CHAPTER 4

RESEARCH METHODS

This chapter provides the scope of the study, data collection, and methods of data analysis. Especially, the concept of stochastic frontier method is presented in detail.

4.1 Scope of the study

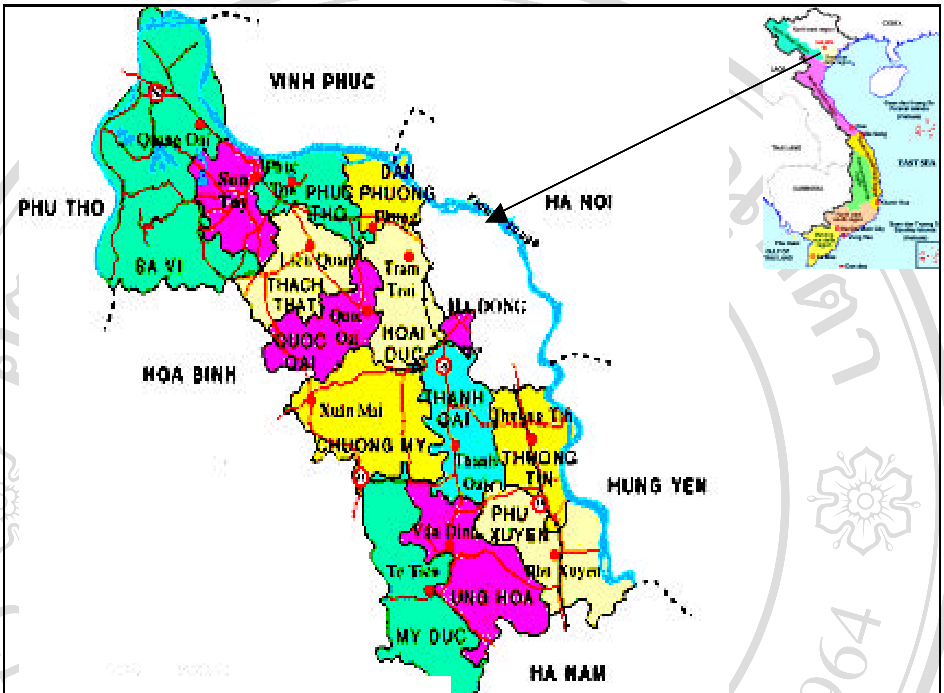
As mentioned in chapter 2, the Red River Delta and Mekong River Delta are two biggest-rice producing regions in Vietnam. Owing to time and budgeting are limited, hence this study only conducted in one province, belonging to the Red River Delta. The criteria for site selection was based on the dominance of crop, i.e., rice, both conventional and hybrid rice varieties. are grown.

In addition, the primary data were cross sectional data for the crop seasons of the year 2002.

4.2 Sampling technique

The multi-stage sampling method was administered in order to get representative research site. First, Hatay province, one province of the Red River Delta, were sellected as for research site. Sellecting of Hatay province based on its contribution to rice production to the Red River Delta. After that, two districts, Quocoai and Phuxuyen belonging to Hatay province were selected. Phuxuyen located in the south while whereas Quocoai located in the north of Hatay province. The distance between two districts is about 50 km.

Second, a random sampling method was used to select the rice households belonging to two districts. The sample size includes 50 rice farms in Quocoai district and 50 rice farms in Phuxuyen district.



Source: *http://www.mpi-oda.gov.vn*

Figure 4.2: Map of Hatay province

4.3 Data collection

4.3.1 Secondary data

In order to get in-depth understanding of the performance of rice production system in Hatay province, a number of relevant publications were reviewed. Studies on rice farming, annual progress reports, and biophysical, socio-economic, and demographic characteristics of the study site wre also collected from various sources. Some of the sources could be listed as follows:

- The statistical yearbooks

- Hatay Agricultural Office

- District Agricultural Offices
- District Extension Offices

4.3.2 Primary data

To get sufficient and accurate data, before carrying out an actual survey, questionnaires were constructed after consultion with local leaders and experienced persons. Then, preliminary testing of questionnaires was conducted in 10 farm households and necessary changes were made.

This study used cross sectional data to be obtaining from field survey. Primary data is eregathered in this study was by interviewing farmers with questionnaire. The questionnaire includes biophysical and socio - economic information. In addition to this, heads of agricultural cooperatives and village provided the general information on the study site. Production data include cropping partern, cultivated area, rice-sown are, rice varieties, rice yields, input utilization, and farm management. Socio-economic data consist of land type, land distribution, farm size, household head information, family size and age, education status, on-farm and off-farm income, labor used on rice farming, and input and output market information.

Using the method of grading extension contact of Song (1997), In the study, extension contact is grading by interviewers and agricultural leaders of the village bases on the number of household member attendanded or paticipated in extension class, mass meeting, combining with level of dealing with learning agricultural transfer technologies in television program, radio, news paper.

Moreover, the extension score of each household also is obtained by answering question about knowledge of fertilizer, pesticide use. Finally, extension score is evaluated for individual household head.

4.4 Data analysis

The study used descriptive analysis and quantitative method to analyze the data.

4.4.1 Descriptive analysis

From survey and secondary data, the descriptive statistical analysis was applied to analyze the rice production s and somepoints of marketing aspect. In addition, budgeting analysis was also address to reflect the profitability of rice production. The budgeting analysis of production in cost and return includes on set of term as follows:

Gross return	= Yield in kg * Price per kg
Net return	= Gross return - Total cost
Material cost	= Cost of seed (own supplied and purchased)
	+ Chemical fertilizer cost
	+ Pesticide cost
	+ Manure cost
Labor cost	= Hired labor cost + Family labor cost
Service fee and land tax	= Land preparation fee
	+ Irrigation fee
	+ Field protection fee
	+ Land tax
Total cost	= Material input cost
	+ Labor cost
	+ Service fee and land tax
Net return	= Gross return- Total cost

Return to family labor = Gross return - All costs except family labor cost

Cost of 1 kg output = Total cost / Total output

In addition, return ratios were calculated, namely: Gross return per Total cost, Net return per Total cost, and Net return per one kilogram output also. A farm- gate of input was used to calculate Gross return. The currency is Vietnam Dong (VND). 1 US$ was approximately 15,300 VND, at the time of survey (October, 2002).

4.4.2 Quantitative method

Rice production the combination of many inputs to produce one homogenous output, i.e, rice and it is affected by many factors including random factor that cannot control by producers and factors affecting by farmers. It is necessary to evaluate the efluence of each affecting factor on rice output. This study took advantages of the stochastic production frontier method in order to estimate rice frontier production function, technical efficiency indies, and factor affecting technical efficiency.

Technical efficiency can be estimated by employing the stochastic frontier with its stochastic frontier production function is defined as:

$$Y_i = f(x_i; b)\exp(v_i - u_i) \qquad i=1,2,\ldots,n \qquad (4.1)$$

The frontier production function is represented by $f(x_i; b)$ and is a measure of a maximum potential output for any particular input vector x_i. Both v and u cause actual production to deviate from this frontier. The random variability in production cannot be influenced by producers represented by " v_i" (e.g environmental factors such as temperature and moisture, etc.). It is identically and independently distributed as $N(0, s_v^2)$ and may be considered as the "normal" error term. In addition, the independence of the error term "u_i" were assumed to be non-negative truncations of the $N(0, s_u^2)$ distribution (or half-normal distribution). The term u is the one-sided error. This implies that each observation is on or below the frontier. "$-u$" is called "*technical in-efficiency*" (Maddala, 1983). The density function for u_i is defined by:

$$f(u) = \frac{2}{\sigma_u (2\pi)^{\frac{1}{2}}} \exp\left(\frac{-u^2}{2\sigma_u^2}\right) \quad (u \geq 0)$$

(4.2)

(Maddala,

1983

.p1995)

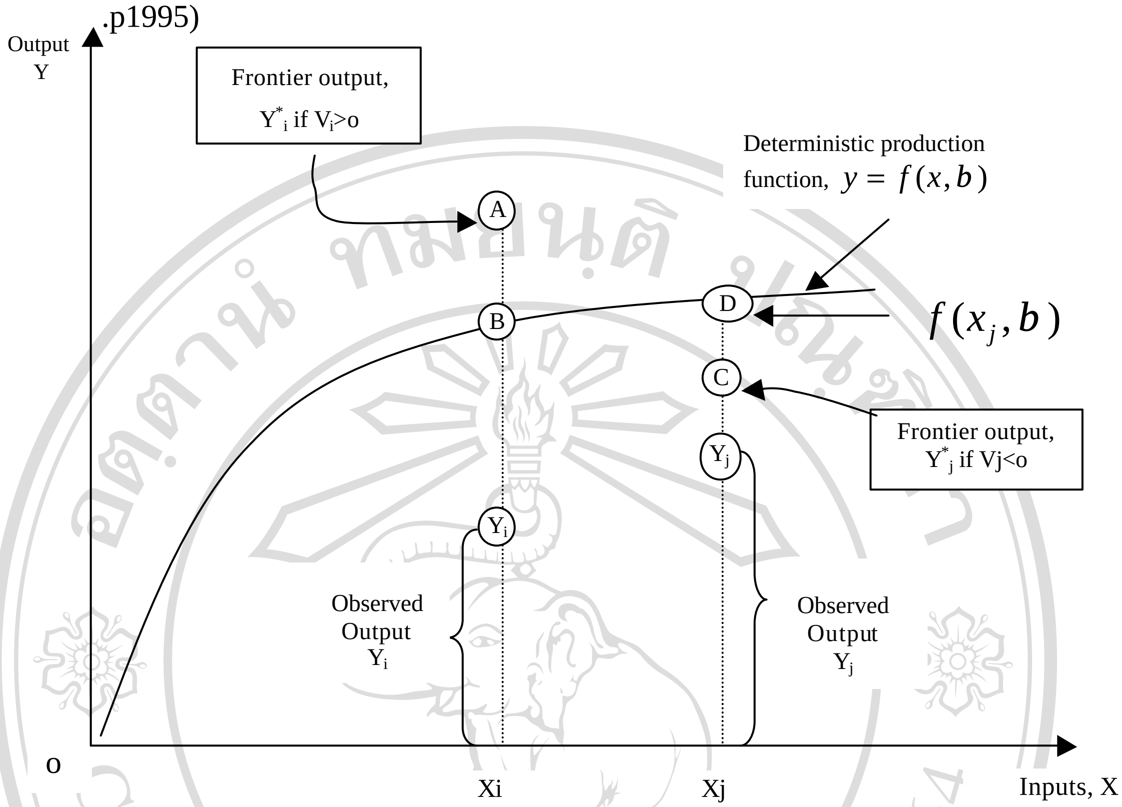


Figure 4.3: Stochastic frontier production function (Battese, 1992)

Note: For firm i

AB = v (random component of favorable conditions)

AY_i = u (inefficiency component)

BY_i = v-u (total deviation from the frontier)

For firm j

CD = v (random component of favorable conditions)

CY_j = u (inefficiency component).

DY_j = v-u (total deviation from the frontier)

The basis structure of the stochastic frontier model 4.1 is depicted in Figure 4.3 in which the activities of two firms, represented by i and j, are considered. Firm i uses inputs with values given by the vector x_i and obtains the input, Y_i, but the frontier output, Y_i^*, exceeds the value on the deterministic production function, $f(x_i;b)$, because its productive is associated with "favorable conditions" for which the random error, V_i is "positive". However, firm j uses inputs with values given by the vector x_j and obtains the output, Y_j, which has corresponding frontier output, Y_j^*, which is less than the value on the deterministic production function $f(x_i;b)$, because its productive activity is associated with "unfavorable conditions" for with the random error, V_j is "negative". In both cases, the observed production values are less than the corresponding frontier values, but the unobservable frontier production values would lie around the deterministic production function associated with the firms involved.

If u and v are distributed independently, and according to Weinstein (1964)

$$f(\varepsilon)=\frac{2}{\sigma}f^*\left(\frac{\varepsilon}{\sigma_e}\right)\left(1-F^*\left(\frac{\varepsilon\lambda}{\sigma_e}\right)\right) \qquad -\infty \le \varepsilon \le +\infty \tag{4.3}$$

where

$\varepsilon = v - u$ is defined as the total deviation from the frontier

$\sigma_e^2 = \sigma_u^2 + \sigma_v^2$ is defined as the total variation of output from frontier which can be attributed to technical efficiency.

$\lambda = \frac{\sigma_u}{\sigma_v}$ is ratio of the two standard error.

f^* (.) and F*(.) are the standard normal density and distribution functions, respectively. This density is asymmetric around zero, with its means and variance given by

$$E(\varepsilon)=E(u)=-\frac{\sqrt{2}}{\sqrt{\pi}}\sigma_u \qquad (4.4)$$

$$V_e = V_v + V_u \qquad (4.5)$$

$$= \left(\frac{\pi-2}{\pi}\right)\sigma_u^2 + \sigma_v^2 \qquad (4.6)$$

The technical efficiency of given firm is defined to be the factor by which the level of production for the firm is less than its frontier output. Given of the stochastic frontier production function 4.1, the technical efficiency for the i th firm is formulated as follows:

$$TE_i = \frac{Y_i}{Y^*} \qquad (4.7)$$

$$TE_i = \frac{f(x_i;\beta)\exp(v_i-u_i)}{f(x_i;\beta)\exp(v_i)} = \exp(-u_i) \qquad (4.8)$$

$$0 \le TE_i \le 1 \qquad \text{(Battese, 1992)}$$

It is considered that lower TE index value represent less efficiency production (or a greater degree of inefficiency). If a firm's technical efficiency is 0.85, it implies that the firm realizes, on the average, 85 percent of the production possible for a fully efficient firm having comparable input values (Battese and Coelli, 1988).

The estimate of production is based on the most efficient observed use of inputs to produce each level of output. The extent to which farm production differs from the frontier provides a measure of technical inefficiency for the sample as a whole or for each firm individually. The causes of technical inefficiency can be investigated by regressing inefficiency on explanatory variables (Ali and Byerlee, 1991). Since u_i's are non-negative random variables, which assumed be independently distributed such as u_{1i} is defined by the truncation (at zero) of normal distributions with mean μ_i and variance σ_u^2 respectively. Seyoum *et al.* (1998) defined each μ_i as a function of some explanatory variables

$$m_i = q_0 + q_1 \mathrm{F}_{1i} + \ldots + q_m \mathrm{F}_{mi} \qquad (4.9)$$

where $F_1, \ldots F_m$ are explanatory variables

The maximum-likelihood estimates for all parameters of the stochastic frontier and inefficiency equation defined by equation 4.1 and 4.9 are simultaneously obtained by using the FRONTIER 4.1 software (Coelli, 1996), which estimates the variace parameters in terms of parameterization.

$$s_e^2 = s_u^2 + s_v^2 \text{ and} \qquad (4.10)$$

$$g = \frac{s_u^2}{s_e^2} \qquad (4.11)$$

Furthermore, the Cobb-Douglas function form was employed as the form of production function in this study. Following are main reasons to explain the choice.Firstly, agricultural input-output relationship usually follows the law of diminishing return. Cobb-Douglas production satisfies this law.Secondly, Cobb-Douglas function is simple and easy to estimation and interpretation.Thirdly, values of variable in the model satisfyd the conditins of the Cobb-Douglas production function (thay must be greater than zero). Finally, the Cobb-Douglas function is the most commonly used functional form to estimate the production frontier.

Apart from of those, hetroscedasticity were tested by White's test and multicollinearity was also detected by using correlationrilation matrix (Gujarati, 1995) with supporting of Limdep 7.0 software. The specifications of empirical models were presented in Chapter 7.