaan data panel memungkinkan pendugaan tingkat inefisiensi produksi secara lebih konsisten dengan cukup menggunakan metoda modifikasi dari LS. Dalam tulisan ini dikemukakan konsep dan penerapan fungsi produksi frontier dengan menggunakan data panel dari usahatani padi sawah dibeberapa daerah produsen padi sawah dikawasan DAS Cimanuk, Jawa Barat. Hasil analisa memperlihatkan bahwa tingkat inefisiensi teknis dalam produksi padi sawah berkisar antara 3,4 – 12 persen, atau rata-rata 6.5 persen. Dengan menggunakan asumsi tertentu, secara kasar dapat diduga jumlah kehilangan hasil produksi padi di Jawa Barat sebesar 0.45 juta ton per tahun.

INTRODUCTION

A production function describes technical relationships that transform inputs into outputs. It also shows the maximum possible output (frontier) attainable from a given combination of inputs. There cannot be any point above the production frontier. The distance a firm lies below its production frontier measures the level of inefficiency.

A production process can be inefficient in two ways. It is technically inefficient if it fails to produce maximum output from a given input bundle. It is price or allocatively inefficient if the marginal revenue product of an input is not equal to the marginal cost of that input, resulting in utilization of inputs in the wrong proportions for given input and output prices. A combination of these two is usually referred to as total productive or economic inefficiency.

There is, however, a discrepancy between the above definition and the one that is statistically estimated. The letter, which is usually referred to as a "non-frontier" or "average" production function and estimated by ordinary least square (OLS), allows some firms to be above the "fitted" function. One can use the

---

*) Researcher, Center for Agro Socioeconomic Research (CASER), Bogor.

"average" function to estimate technical inefficiency under a certain restrictive assumption. The result, however, cannot be called a pure measure of technical inefficency since it also includes random variability.

The frontier production function is designed to bridge the gap by introducing the error term to represent an (in) efficiency measure. This paper, using a stochastic production frontier concept, aims to estimate the level of inefficiency on the rice farms in West Java. A panel (longitudinal) data set, that is a cross-section of individual observed over time, is used in the analysis.

ANALYTICAL FRAMEWORK

Stochastic Production Frontiers and Technical Efficiency

Forsund, Lovell and Schmidt (1980), in their survey on frontier production functions, distinguished four types of production frontier namely (1) deterministic non-parametric frontier, (2) deterministic parametic frontier, (3) deterministic statistical frontier, and (4) stochastic statistical frontier. This section briefly reviews the stochastic statistical production frontier.

The stochastic statistical production frontier, which is commonly known as stochastic production frontier, was developed independently by Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977). In the deterministic frontier model the variation in firm performance relative to the frontier is attributed to inefficiency, thereby ignoring the possibility of variation due to factors not under the control of any firm such as weather variation, machine breakdown, luck, which is usually referred to as statistical noise (Forsund *et al.*, 1980). Combining these two together, as in deterministic frontiers, and labelling it as inefficiency is not appropriate. The statistical noise needs to be separated from the controllable factors that are designated as inefficiency. This is the essential idea behind the stochastic frontier model.

This model, which can also be referred to as a composed error model, is stochastic in a sense that it capture exogeneous shocks beyond the control of firms. The model is described as follow:

$$Y_i = f(X_i;B) + e_i \quad \text{(E.1)}$$

where Y_i is the maximum amount of output obtainable from X_i, a vector of non-stochastic productive inputs of the i^{th} firm, and B is a vector of unknown parameters to be estimated. In addition

$$e_i = v_i - u_i \quad \text{(E.2)}$$

where v_i, the error component representing random noise, is assumed to be distributed normally with zero mean and variance of σ_v^2, while u_i, the non-negative error component representing technical inefficiency, is assumed to be distributed either with a "half normal" density or with an exponential density, both with mode at $u=0$.

The major weakness of the model is its difficulty in estimating individual inefficiency although average inefficiency for the population and its variance are available as

$$E(u) = \underline{u} = \sigma_u(2/\pi)^{0.5} \qquad \text{(E.3)}$$

$$V(\underline{u}) = \sigma_u^2 (1 - 2/\pi) \qquad \text{(E.4)}$$

Jondrow, Lovell, Materov and Schmidt (1982) show that by assuming v_i in normal and u_i is positive half-normal, v_i and u_i are independent, and that inefficiency is independent of the regressors, then an estimate of the individual inefficency for each firm can be obtained, although not consistently. The estimate u_i is based on the conditional mean of u_i given e_i, that is:

$$E(u|e_i) = \sigma^*[f(e_i\, s/\sigma^*)/(1-F(e_i\, s/\sigma^*))] - e_i/\sigma^* \qquad \text{(E.5)}$$

where

$$\sigma^* = (\sigma_u^2 \sigma_v^2)(\sigma_u^2 + \sigma_v^2)^{-1}$$

$$s = \sigma_u^2 / \sigma_v^2$$

The symbols f and F represent the standard normal density (pdf) and cumulative distribution function (cdf), respectively. By replacing e_i with its estimate (e_i) and σ^*, s by their estimates, one can estimate u_i. Using this method Bagi and Huang (1983) as cited in Seale (1985), estimate individual efficiency for 193 farms in Tennessee. More recently, Battese and Coelli (1988) presented a generalization of this method for a given panel data of the Australian dairy farms, which will be discussed in the next section.

Panel Data and Stochastic Production Frontiers

The estimation of individual technical inefficiency from a set of panel data was done by Hoch (1955; 1962). He, however, used an "average" (non-frontier) function rather than frontier function. Hoch assumed that firms maximize anticipated profit, and then estimated the production parameters using covariance analysis.

There are graet potential advantages for modifying existing frontier models to allow the use of panel data. Schmidt and Sickles (1984) pointed out three difficulties in applying stochastic production frontier models using cross section data. First, one can estimate technical inefficiency of each firm but not consistently.

Second, separation of inefficiency measures from statistical noise depends on specific assumptions about the distribution of technical inefficiency. Third, the assumption that inefficency is independent of regressors is not valid if a firm knows its level of technical inefficiency. These difficulties will analogously be found in using cross-section stochastic profit frontiers.

With the availability of.panel data these problems can be avoided. First, if there are T observations on each firm, then the technical inefficiency of a particular firm can be estimated consistently as T tends to infinity. Second, any distribution of technical inefficiency need not be assumed if these are treated as firm-specific effects. Third, no assumption is needed regarding the independence of technical inefficiency and the regressors.

This section draws heavily from Schmidt and Sickles (1984) article, with some modification in notation. Consider a production function as:

$$Y_{it} = a_o + X_{it}\, a + v_{it} - u_i \quad \text{(E.6)}$$

Here, i = 1,2 ...N indexes firms and t = 1,2 ...T indexes time period. The value y_{it} is output of the i^{th} farm at time t, whereas x_{it} is a vector of K inputs. The v_{it} are assumed uncorrelated with regressors and distributed iid $N(0, \sigma_v^2)$. The u_i represent technical inefficiency and $u_i \geqslant 0$ for all i. It is also assumed that u_i is i.i.d. with mean U and variance σ_u^P and is independent of the v_{it}. A particular distribution for u_i may or may not be assumed. Furthermore, the u_i may or may not be assumed to be correlated with regressors.

For T=1 (a single cross section) the model is the stochastic frontier of Aigner, Lovell and Schmidt (1977). For T > 1, it is a straight forward generalization of that model, and it fits the usual framework in the panel-data literature with individual effects but no time effects. The only difference from the standard panel data model is that individual effects are one-sided.

The equation (3.6) can be rewritten in two ways. First, let $E[u_i] = \underline{u} > 0$, and define

$$a_o^* = a_o - \underline{u} \text{ and } u_i^* = u_i - \underline{u}$$

so that u_i^* are iid with mean 0, Equation (3.6) can then be rewritten as

$$Y_{it} = a_o^* + X_{it}\, a + v_{it} - u_i^* \quad \text{(E.7)}$$

with the error terms v_{it} and u_i^* have zero mean. Most of the results of panel data literature can be applied directly, except those that hinge on normality.

Secondly, define

$$a_{oi} = a_o - \underline{u} = a_o^* - u_i^*$$

and then rewrite the model into

$$Y_{it} = a_{oi} + X_{it}'a + V_{it} \quad \text{(E.8)}$$

This is exactly a variable intercept model. This model can be estimated using either a dummy variable estimator or Generalized Least Square (GLS) estimator.

Fixed Effect Model: Dummy Variable Estimator

This estimator treats the u_i as fixed, that is, it estimates a separate intercept for every individual firm as in (E.8). This can be done by suppressing the constant term and adding a dummy variable for each of the N frams or, equivalently, by keeping the constant term and adding (N-1) dummies. Another equivalent procedure is to apply the within transformation, that is, to apply OLS after transforming the data in terms of deviations from the farm means.

The advantage of the within estimator is that its consistency does not hinge on uncorrelatedness of the regressors and the effects. It also does not depend on the distribution of the effects, since in treating them as fixed it simply proceeds conditionally from whatever their realizations are. The estimates of a is consistent as either N or T tends to infinity. Consistency of he individual estimated intercept requires $T \rightarrow \infty$.

A considerable disadvantege of the within estimator is that it is impossible to include in the specification the time invariant regressors even though they vary across farms. In this case the estimated individual effects will include the effects of all variables that are fixed within the sample at the farm level, possibly including some that are not in any sense a representation of inefficiency (Schmidt and Sickles, 1984).

In the case of the frontier function, if N is large, we can use the fact that $u_i \geqslant 0$ to appropriately normalize the individual effects (u_i) and the overall constant (a_o). If N estimated intercept are $\hat{a}_{o1}$, $\hat{a}_{o2}$,....$\hat{a}_{oN}$, simply define

$$\hat{a}_o = \max (\hat{a}_{oi}) \qquad \text{(E.9)}$$

$$\hat{u}_i = \hat{a}_o - \hat{a}_{oi} \qquad \text{(E.10)}$$

This definition amounts to counting the most efficient firm in the sample as 100 percent efficient. The estimates a_o and u_i are consistent as N and T go to infinity.

Random Effect Model: GLS Estimator

With σ_v^2 and σ_u^2 known, the GLS estimator of a_o* and a of equation (E.7) is consistent,as either N ot T approaches infinity. It is more efficient than the within estimator in the case of $N \rightarrow \infty$ and T fixed, but this difference in efficiencies disappears as $T \rightarrow \infty$. When σ_v^2 and σ_u^2 are not know, their consistent estimates need to be estimated. Consistent estimation of σ_u^2 requires $N \rightarrow \infty$. Thus the strongest case for GLS is when N is large and T is small. If the opposite is true the GLS is useless, unless σ_u^2 were known a **priori**.

Given estimates of $\hat{a}$, we can recover estimates of the individual firm intercepts (a_{oi}) from the residuals, that is, the mean (over time) of the residuals of each individual firms.

$$\hat{a}_{oi} = 1/T\ \Sigma_t\ \hat{e}_{it} \qquad \text{(E.11)}$$

These estimates are consistent as $T \longrightarrow \infty$, provided that estimates of a are consistent. Note that $\hat{a}_{oi}$ can be decomposed into $\hat{a}_o$ and u_i, for which concistency requires $N \longrightarrow \infty$ and consistency of the $\hat{a}_{oi}$. Another way to estimate the individual effects (inefficiency) is by using Battese and Coelli (1986) method, which is actually a generalization of the method suggested by Jondrow *et.al* (1982) as described in equation (3.25). The Battese and Coelli method is presented in a slightly different notation as follows:

$$\hat{u}_i = 1 - \frac{[1 - F(\sigma * - m_i/\sigma *)]}{[1 - F(-m_i/\sigma *)]} \exp(-m_i + \sigma */2) \qquad \text{(E.12)}$$

where

$$\sigma * = \sigma_u^2\ \sigma_v^2\ (\sigma_v^2 + T\ o_u^2)^{-1}$$

$$m_i = -(\sigma_u^2\ \underline{e}_i).(\sigma_u^2 + \sigma_v^2/T)^{-1}$$

$$\underline{e}_i = \hat{a}_{oi} - \underline{u}$$

Note that u and a_{oi} have been described in equation (E.3) and (E.11) respectively, while F is a symbol for standard normal cumulative distribution function (cdf).

The important advantage of GLS estimator relative to within estimator in the present context is not efficiency, but rather the ability to include the time invariant regressors. In cases where time-invariant regressors are relevant, this is important so that their effects do not contaminate measured efficiency.

Choice Between Fixed or Random Effect Model

The choice between these two has nothing to do with the frontier model as such. The only problem with the FE framework in the context of a production or cost (profit) frontier is that the firm-specific effects pick up the effect of variable that differ across firms but are invariant over time. These effects are not in any sense a representation of inefficiency. This might be one of the reasons why estimated inefficiencies in the FE models are much greater than in the RE models.

One way to decide whether to use a fixed effects or random effects model is to test the null hypothesis that there is no correlation between the individual effects and the included explanatory variables against the alternative hypothesis that such correlation exists. For this purpose, we can use either Hausman test (1978) or an asymptotically equivalent test suggested by Mundlak (1979). If the null hypothesis holds we use the random-effects model, otherwise we use the fixed-effect model.

DATA SET AND MODEL SELECTION

The Data Set

The data set used in this study was collected by the Agro Economic Survey, as part of the Rural Dynamic Study in the rice production area of the Cimanuk River Basin, West Java, and obtained from the Center for Agro Economic Research, Ministry of Agriculture, Indonesia.

The survey area, which is the rice production area in the Cimanuk river basin, is characterized by irrigated rice farms and an almost uniform agroclimate. It covers six **desa** (village) located in five **kabupaten** (the administrative unit between district and province level), namely: desa Wargabinagun in kabupaten Cirebon, desa Lanjan in kabupaten Indramayu, desa Gunung Wangi and Malausma in kabupaten Majalengka, desa Sukaambit in kabupaten Sumedang and desa Ciwangi in kabupaten Garut.

In 1977, the survey wan conducted twice, that is at the beginning and the end of the year. The first survey gathered information on farming practices in the wet season of 1975/1976 and dry season of 1976. The second survey covered farm household activities in the wet season of 1976/1977. A similar survey was undertaken in 1978 to cover farm management activities in the dry season 1977. The resurvey of 1983 to the same areas and same farmers was conducted with a different emphasis on labor utilization, asset holding, and land tenure arrangements.

The data net generated by the survey is commonly referred to as a panel data set, since the individual farmer was observed over time. This data set will be used in its advantage manner, that is in the framework of panel data analysis discussed in the next chapter. To date, several studies have used these data. However, these studies, Sugianto (1982), Hutabarat (1986) and Gunawan (1987) among others, analyzed these data using separate cross-sectional analysis or by simply pooling the data.

The analysis of this study uses the so called balance design, where individuals are obeserved for the same lengths of time. Using the individual identification number to check and match individual respondents, only **171 respondents** were found to have been continuously recorded for **six seasons** (Table 1). Some respondents, for various reasons, were replaced by the new ones in the next survey. Others were not recorded in a particular planting season since they were absent. All these respondents were excluded from the analysis.

Table 1. Number of respondents in each Sample Village

Desa (Village)	Kabupaten (Regency)	Number of observation
Wargabinangun	Cirebon	19
Lanjan	Indramayu	24
Gunung Wangi	Majalengka	37
Malausma	Majalengka	33
Sukaambit	Sumedang	22
Ciwangi	Garut	36
Total observations		171

Model Specification and Functional Form

The total output per farm, measured in kilograms of rough rice, is the dependent variable, while the total quantity of seed, fertilizer labor and farm size of the corresponding farm household, are the independent variables. By assuming that farmer is maximizing anticipated profit, all of the production inputs can be treated as exogenous variables, and therefore, the distinction of whether a particular input is variable or fixed is irrelevant.

In logarithmic form, the per farm Cobb-Douglas production function to be estimated is specified as follows:

$$\begin{aligned} \ln Y_{it} = {} & \ln a_o + \Sigma_k\, a_k \ln X_{kit} + a_6 DP_{it} + a_7 DV1_{it} + a_8 DV2_{it} \\ & + a_9 DSS + a_{10} DSIZE + a_{11} DR1 + a_{12} DR2 \\ & + a_{13} DR3 + a_{14} DR4 + a_{15} DR5 + v_{it} - u_i \ldots\ldots\ldots\ldots\ldots \quad \text{(E.13)} \end{aligned}$$

where:

i	=	1,2,.....171 subscript for individual observations
t	=	1,2,.....6 subscript for time
k	=	1,2,.....5 subscript for production inputs
v	=	the error component represents random noise, and is assumed to be distributed normally with zero mean and variance of σ_v^2.
u	=	the non-negative error component representing technical inefficiency.
lnY	=	lnKGOUT: total production in the form of rough rice in kilogram.
lnX1	=	lnKGS : the amount of seed (kg)
lnX2	=	lnKGN : the amount of urea (kg)
lnX3	=	lnKGP : the amount of TSP (kg)
lnX4	=	lnLAB : the amount of labor (hours)
lnX5	=	lnHA : area planted with rice (ha)

DP	:	dummy variable of pesticide use, equals 1 if farmer uses pesticides and equals 0 otherwise
DV1	:	dummy HYV variety, equals 1 if HYV, zero otherwise
DV2	:	dummy of Mixed Varienties (MV), equals 1 if mixed varieties are used, zero otherwise.
Note	:	traditional variety (TV) is the control
DSS	:	dummy variable of season, equals 1 if wet season, zero otherwise
DSIZE	:	dummy variable of farm size, equals 1 if farm size greater than 0.5 hectar, zero otherwise
DR1	:	dummy village, equals 1 if desa Lanjan kabupaten Indramayu, zero otherwise
DR2	:	dummy village, equals 1 if desa Gunung Wangi kabupaten Majalengka, zero otherwise
DR3	:	dummy village, equals 1 if desa Malausma kabupaten Majalengka, zero otherwise
DR4	:	dummy village, equals 1 if desa Sukaambit kabupaten Sumedang, zero otherwise
DR5	:	dummy village, aquals if desa Ciwangi kabupaten Garut, zero otherwise
Note	:	Wargabinangun (kabupaten Cirebon) is the control village

Estimation Methods

Three estimators were used in this analysis, namely the ordinary least square (OLS), the dummy variable (within) and the EGLS estimator. Recall that only the last two estimators yield a frontier function, while the OLS, which is intended for comparison purposes, gives the usual non frontier function. The OLS estimator is obtained simply by applying the ordinary least squares method to the pooled data. The dummy variable estimator, hereafter referred to as within estimator, is obtained by applying OLS to the transformed data, that is after transforming the data in terms of deviations from individual means. The EGLS estimator, which can be viewed as a weighted average of the within and the between estimator, is calculated by first transforming the data in terms of deviations from a **fraction** of the individual means and than running OLS on the transformed data.

One way to decide whether to use fixed effects (FE) or random effects (RE) model is to test the null hypothesis that there is no correlation between the individual effects and the included explanatory variables against the alternative hypothesis that such correlation exists. For this purpose, we can use either Hausman test (1978) or an asymptotically equivalent test suggested by Mundlak (1978).

The individual level of technical inefficiency is measured using Battese and Coelli (1986) method as described in (E.12). The methods described in (E.9 and E.10) for FE model and in (E.11) for RE model will also be employed for comparison purposes.

RESULTS AND DISCUSSION

Three statistical tests were performed. The first test is related to the fixed effect (FE) specification to test the null hypothesis that individuals have the same intercept against the alternative hypothesis that their intercepts are not the same (see Judge *et.al.,* 1982). The computed F-statistic equal 1.4818 and the critical $F_{0.05}(170, 845)$ equals 1.2214. Thus, the null hypothesis is rejected at the 0.05 significance level.

The second test is the LM test, related to random effect (RE) specification to test the presence of individual random effects by testing the null hypothesis that σ_u^2 equals zero (see Judge *et.al.,* 1982). The computed LM statistic equals 9.5864 and the critical $Chi^2(1)_{0.05}$ equals 3.8415. Since the LM statistic is larger than its critical value, we can therefore reject the null hypothesis. This implies that individual random effects exist.

The thrid test was perfomed on the null hypothesis that there is no correlation between the individual effects and the included variables against the alternative hypothesis that such correlation exists. The results determine whether to use RE or FE specification. The Mundlak test (1978) was used in this case. The computed F-statistic is 1.5378 while the critical $F_{0.05}$ (8, 1009) is 1.9384. Thus the RE model rather than the FE model is justified statistically. Note that in performing the Mundlak test, only the time-varying variables which are statistically significant were included.

In order to examine the difference between OLS, within and GLS estimators, we present the estimation results in appendix 1. The individual intercepts of the within estimator are not presented here. The results show that the parameter estimates of the GLS lie between the corresponding parameter estimates of the other two estimators. The GLS estimator gives the best fit of the production function compared to the others. The computed coefficient of multiple determination (R^2) of the GLS estimator equals 0.9967, which is higher compared to 0.8843 of the OLS and 0.7497 of the within estimator.

The sign of the cofficients estimated by OLS and GLS is the same and the magnitudes are very close to each other. This is not too surprising, since the $(1 - w)$ is very small, and therefore running LS regression on the transformed data (GLS estimator) yields similar results to running LS regression on the original data (OLS estimator). Note that "w" measure the weight given to the between-individual variation. In the covariance (within) estimator, this variation is completely ignored (w = 0), while in the OLS estimator this variation is completely incorporated (w = 1).

The results also show little difference in magnitude between the GLS estimator and corresponding estimate of the within estimator, and no difference in the sign of the estimate. This is consistent with the result of the model specification test previously described. If the null hypothesis the E{X'u} = 0 holds, and in fact it is not rejected by the test, the GLS estimator should not be very different from the within estimator. The significant difference between these two estimators indicates that the alternative hypothesis holds. This is the basic idea underlying the Huasman test, or equivalently the Mundlak test.

Production Elasticity

This following discusion will focus only on the GLS estimates. However, some comparisons to the within estimates might be made. Before, proceeding to production elasticities, let us first interpret the coefficient estimates of the dummy variables. The dummy variable for pesticide use is not significant, indicating the use of pesticide does not have any effect on the level of production. It was reported that during survey periods no significant crop damage due to insect attack or plant diseases occured in the study area. The dummy variables of HYV and MV are significant at the 0.01 level. Thus, farmers with HYV and MV produce more output than TV farmers. This is consistent with a **priori** expectation.

The season dummy has a positive sign and is significant at the 0.05 significance level. This indicates that the level of production is greater in the wet season than in the dry season. This is understandable because the soil moisture content in the wet season is usually more optimal for plan growth than in the dry season. Thus, the lack of water during the dry season is possibly the key seasonal yield difference. The difference in the level of production is represented directly by the coefficient for the season dummy.

The regression coefficient for the farm size dummy is not significant, indicating that there is no significant difference in productivity between small and large farmers. The region dummies representing individual non-specific time invariant variables such as climate and soil quality, are not significantly different from zero. This result indicates that there is no signifant difference, statistically, in the level of production between regions. The natio-wide rice intensification program (BIMAS/INMAS), which has already been implemented intensively since the early 1960's, particularly in West Java, could be the main reason.

The interpretation of a Cobb-Douglas production function is very simple and straightforward, since the regression coefficients directly represent the production elasticities or the corresponding independent variables (inputs). Appendix 1 shows that all input coefficients have correct signs and are significantly different from zero at the 0.01 level. The seed coefficient of 0.1300, indicates that a one percent increase in quantity of seed, other things being fixed (*ceteris paribus*), will result in 0.13 percent increase in the level of rice production.

The production elasticities with respect to urea and TSP fertilizer equal 0.1110 and 0.0778, respectively, and both are significant at the 0.001 level. The result shows that rice is more responsive to N fertilizer than P fertilizer. The slow decomposition of phosphorus relative to nitrogen in the soil could be the reason for this difference. Thus the paddy fields may have sufficient phosphorus, but insufficient nitrogen, resulting in a relatively smaller yield response from additional phosphorus compared to yield response from additional nitrogen.

The production elasticity with respect to labor is 0.22. This means that a one percent increase in labor hours will increase the production level by 0.22 percent. Similarly, the interpretation of the production elasticity with respect to land is that a one percent increase in area cultivated per household will result in 0.47 percent of increase in the level of production, indicating the condition of diminishing marginal returns to land.

Production Efficiency

Let us now turn to technical inefficiency measures. Individual level of technical inefficiency is estimated using three different methods. The first method obtains individual technical inefficiency by differencing the individual intercepts from the intercept of the most efficient farm, as in equation (E.9 and E.10). The second method is based on the residual of the EGLS estimator as presented in equation (E.11). The third method is the one suggested by Battese and Coelli (1985) as presented in equation (E.12). Note that these methods, particularly for the first two, give consistent estimates of individual effects (u_i) only if both N and T are large. Since in our case T is relatively small, the consistency of the estimates of u_i is questionable. The first two methods do not use any distributional assumptions regarding u_i, while the third method uses a half-normal distribution with mode at $u_i = 0$.

The first two methods give much larger estimates than the third one. In the case of the first method, the estimate of u_i may always be larger than the others since it includes the effects of time invariant variables which could not be included in the model. Given the obvious disadvantage of the first two methods, their estimation results are not discussed. However, it is important to note that although the magnitudes of these three estimators are quite different, the individual ranks based on the level of efficiency are quite similar.

Estimated technical (in) efficiency of individual farms is not presented in this paper. The range of technical inefficiency using the third method (E.12) is 3.4 percent – 12 percent, with the mean of 6.5 percent. With mean is 7 percent if the calculation is based on equation (E.3). This figure tells us that rice farms in West Java are, on the average, 6.5 percent technically inefficient or 93.5 percent technically efficient. One could also interpret that the fitted individual production function is 3.4 percent – 12 percent, or 6.5 percent on the average, below the frontier production function. Given the intensive nature of wetland rice farming in West Java (since rice production technology has long been and widely implemented in this region) this relatively small figure of technical inefficiency is reasonable. The figure, presumably would be much greater in the case of dryland rice farming since the government has given relatively less attention to intensifying dryland farming.

Policy implications of the efficiency measure are debatable. Some advocates to frontier analysis claim that a firm can move from the interior of the production function surface to the frontier without any cost to the firm. They assert that better use of the existing technology in terms of cultivation and crop-management practices will definitely increase yield. They do not, however, specifically address the question of how this can be achieved. On the other hard, nonadvocates would argue that free correction is very unrealistic, since movement to frontier requires adjustments of factors of production including management skills which could be regarded as a fixed factor. Improving management skills is of course not without cost.

Table 2 shows that no significant difference in the level technical inefficiency between small farms and large farms; small farms may or may not be more technically efficient than large farms or vise versa. This is consistent with the regression results which yield a coefficient on the farm size dummy variable not significantly different from zero. Therefore, it seems reasonable to interpret the mean of technical inefficiency level (6.5%), as a per hectare output loss due to technical inefficiency. Alternatively, we can simply reestimate a per hectare production frontier and find both the mean and the individual level of technical inefficiency.

Assuming that the estimated frontier production represents the existing rice production technology in West Java, one could then roughly estimate the total annual rice production losses due to technical inefficiency in this region. Note, however, that this is very rough estimate and is solely intended as an illustration purposes. Given rice yield figure (4000 kg/ha) and the annual harvested area of rice farms in West Java (1.74 million hectares) in 1983, we get the estimated figures of 260 kg per hectare and 0.45 million ton annually of rice production losses. Thus, better use of existing technology of rice production provides an opportunity to somewhat increase rice yield and total rice production in West Java.

Table 2. Frequency distribution of farmers based on the level of technical inefficiency from Cobb-Douglas production frontier

Range of technical inefficiency	% farms		% from total frame
	≤ 0.5 ha	> 0.5 ha	
≤ 5%	13.4	13.6	13.5
5 % < u ≤ 10%	83.5	84.1	83.6
10% < u ≤ 15%	3.1	2.3	2.9
Total %	100	100	100
Total farms	127	44	171

Let us now evaluate allocative efficiency of the input use. A farm is allocatively efficient if the input use maximizes profit, that is if the value of marginal product (VMP) of particular input equals its marginal factor cost (MFC). This condition implies that at the point of profit maximization, the ratio (d) of VMP to MFC for each input is equal to one. This also means that the last dollar spent on each input must return exactly one dollar, and most if not all previous units will have given back more than a dollar (Debertin, 1986). The accumulation of the excess dollars in returns over costs represents the profits or net revenues accruing to the farm.

A simple evaluation for allocative inefficiency can be conducted by calculating the "d" ratio for individual farm based on the estimated production function and the price levels reported in the survey. There is no intention to interpret the level of allocative (in) efficiencies for an individual farm. Seasonal average and grand average values of the "d" ratio for each input are presented in Table 3. On the (grand) average the "d" ratio for seed, urea, TSP, and labor are 10.34, 3.09, 14.60, and 0.89 respectively, indicating underutilization or seed and fertilizers and overutilization of labor. This, however, does not necessarily mean that farmers do not attempt to maximize profits, rather it may mean that farmers, for various reasons, were not able to maximize profit.

Table 3. Allocative inefficiency measure ("d" ratio) in rice production in West Java.

Season	Seed	Urea	Phosphate	Labor
W-75/76	12.04	3.74	16.74	1.06
D-76	11.45	3.79	19.38	0.93
W-76/77	8.58	2.46	17.00	0.78
D-77	7.39	2.97	16.11	0.74
W-82/83	10.67	2.58	6.03	0.85
D-83	11.92	3.04	12.37	1.03
Average	10.34	3.09	14.60	0.89

CONCLUSION

This paper has demonstrated the application of the stochastic production frontier model for estimating technical efficiency on the rice farm in West Java. This paper also described the potential advantages for modifying existing frontier models to allow the use of panel data. Some major findings and conlusion are as follows:

1. There is no statistically significant difference in the level of production between regions. This is not surprising given the fact that Bimas/Inmas program have been intensively implemented in nearly all of the West Java province, particularly in the regions covered in the survey which are major rice production areas.
2. The dummy variable for farm size (i.e. small or large farm category), is not significantly different from zero, indicating that there is no significant difference in the productivity level between small and large farms. In other words, small farms may not be more productive than large farms.
3. The dummy variable for season has a positive sign and is statistically significant at 0.05 significance level, indicating that the level of production for wet season is greater than that for the dry season. The lack of water availability during the dry season due to in appropriate irrigation facilities is the key reason for this seasonal yield difference, as reported in the survey.
4. The dummy variables for varieties, HYV and MV, are statistically significant at 0.01 significant level, with positive signs. This strongly indicates that farms with HYV and MV produce more output that TV farms, which in consistent with a **priori** expectations.
5. The production elasticities with respect to seed, labor, land urea and phosphate fertilizer are 0.1304, 0.2211, 0.4676, 0.1110 and 0.0778, respectivelly, and are statistically at 0.01 significance level. Thus, a one percent increase in the amount of each ot these inputs, *ceteris paribus,* will increase the level of production by that percentage amount, respectively. The production elasticity with repect to urea fertilizer is greater than the elasticity with respect to phosphate fertilizer. This finding could be used to support the argument for differentiating the prices of these two fertilizers, if the government's primary concern is to gradually reduce fertilizer susidies.
6. The range of individual technical inefficiency is 3.4 percent – 12 percent with the mean 6.5 percent. These figures simply tell us that the rice farms in West Java are, on the average, 6.5 percent technically inefficient or 93.5 percent technically efficient. Using rice yield for the 1983 dry season (4000 kg/hectare) and the figure of total annual harvested area in West Java (1.74 million hectare), the estimate of yield loss was 260 kg per hectare, and the total quantity of production loss would be about 0.45 million tons annually.

7. This study confirms that farmers in the study area are not able to optimally allocate the production inputs. There is a tendency that farmers underutilize both seed and fertilizers, but overutilize labor.

BIBLIOGRAPHY

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.

Battese, G.E. and T.J. Coelli. 1988. Prediction of Firm-Level Technical Efficiencies With A Generalized Frontier Production Function and Panel Data. Journal of Econometrics, 38: 387-399.

Erwidodo, 1990. Panel Data Analysis on Farm Level Efficiency, Input Demand and Output Supply of Rice Farm in West Java, Indonesia. Unpublished PhD thesis, Michigan State University, East Lansing.

Forsund, F.R., C.A.K. Lovell and P.Schmidt. 1980. A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurements. Journal of Econometrics 13: 5-25.

Gunawan, M. 1988. Adoption and Bias of New Agricultural Innovation in Jawa Barat, Indonesia. Unpublished Ph.D Thesis, Univ. of Minnesota, St. Paul, Minnesota.

Hausman, J.A. 1978. Specification Test in Econometrics. Econometrica, 46: 1251-1272.

Hoch, I. 1962. Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data. Econometrica, Vol.30, No.1: 34-53.

Hutabarat, B. 1985. An Assessment of Farm-Level Input Demands and Production Under Risk on Farms in Cimanuk River Basin, Jawa Barat, Indonesia. Unpublished Ph.D thesis, Iowa State Univ. Ames, Iowa.

Jondrow, J, C.A.K. Lovell, I.S. Materov, and P.Schmidt. 1982. On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. Journal of Econometrics, 19: 233-238.

Judge, G.E, R.C. Hill, W.E. Griffiths, H. Lutkepohl, and T-C Lee. 1982. Introduction to the Theory and Practice of Econometrics. John Wiley and Sons, N.Y, USA.

Meeusen, W. and J van den Broeck. 1977. Efficiency Estimation From Cobb-Douglas Production Functions With Composed Error. International Economic Review, Vol.18, No.2: 435-444.

Mundlak, Y. 1978. On the Pooling of Time Series and Cross Section Data. Econometrica, 46: 69-85.

Schmidt, P. and R.C. Sickles. 1984. Production Frontiers and Panel Data. Journal of Business and Econometrics, Vol.2, No.4: 367-374.

Seale, J.L. 1985. Fixed Effect Cobb-Douglas Production Functions For Floor Tile Firms, Fayoum and Kalyubiya, Egypt, 1981-1983. Unpublished Dissertation, Michigan State University.

Sugianto, T. 1982. The Ralative Economic Efficiency of Irrigation Rice Farms, West Java, Indonesia. Unpublished Ph.D thesis, Univ. of Illinois, Urbana- Champaign, Illinois.

Appendix 1. Estimated parameters of the Cobb-Douglass production function

Independent variables	Estimation methods		
	OLS	Within	GLS
Constant	5.0868***	-	5.0690***
	(0.1916)		(0.1938)
LKGS	0.1339***	0.1176***	0.1304***
	(0.0271)	(0.0271)	(0.0271)
LKGN	0.1175***	0.0878***	0.1110***
	(0.0175)	(0.0193)	(0.0179)
LKGP	0.0735***	0.0912***	0.0778***
	(0.0114)	(0.0116)	(0.0115)
LLAB	0.2159***	0.2378***	0.2211***
	(0.0288)	(0.0296)	(0.0290)
LHA	0.4759***	0.4323***	0.4676***
	(0.0318)	(0.0333)	(0.0321)
DP	0.0066	0.0325	0.0127
	(0.0284)	(0.0293)	(0.0286)
DV1	0.1743***	0.1768***	0.1756***
	(0.0385)	(0.0377)	(0.0383)
DV2	0.1389***	0.1792***	0.1477***
	(0.0541)	(0.0531)	(0.0539)
DSIZE	0.0198	0.0881**	0.0349
	(0.0359)	(0.0400)	(0.0368)
DSS	0.0496**	0.0555***	0.0503**
	(0.0218)	(0.0196)	(0.0211)
DR1	-0.0505	-	-0.0519
	(0.0435)		(0.0499)
DR2	-0.0403	-	-0.0465
	(0.0546)		(0.0591)
DR3	-0.0640	-	-0.0736
	(0.0575)		(0.0621)
DR4	0.0240	- 0.0118	
	(0.0527)		(0.0585)
DR5	0.0801	-	0.0734
	(0.0557)		(0.0602)

Appendix 1. (continued)

Independent variables	Estimation methods		
	OLS	Within	GLS
σ_v^2	-	0.1069	0.1069
σ^2	-	-	0.1526
σ_u^2	-	-	0.0075
$w = \sigma_v / \sigma$	-	-	0.8372
$\theta = 1 - w$	-	-	0.1628
E(u)	-	-	0.071
F-Statistic	514.70	304.34	19208.10
R^2	0.8843	0.7497	0.9967
N (individuals)	171	171	171
T (seasons)	6	6	6

Variable definition can be seen in chapter 4 or appendix 5.4
Figure in parentheses are standard deviations
*** statistically significant at $\alpha = 0.01$
** statistically significant at $\alpha = 0.05$
* statistically significant at $\alpha \doteq 0.10$