Chapter Three

MODELS METHODOLOGY AND DATA

This chapter deals with the methodology adopted for the present study. The analytical foundations of measurement of economic efficiency, formulation of the stochastic production frontier and sampling design and collection of primary data are expressed in detail. The analytical microeconomic foundations of measurement of technical efficiency are explained in the first sub-section, that is, section 3.1. Concepts of production frontier, input and output based measures of technical efficiency are elaborated in this section using standard microeconomic tools. Some elementary real analysis is used in line with graduate level standard microeconomic texts. The second subsection, that is, section 3.2 is dedicated to the concept of the stochastic production frontier and its conceptual development since its inception. The next three sections 3.3. 3.4. and 3.5 are devoted respectively to the basic stochastic production frontier model of Aigner et al., (1977) and the specific econometric strategy adopted in the present study. All variables selected and constructed are first defined. Next, the rationale behind each variable construction, i.e., the measurement methods for each variable selected for stochastic frontier analysis is elaborated in section 3.6. Finally the sample selection and data collection methodologies are narrated. Data for the present research is entirely primary in nature. The details of how the sample offishing teams was selected from the study area are described in section 3.7. The details of interview with the catchers for data collection are finally described.

3.1 Analytical Foundations of Measurement of Economic Efficiency

Prior to describing the econometric strategy adopted to measure firm level economic efficiency or inefficiency it is essential to present the conceptual (microeconomic) framework of economic efficiency. We begin with the physical structure of different production technologies with the help of graphical analysis. A production technology that transferring inputs $x=(x_1,x_2,.......,x_N)\in R_+^N=\{x:x\in R_+^N,x\geq 0\}$ into output $y=(y_1,y_2,....,y_M)\in R_+^N$ can be represented by the output correspondence *L* or the graph of the technology GR. The output correspondence *P*: $R_+^N\rightarrow 2^{R_+^N}$ $[2^{R_+^N}\Rightarrow\{A:A\subseteq R^M\}]$ (*A* is a subset of the Euclidian space of dimension *M*) maps input $x\in R_+^N$ into subset $P(x)\subseteq R_+^N$ of output. The set $P(x)$ is called output set and it denotes the collection of all output vectors $y\in R_+^M$ that are obtainable from the input vector $x\in R_+^N$ (Neogi, 2005). The input correspondence $L:R_+^M\rightarrow 2^{R_+^N}$ maps the output $y\in R_+^M$ into subset $L(y)\subseteq R_+^N$ of inputs.

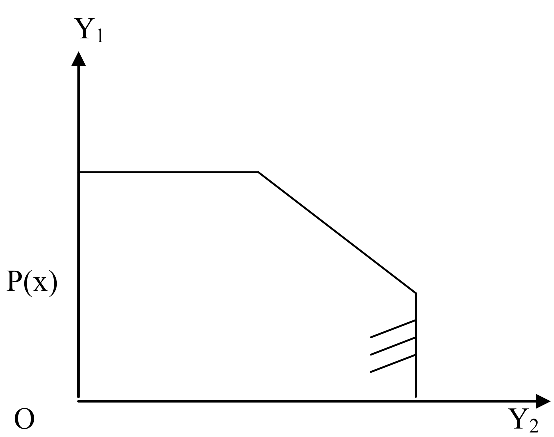

Figure 3.2.1

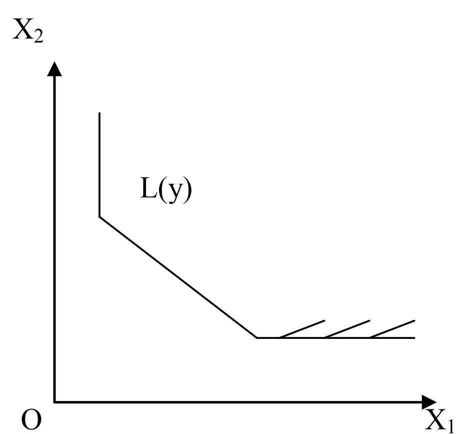

Figure 3.2.2

The input set $L(y)$ denotes the collection of all input vectors $x \in R_{+}^{N}$ that yields at least output vector $y \in R_{+}^{M}$. The input and output correspondence can be derived from one another by means of the relationships $L(y) = \{x : y \in P(x)\}$ and $P(x) = \{y : x \in L(y)\}$.

Now the graph of the production technology is the collection of all feasible input-output vectors, i.e.

$$GR = \{(x, y) \in R_{+}^{N+M} : y \in P(x), x \in R_{+}^{N}\} \text{ and}$$

$$GR = \{(x, y) \in R_{+}^{N+M} : x \in L(y), y \in R_{+}^{M}\} .$$

Thus the input-output correspondence can be derived from the graphs above (figure 3.2.1 and 3.2.2) as, $P(x) = \{y : (x, y) \in GR\}$ and

$$L(y) = \{x : (x, y) \in GR\} .$$

Figure 3.2.3 models both inputs and output substitution in addition to modeling input-output transformation. The input set, the output set and the graph represents the technology in terms of input quantities and output quantities.

Let us introduce the concept of production frontier as a functional characteristic of the boundary of the graph of the production technology. The boundary of the graph represents the maximum possible output obtained from a given level of input or minimum input use for a given level of output. In a single output-multiple input case the production frontier can be defined as

$$\begin{aligned} f(x) &= \max \{ y : y \in P(x)\} \\ &= \max \{ y : x \in L(y)\} \end{aligned}$$

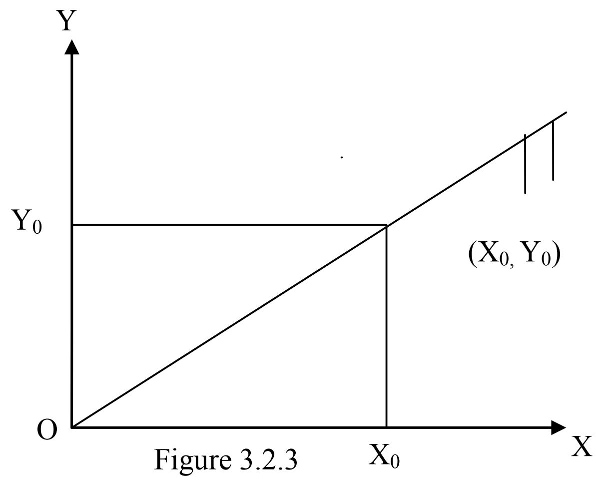

Figure 3.2.3

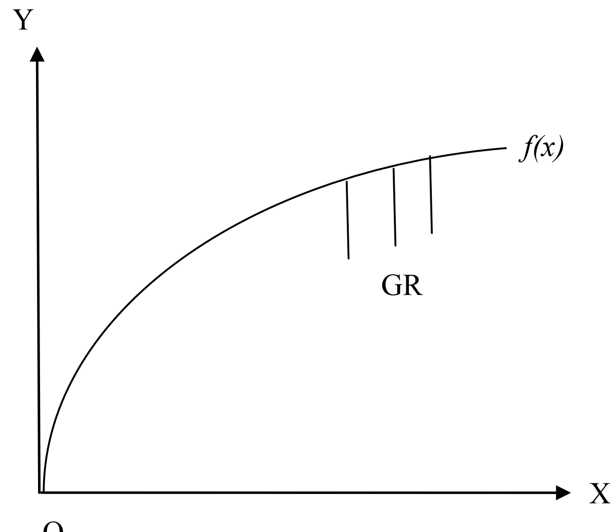

Figure 3.2.4

In figure 3.2.4 the production function *f*(*x*) describes the maximum output that can be obtained with any given input vector. The different combinations of inputs and outputs fall on or below the production frontier. The basic idea of efficiency is to measure the distance of a particular combination of input and output of a production unit from the respective production frontier (Neogi, 2005).

There are two concepts of a distance function. The input distance function measures the maximum possible conservation of input to reach the boundary of production frontier. The concept can be illustrated in Figure 3.2.5 [(a) and (b)].

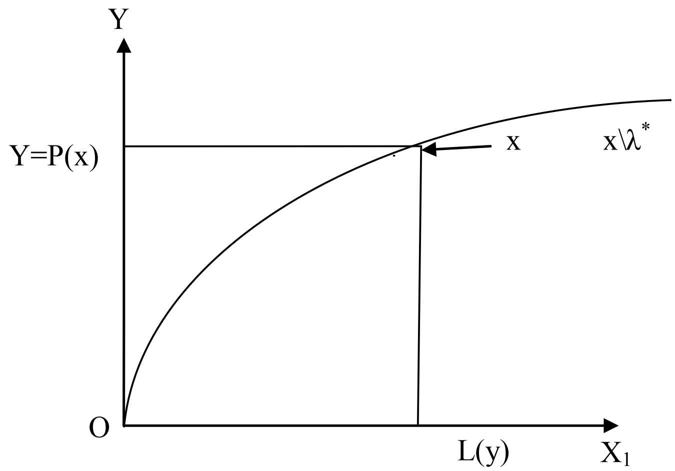

Figure 3.2.5 (a)

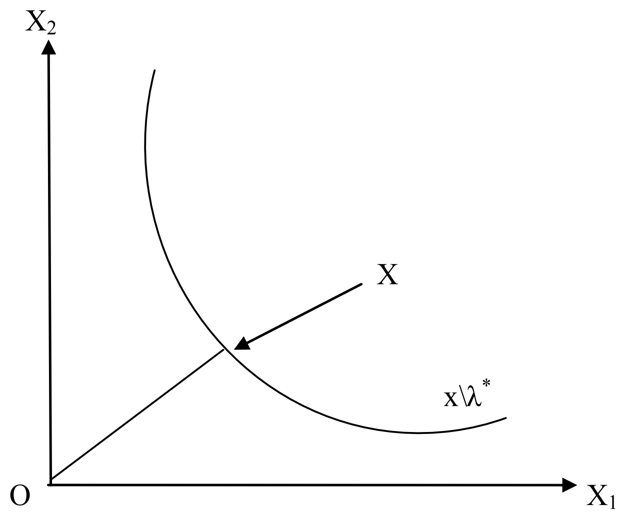

Figure 3.2.5 (b)

An input distance function can be defined as $D_i(y,x) = \max\{\lambda : x/\lambda \in L(y)\}$ where λ is the contraction factor by which inputs can be reduced to produce output y. Figure 3.2.5 is the graphical representation of the input distance function.

An output distance function can be defined as $D_0(y,x) = \min\{\mu : y/\mu \in P(x)\}$ where μ is the output expanding factor i.e., the proportion in which output can be maximized (with $\mu < 1$) with a given level of input. A graphical representation of the output distance function is given in figure 3.2.6 [(a) and (b)].

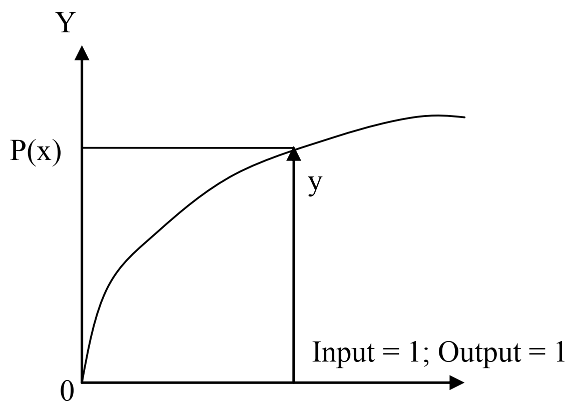

Figure 3.2.6 (a)

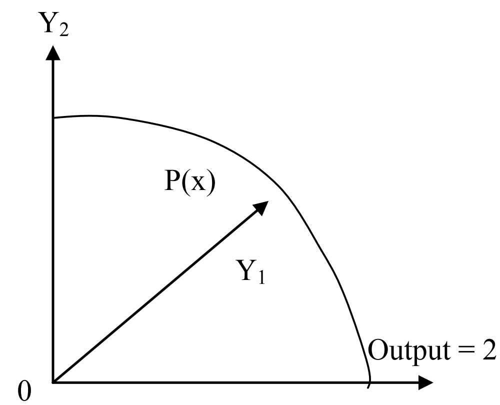

Figure 3.2.6 (b)

Efficiency measurement is based on the estimates of the best practice frontier production function which is a natural reference or basis of efficiency measure. Efficiency may be of three types: (i) technical, (ii) economic and (iii) scale.

Efficiency measure provides a description of the structure of an industry and is hence a very important step for identifying the causes of inefficiencies. Figure 3.2.7 (a) describes the concept of feasible production set which is the set of all input-output combinations which are feasible. The set consists of all points between the production frontier and the *X*- axis. The points along the production frontier line *OP* define the efficient subset of the feasible production set. If the firm operating at point A moves to point B, the firm can achieve output augmenting efficiency. Similarly if the firm moves from point A to point C, it will be technically efficient from the input saving perspective (measure). Point D in figure 3.2.7 (a) gives the technically optimal scale where output per unit of input is maximized. Figure 3.2.7 (b) represents the corresponding points of figure 3.2.7 (a) in an isoquant frame.

Now the output based measure of technical efficiency E_1 is computed by comparing an observed point of input requirement to produce output Y_a with the input requirement on the frontier production function corresponding to that level of output. In the input coefficient space this means comparing an observed input coefficient point with the point on the transformed isoquant of the frontier function corresponding to the observed output with observed factor proportions.

In figure 3.2.7 (b) this can be stated as $E_1 = \frac{OC}{OA}$. Another measure E_2 is obtained by comparing an observed point of input requirements for an observed output Y_a with the output Y_b obtained on the frontier production function at the same level of input. In 3.2.7 (b) this can be represented by the ratio $E_2 = \frac{OB}{OA}$.

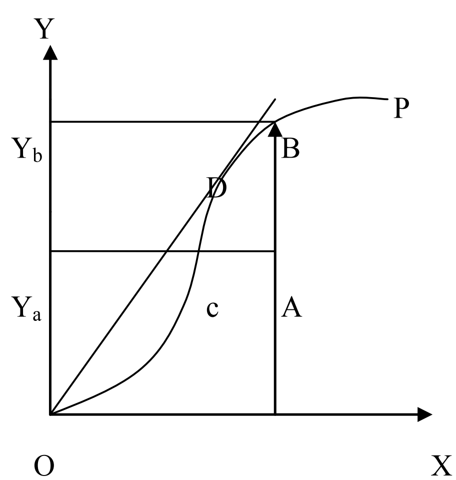

Figure 3.2.7 (a)

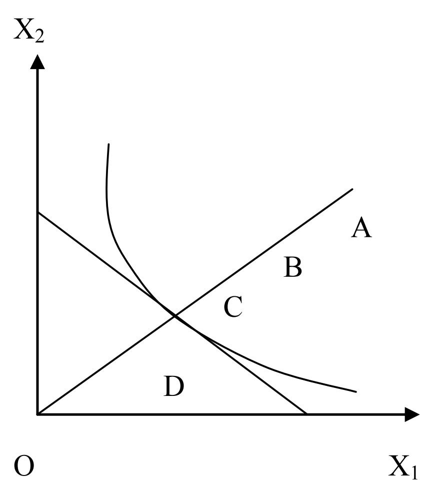

Figure 3.2.7 (b)

We now define the efficiencies in terms of distance functions. We consider the case of multiple inputs and single output. The input oriented measure of technical efficiency is given by the function $TE_i(y,x) = \min\{\lambda : y \leq f(\lambda x)\}$ and the output oriented technical efficiency is measured as $TE_0(y,x) = \max\{\mu : y\mu \leq f(\lambda x)\}$. Now the input oriented technical efficiency can be described as a measure of maximum radial contraction in *X* that enables to produce *Y* and $\lambda < 1$. Output oriented technical efficiency is the maximum radial expansion in *Y* for a given set of input *X*.

We must now introduce costs and input prices in measuring firm level efficiency. When input prices are introduced in explaining production technology it will be possible to measure the efficiency of units in terms of costs and allocation of inputs. A cost frontier is defined as the locus of minimum possible costs to produce a given level of output. A cost function is defined as $C(y,w) = \min\{W : x \in L(y)\}$, where $W = \sum w_i x_i$; w_i is the price of inputs.

The measure of cost efficiency is defined as $CE(y, x, w) = C(y,w)/W$. In other words it is the ratio of minimum possible costs to actual costs. In other words it is the ratio of minimum possible costs to actual costs (Neogi, 2005).

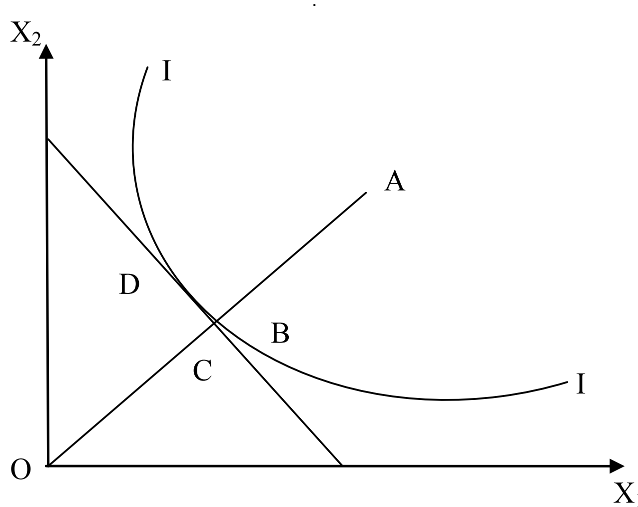

Figure 3.2.8

Let A be a set of inputs required to produce a given level of output as indicated by the isoquant $I\,I$. At point A the firm is neither technically efficient nor cost efficient. If the firm can move down to a point along the ray through the origin where it cuts the iso-cost line, the intersecting point will be cost efficient.

We define cost efficiency CE as the ratio of minimum costs of production with given input prices to observed cost. From figure 3.2.8 we can write $CE = \frac{OC}{OA}$. However all cost efficient points may not lie on the isoquant. That is, all cost efficient points are not technically efficient. For example, input combination at C in figure 3.2.8is cost efficient but not technically efficient. On the other hand point B is technically efficient but not cost efficient. Allocative efficient point is a point which gives both technical and cost efficient combination of inputs. Point D in figure 3.2.8 is a point where the firm is technically efficient as well as cost efficient. Hence the point where allocative efficiency is attained must be a point of tangency between the iso-cost line and the isoquant. The measure of allocative efficiency is defined as $AE(y, x, w) = CE\,(y, x, w) / TE\,(y, x)$

The measure of input allocative efficiency is given by the ratio of cost efficiency to input oriented technical efficiency.

3.2 Formulation of the Stochastic Production Frontier

To begin with we assume that cross-sectional data on the quantities of m inputs used to produce a single output are available for each of N firms or producers in an industry. A production frontier model can be written as

$$y_i = f(x_i\,; \beta).\, TE_i\,, \qquad (3.2.1)$$

Where y_i is the scalar of output of the i^{th} firm, i = 1, 2… N, x_i is a vector of m inputs used by firm i, $f(x_i;\beta)$ is the production frontier, and β is a vector of technology parameters to be estimated. TE_i is the output oriented measure of technical efficiency of firm i. From the definition of technical efficiency developed in section 4.1, we can write

$$TE_i = \frac{y_i}{f(x_i;\beta)}, \qquad (3.2.2)$$

which defines technical efficiency as the ratio of observed output to maximum feasible output? y_i achieves its maximum feasible value of $f(x_i;\beta)$ if and only if $TE_i = 1$. Otherwise TE_i < 1 provides a measure of the shortfall of observed output from maximum feasible output. In equation (3.2.1) the production frontier $f(x_i;\beta)$ is a deterministic frontier. Consequently, in equation (3.2.2) the entire shortfall of observed output y_i from maximum feasible output $f(x_i;\beta)$ is attributed to technical inefficiency. Such a formulation disregards the fact that output can be affected by random shocks that are beyond the control of firms. To incorporate firm specific random shocks into the analysis requires the specification of a stochastic production frontier. To do so we rewrite equation (3.2.1) as

$$y_i = f(x_i;\beta).\exp\{v_i\}.TE_i, \qquad (3.2.3)$$

where $f(x_i;\beta).\exp\{v_i\}$ is the stochastic production frontier. The stochastic production frontier consists of two parts: a deterministic part $f(x_i;\beta)$ common to all producers and a producer specific part $\exp\{v_i\}$ which captures the effect of random shocks on each producer. Hence in case of a stochastic production frontier equation (3.2.2) becomes

$$TE_i = \frac{y_i}{f(x_i;\beta).\exp\{v_i\}}, \qquad (3.2.4)$$

which defines technical efficiency as the ratio of observed output to maximum feasible output in an environment characterized by $\exp\{v_i\}$. Now y_i achieves its maximum possible value of $f(x_i;\beta).\exp\{v_i\}$ if and only if $TE_i = 1$. Otherwise $TE_i < 1$ provides a measure of the shortfall of observed output from maximum feasible output in an environment characterized by $\exp\{v_i\}$, which is allowed to vary across producers. Technical efficiency may be estimated using either the deterministic production frontier or the stochastic production frontier models as given in equations (3.2.1) and (3.2.3). In this study we prefer to use the stochastic production frontier model because the deterministic frontier ignores the effect of random shocks on the production process. The deterministic frontier runs the risk of improperly attributing unmodeled environmental variation to variation in technical efficiency (Kumbhakar, 2000).

3.3 The Basic Cross-Sectional Model and Method of Estimation

Aigner, Lovell and Schmidt (ALS) (1977) and Meeusen and van den Broeck (MB) (1977) simultaneously introduced the stochastic production frontier models. These models allow for technical inefficiency and also acknowledge the fact that random shocks outside the control of producers can affect output. The biggest advantage of the stochastic production is that the impact on output of shocks due to variation in labor and machinery performance, vagaries of the weather, and plain luck can in principle be separated from the contribution of variation in technical efficiency.

We assume that $f(x_i;\beta)$ takes the log-linear Cobb-Douglas form so that the stochastic production frontier can be written as

$$\ln y_i = \beta_0 + \sum_j \beta_j \ln x_{ji} + v_i - u_i \,, \qquad (3.3.1)$$

where v_i is a two sided random statistical noise component and u_i is a non-negative ($u_i \geq 0$) technical inefficiency component of the error term. Since the error term in (3.3.1) has two components, the stochastic frontier model is often referred to as a composed error model. The noise component v_i is assumed to be independently and identically distributed (iid), is symmetric and is distributed independently of the one sided technical inefficiency component u_i. Thus the error term in (3.3.1) $\varepsilon_i = v_i - u_i$ is asymmetric since $u_i \geq 0$. Assuming that v_i and u_i are distributed independently of x_i, estimation of (3.3.1) by OLS provides consistent estimates of β_j s, but not of β_0, since $E(\varepsilon_i) = -E(u_i) \leq 0$. Moreover OLS does not provide producer specific technical efficiency.

However OLS provides a simple test for the presence of technical inefficiency in the given data. If u_i = 0, then $\varepsilon_i = v_i$, the error term is symmetric and the data do not support the presence of technical inefficiency. However if u_i> 0, then $\varepsilon_i = v_i - u_i$ is negatively skewed and there is evidence of technical inefficiency in the data. This implies that a test of the presence of inefficiency in the data can be directly based on the OLS residuals. Schmidt and Lin (1984) proposed the test statistic $b_1^{1/2} = \frac{m_3}{(m_2)^{3/2}}$, where m_2 and m_3 are the second and third sample moments of the OLS residuals. Since v_i is symmetrically distributed, m_3 is the third sample moment of u_i. Thus m_3 < 0 implies that the OLS residuals are negatively skewed and it suggests the presence of technical inefficiency. m_3 > 0 implies that the OLS residuals are positively skewed which is meaningless in this context. Hence positive skewness of OLS residuals indicates that the model is misspecified. Since the distribution of $b_1^{1/2}$ is not extensively published, Coelli (1995) proposed an alternative test statistic that is asymptotically distributed as $N(0,1)$. Negative skewness of OLS residuals occurs when m_3 < 0, a test of hypothesis that $m_3 \geq 0$ is appropriate. Under the null hypothesis of zero skewness of errors in equation (3.3.1), the test statistic $\frac{m_3}{(6m_2^3/N)^{1/2}}$ is asymptotically distributed as $N(0,1)$. The advantage of this test is that it is based on OLS residuals. The disadvantage is that it is based on asymptotic theory and especially in large scale manufacturing number of firms is small in a cross section. We shall consider hypothesis tests of the absence of technical inefficiency based on maximum likelihood estimators.

In order to estimate the model in (3.3.1) Aigner *et al.,* (1977) made specific distributional assumptions regarding the individual components of the composed error term ε_i. In particular the following assumptions are made.

(i) $v_i \sim iid\, N(0,\sigma_v^2)$

(ii) $u_i \sim iid\, N^+(0,\sigma_u^2)$, that is u_i is non-negative and distributed as half normal,

(iii) v_i and u_i are distributed independently of each other and of the regressors.

Assumption (i) is conventional and is maintained throughout our analysis. Assumption (ii) is based on the proposition that the modal value of the technical inefficiency term is zero. It is over simplistic and the distribution of the sum of *v* and *u*, under the distributional assumptions in (i) and (i) are easy to derive. The second part of assumption (iii) is a bit problematic as because if producers know something about their technical efficiency it can influence their choice of inputs. The density functions of *v* and *u* respectively are:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v}.\exp\left\{-\frac{v^2}{2\sigma_v^2}\right\} \quad (3.3.2)$$

and $$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u}.\exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \quad (3.3.3)$$

with $u_i \geq 0$, i.e. it is non-negative half normal. Given the independence assumption, the joint density function of *v* and *u* is the product of their individual density functions and hence

$$f(u,v) = \frac{2}{2\pi\sigma_u\sigma_v}.\exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right\}. \quad (3.3.4)$$

Since $\varepsilon_i = v_i - u_i$, the joint density functions of u and ε is

$$f(u,\varepsilon) = \frac{2}{2\pi\sigma_u\sigma_v}.\exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\} \tag{3.3.5}$$

The marginal density function of ε is obtained by integrating u out of $f(u,\varepsilon)$, which gives

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(u,\varepsilon)\,du \\ &= \frac{2}{\sqrt{2\pi}\sigma}.\left[1-\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right].\exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\} \\ &= \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right).\Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right) \end{aligned} \tag{3.3.6}$$

where $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}, \lambda = \sigma_u / \sigma_v$ and $\phi(.)$ and $\Phi(.)$ are the standard normal density and cumulative distribution functions respectively. The reparameterization from σ_u^2 and σ_v^2 to σ and λ is convenient as because λ provides an indication of the relative contribution of u and v in ε. As $\lambda \to 0$ either $\sigma_v^2 \to +\infty$ or $\sigma_u^2 \to 0$, and the symmetric error component dominates over the one sided error component in the determination of ε. As $\lambda \to \infty$ either $\sigma_u^2 \to +\infty$ or $\sigma_v^2 \to 0$ and the one sided error component dominated over the symmetric error component in determination of ε. In the first case we have the OLS production function model with no technical inefficiency, whereas in the second case we are back to a deterministic production frontier model with no noise.

The Aigner *et al.* (1977) stochastic production frontier model with normal-half normal composed error distribution contains basically two parameters, σ_u and σ_v or else σ and λ. The distribution parameters σ and λ are to be estimated along with the technological parameters β. But before that we must test the hypothesis that $\lambda = 0$, where the test is based on the maximum likelihood estimate of λ. A likelihood ratio test may be conducted to test the hypothesis that $\lambda = 0$ but since the hypothesized value of λ lies on the boundary of the parameter space, it is difficult to interpret the test statistic. However Coelli (1995) has shown that in this case the appropriate one sided likelihood ratio test statistic is asymptotically distributed as a mixture of χ^2 distributions rather than as a single χ^2 distribution. The critical values of such a test are obtained in table Kodde and Palm (1986, table 8.1). If the null hypothesis is true then the production function is equivalent to the traditional OLS average production function where firms are assumed to be fully technically efficient.

The marginal density function $f(\varepsilon)$ is asymmetrically distributed with mean and variance $E(\varepsilon) = -E(u) = -\sigma_u \sqrt{\frac{2}{\pi}}$ and $V(\varepsilon) = \frac{\pi - 2}{\pi}.\sigma_u^2 + \sigma_v^2$. ALS suggested [1 – $E(u)$] as an estimator of mean technical efficiency. But Lee and Tyler (1978) proposed

$$E(\exp\{-u\}) = 2[1 - \Phi(\sigma_u)].\exp\left\{\frac{\sigma_u^2}{2}\right\} \qquad (3.3.7)$$

which is preferred to [1 – $E(u)$] since [1 – u] includes only the first term in the power series expansion of exp $\{-u\}$. Also unlike [1 – $E(u)$], $E(\exp\{-u\})$ is consistent with the definition of technical efficiency developed in (3.2.4).

Using the marginal density function $f(\varepsilon)$ from (3.3.6) the log likelihood function for a sample of N firms in an industry is

$$\ln L = (cons\tan t) - N\ln\sigma + \sum_i \ln\Phi\left(-\frac{\varepsilon_i\lambda}{\sigma}\right) - \frac{1}{2\sigma^2}\sum_i \varepsilon_i^2 \qquad (3.3.8)$$

The log likelihood function in equation (3.3.8) can be maximized with respect to the parameters to obtain maximum likelihood estimates of all parameters of the model. These estimates are consistent as $N \rightarrow +\infty$.

Battese and Corra (1977) parameterization is more convenient from the estimation point of view. Letting $\gamma = \sigma_u^2/\sigma^2$ we see that $\gamma \in [0,1]$. The log likelihood function with this reparameterisation is

$$\ln L = -\frac{N}{2}(\ln 2\pi + \ln\sigma^2) + \sum_i \ln\Phi(-z_i) - \frac{1}{2\sigma^2}\sum_i \varepsilon_i^2 \qquad (3.3.9)$$

Where $z_i = \left[\frac{\varepsilon_i}{\sigma}\right]\sqrt{\frac{\gamma}{1-\gamma}}$.

The next step is to estimate technical efficiency of each producer. We have estimates of $\varepsilon_i = v_i - u_i$, which contain information on u_i. The task is to extract the information that ε_i contains on u_i. A solution to the problem is obtained from the conditional distribution of u_i given ε_i, which contains whatever information ε_i has concerning u_i. Jondrow, Lovell, Materov and Schmidt (1982) showed that if $u_i \sim N^+(0,\sigma_u^2)$, the conditional distribution of u given ε_i is

$$f(u/\varepsilon)=\frac{f(u,\varepsilon)}{f(\varepsilon)}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_*}\cdot\frac{\exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right\}}{\left[1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]} \qquad (3.3.10)$$

where $\mu_* = -\varepsilon\sigma_u^2/\sigma^2$ and $\sigma_*^2=\sigma_u^2\sigma_v^2/\sigma^2$. Since $f(u/\varepsilon)$ is distributed as $N^+(\mu_*,\sigma_*^2)$, either the mean or the mode of this distribution can serve as point estimator of *u*. The mean is given by

$$E(u_i/\varepsilon_i)=\mu_{*i}+\sigma_*\left[\frac{\phi(-\mu_{*i}/\sigma_*}{1-\Phi(\mu_{*i}/\sigma_*)}\right]$$

$$=\sigma_*\left[\frac{\phi(\varepsilon_i\lambda/\sigma)}{1-\Phi(\varepsilon_i\lambda/\sigma)}-\left(\frac{\varepsilon_i\lambda}{\sigma}\right)\right] \qquad (3.3.11)$$

Once point estimates of u_i are obtained, estimates of firm specific technical efficiency can be computed from

$$TE_i=\exp\{-\hat{u}_i\} \qquad (3.3.12)$$

where $\hat{u}_i$ is $E(u_i/\varepsilon_i)$. Battese and Coelli (1988) proposed the alternative point estimator for TE_i as,

$$TE_i=E(\exp\{-u_i\}/\varepsilon_i)=\left[\frac{1-\Phi(\sigma_*-\mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)}\right].\exp\left\{-\mu_{*i}+\frac{1}{2}.\sigma_*^2\right\} \qquad (3.3.13)$$

The point estimators in (3.3.12) and (3.3.13) can give different results since the two formulae are unidentical. We use the later in the present study. But it is to be noted that regardless of which estimator is used the estimates of technical efficiency are inconsistent simply because the variation associated with the distribution of (u_i / ε_i) is independent of *i*. However this is the best estimation strategy with cross sectional data.

3.4 The Normal-Truncated Normal Model

The normal–half normal model may be generalized by allowing *u* to follow a truncated normal distribution. This model was introduced by Stevenson (1980). The distributional assumption on *v* remains the same. Only the assumption on *u* is changed as $u_i \sim iid\, N^+(\mu, \sigma_u^2)$. The third assumption of the ALS model is also maintained. The truncated normal distribution assumed for *u* generalizes the one parameter half normal distribution by allowing the normal distribution which is truncated below at zero to have a non-zero mode. Thus this model contains an additional parameter μ to be estimated and hence provides a somewhat more flexible representation of the pattern of inefficiency in the data. The truncated normal density function for $u \geq 0$ is given by

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma_u \Phi(-\mu/\sigma)} . \exp\left\{ -\frac{(u-\mu)^2}{2\sigma_u^2} \right\} \qquad (3.4.1)$$

Here μ is the mode of the normal distribution which is truncated below at zero. If μ =0, then the density function collapses to the half normal density function. μ may be of either sign. The estimation strategy is the same as in ALS model but with minor changes in the parameterization. The log likelihood function for a sample of *N* firms is

$$\ln L = (constant) - N\ln\sigma \;\; - N\ln\Phi\left(-\frac{\mu}{\sigma_u}\right)$$
$$+\sum_i \ln\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon_i\lambda}{\sigma}\right) - \frac{1}{2}\sum\left(\frac{\varepsilon_i + \mu}{\sigma}\right)^2 \qquad (3.4.2)$$

where $\sigma_u = \lambda\sigma / \sqrt{1+\lambda^2}$, $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$ *and* $\lambda = \sigma_u / \sigma_v$. This log likelihood function can be maximized to obtain maximum likelihood estimators of all parameters in the model. It can be shown that the conditional distribution $f(u/\varepsilon)$ is distributed as $N^+(\tilde{\mu}_i, \sigma_*^2)$ where $\tilde{\mu}_i = (-\sigma_u^2\,\varepsilon_i + \mu\sigma_v^2)/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\,\sigma_v^2/\sigma^2$. Thus either the mean or the mode of $f(u/\varepsilon)$ may be used to estimate the technical efficiency of each firm. The mean of the conditional distribution of u is given by

$$E(u_i/\varepsilon_i) = \sigma_*\left[\frac{\tilde{\mu}_i}{\sigma_*} + \frac{\phi(\tilde{\mu}_i/\sigma_*)}{1-\Phi(\tilde{\mu}_i/\sigma_*)}\right] \qquad (3.4.3)$$

Point estimates of technical efficiency of each firm can be obtained by means of

$$TE_i = E(\exp\{-u_i\}/\varepsilon_i) = \left[\frac{1-\Phi[\sigma_* - (\tilde{\mu}_i/\sigma_*)}{1-\Phi(-\tilde{\mu}_i/\sigma_*)}\right].\exp\left\{-\tilde{\mu}_i + \frac{1}{2}.\sigma_*^2\right\} \quad (3.4.4)$$

This produces unbiased but inconsistent estimates of technical efficiency (Kumbhakar and Lovell, 2000).

3.5 The Econometric Approach for the Present Study

In order to measure technical efficiency at the fishing team level along with its non-input determinants, the present study adopts a Cobb-Douglas stochastic frontier model with inefficiency effects following Battese and Coelli (1995). In other words the stochastic production frontier and the inefficiency effects parameters are simultaneously estimated, given appropriate distributional assumptions. This was originally proposed by Kumbhakar *et al.,* (1991), Reifschneider and Stevenson (1991), and Haung and Lui (1994). Battese and Coelli (1995) is an improvement over the previous methods as it is based on panel data. Moreover this one-stage maximum likelihood approach is statistically consistent with the Kumbhakar *et al.,* (1991) approach and leads to more efficient inference with respect to the parameters (Coelli and Battese, 1996). The approach has been applied empirically by, Coelli and Battese (1996), Battese and Broca (1997).

Acceptably, a Cobb-Douglas form restricts the flexibility of the fish catch technology by imposing the elasticity of scale to be constant and the elasticity of input substitution to be unity. The trans-log production function often creates practical problems in estimation. First with several inputs there is an obvious loss of degrees of freedom (incorporation of log of inputs, square of log of inputs and cross product of log of inputs). Second there is the obvious econometric risk of muticollinearity among the various explanatory variable columns. Although its parameters may be estimated by Seemingly Unrelated Regression Equations (SURE) method, it is ineffective in estimating the stochastic production frontier parameters which requires direct estimation of the production frontier.

The stochastic production frontier developed separately by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) decomposes the error term of the usual econometric production function model into a white random noise component and a one sided inefficiency random component. For the present, we assume a cross-sectional stochastic production frontier model (specified in Kumbhakar *et al,* 1991) as

$$\ln y_i = \ln f(x;\beta) + v_i - u_i \qquad (3.5.1)$$

$$u_i = \gamma' z_i + \varepsilon_i \qquad (3.5.2)$$

The random noise component in the production process is introduced through the error component v_i which is $iid\, N(0, \sigma_v^2)$ in equation (3.5.1). The second error component which captures the effects of technical inefficiency has a systematic component $\gamma' z_i$ associated with the firm specific variables and exogenous variables along with a random component ε_i. Inserting equation (3.5.2) in (3.5.1) gives the single stage production frontier model

$$\ln y_i = \ln f(x_i;\beta) + v_i - (\gamma' z_i + \varepsilon_i) \qquad (3.5.3)$$

The condition that $u_i \geq 0$ requires that $\varepsilon_i \geq -\gamma' z_i$ which does not require $\gamma' z_i \geq 0$ for each producer. It is now necessary to impose distributional assumptions on v_i and ε_i and to impose the restriction $\varepsilon_i \geq -\gamma' z_i$ in order to derive the likelihood function.

Kumbhakar *et al* (1991) imposed distributional assumptions on v_i and u_i and ignored ε_i. They assumed that $u_i \sim N^+(\gamma' z_i, \sigma_u^2)$ i.e., the one-sided technical inefficiency error component has truncated normal structure with variable mode depending on z_i. It is still not necessary that $\gamma' z_i \geq 0$. If $z_{1i} = 1$ and $\gamma_2 = \gamma_3 = \cdots\cdots \gamma_Q = 0$, this model collapses to Stevenson's (1980) truncated normal stochastic frontier model with constant mode γ_1, which further collapses to the Aigner, Lovell and Schmidt (1977) half normal stochastic frontier model with zero mode if $\gamma_1 = 0$. Each of these restrictions are statistically tested. Finally if u_i and v_i are independently distributed, all parameters of equation (3.5.1) can be estimated by using maximum likelihood estimation method. The log likelihood function is a simple generalization of that of Stevenson's (1980) truncated normal model having constant mode μ, with only one change. Constant mode μ is now replaced by the variable mode $\mu_i = \gamma' z_i$, so that the log likelihood function is

$$\ln L = constant - \frac{N}{2}\ln(\sigma_v^2+\sigma_u^2) - \sum_{i=1}^{N}\ln\Phi\left(\frac{\gamma' z_i}{\sigma_u}\right) + \sum_{i=1}^{N}\ln\Phi\left(\frac{\mu_i^*}{\sigma^*}\right) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{(e_i+\gamma' z_i)^2}{\sigma_u^2+\sigma_v^2}\right)$$

-------------- (3.5.4)

Where $\mu_i^* = \dfrac{\sigma_v^2\,\gamma' z_i - \sigma_u^2\,e_i}{\sigma_v^2+\sigma_u^2}$, $\sigma^{*2} = \dfrac{\sigma_v^2\,\sigma_u^2}{\sigma_v^2+\sigma_u^2}$

and $e_i = \ln y_i - \ln f(x_i;\beta)$ are the residuals obtained from estimating equation (3.5.1) simply by OLS. The log likelihood function of (3.5.2) can be maximized to obtain ML estimates of $(\beta, \gamma, \sigma_v^2, \sigma_u^2)$. These estimates can then be used to obtain producer specific estimates of technical efficiency, employing the Jondrow, Lovell, Materov and Schmidt (1982) approach to find the best point estimates of technical efficiency. These estimates are either

$$E(u_i/e_i) = \mu_i^* + \sigma^*\,\frac{\phi(\mu_i^*/\sigma^*)}{\Phi(\mu_i^*/\sigma^*)} \qquad (3.5.5)$$

Or

$$M(u_i/e_i) = \begin{cases} \mu_i^* & if\ \mu_i^* \geq 0 \\ 0 & otherwise. \end{cases} \qquad (3.5.6)$$

Once technical efficiency has been estimated, the effect of each exogenous or environmental variable on technical efficiency can be calculated from either

$[\partial E(u_i / e_i)/\partial z_{ik}]$ *or* $[\partial M(u_i / e_i)/\partial z_{ik}]$. Battese and Coelli (1995) model is an improvement over the Kumbhakar *et al* (1991) model as, (i) it is based on panel data and (ii) the non-negativity requirement $u_i = (\gamma' z_i + \varepsilon_i) \geq 0$ is modeled as $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ with the distribution of ε_i bounded below by the variable truncation point $-\gamma' z_i$. Battese and Coelli (1995) verified that this new distributional assumption on ε_i is consistent with the assumption on u_i that $u_i \sim N^+(\gamma' z_i, \sigma_u^2)$. The trans-log (transcendental logarithmic) production function is a more flexible functional form from which the Cobb-Douglas production function can be obtained as a special case. The special advantage is that homogeneity restrictions are not directly imposed as is done in case of a Cobb-Douglas form. However for practical reasons direct estimation of trans-log production function parameters may not be possible due to the presence of strong or near–perfect multi-collinearity among the explanatory variable columns. The parameters of a trans-log function may still be estimated by Seemingly Unrelated Regression Equation Method (SURE) but that would not serve the purpose of frontier production function estimation (especially of the Battese and Coelli, 1995 form) which requires a direct single step estimation of the production function parameters as well as frontier and inefficiency effects parameters.

The present study thus prefers the Cobb-Douglas functional forms with three endogenous inputs to specify the underlying technological relationship between inputs and output. The three inputs are described in detail below. Two separate estimations of production frontiers with inefficiency effects models are applied for single and paired boat fishing teams with multiple catchers and single boats with single catchers using nets.

$$\ln(Y_i) = \ln\beta_0 + \beta_1 \ln L_i + \beta_2 \ln B_i + \beta_3 \ln N_i + (v_i - u_i) \quad (3.5.7)$$

Here (3.5.7) is the Cobb-Douglas frontier model specification having three inputs – labour (L), boat (B) and net (N). Exact description of all relevant variables used in the study is imperative.

3.6 Variable Construction and Measurement

Output (Y_i)

In the present study output (Y_i) is basically rupee value or money value of fish catch at the fishing team level (converted to monthly figures, i.e. Rupees per month) recorded at the time of sale during survey. This catch level pertains to the effort during a single day by either the team or the individual. Repeated observation on catch level for the same team/individual was possible in the present study which is obviously a drawback. However since the peak fishing season was selected for the survey, the catch level recorded during interview may be taken as an effective estimate of average catch per day. This figure in rupee per day terms was converted to rupee/month simply by multiplying by 30. In case of team of catchers (2 or 3catchers) with either single or paired boats, the catch (or harvest) is basically a team level catch (or harvest). In case of single catchers the catch is at the individual level as the catcher does not have a partner or accompanying person. In sum the catch is measured in terms of rupees/month at the team level or individual level, as the case may be. The details of catch methods and catch teams are outlined in section 3.6.

Measurement of Inputs – Labour, Boat and Net

The three inputs labour, boat and net are all measured physically and not in nominal rupee terms. This is not a technical or statistical problem as the same method is applied across all fishing teams in the sample. **Labour** (*L*), a flow input, is measured in terms of labour hours per month for the i^{th} team. For instance members of a team comprising of 2 members may each spend 5 hours daily on fish catch implying that for this team daily labour hours spent is 10. If this team engages in fishing 5 days per week then weekly labour hours spent by this team equal 50 or in other words monthly labour hours equal approximately 200. A word of caution is needed here as the same fishing team may have 2 to 3 members on different days of a week. In other words for practical reasons the number of catchers may vary slightly across fishing days. In such cases an approximation is most effective.

Fishing tools and equipments play the role of fixed capital or durable capital equipments. These are stock variables or stock inputs in contrast with labour which is a flow input. **Boat size** (B) in terms of length of the boat (in meters) is taken as a proxy physical measure of boat capacity. Alternatively maximum human carrying capacity could have been taken as a proxy measure of boat. However for practical reasons maximum carrying capacity of a boat may be difficult to ascertain. The larger the boat the greater would be the catch effort, so that length of the fishing boat.

In the present study length of **net** (N) in meters is taken as a proxy measure of net capacity for each team. Arguably in open access fishing net is a heterogeneous input like boat as because net density or gapping may vary across catchers given the fact that in most places different catchers specialize in different types of fish. For the fishing teams of Sone Beel included in the present sample, the nets are almost of identical density or gapping and do not differ much across teams. From the field level experience during the survey and corresponding interview it was found that almost all fishing teams use a particular type of net and on the ground net gapping does not vary much across teams. Hence length of the net is taken as a measure of net capacity.

In sum, the production function for the present study models monthly output in value terms (value of catch per month) as the outcome of catch effort due to labour (L), boat (B) and net (N). In many similar studies it is customary Effort Index is not constructed. Depreciation of fixed inputs like tools and equipments (i.e., boat and net) is ignored.

There are some single catchers or one member teams that do not use net. For single catcher fishing teams using non net inputs like dori (DR) and kathi (KT) are measured in physical terms. Dari, which implies a cylindrical drum traps, measured in numbers and kathi, a vertical slit trap is measured length in meters. These two are found to be complementary inputs as these are jointly used for fishing.

Inefficiency Effects Variables

Finally the inefficiency effects components in the composed error term ($v_i - u_i$) needs to be elaborated in the context of the present study where z_i are all non-input inefficiency effects variables (refer to equations 3.5.2 and 3.5.3). In the present study the inefficiency effects component has the following form.

$$\gamma' z_i = \gamma_1 + \gamma_2 z_{2i} + \gamma_3 z_{3i} + \gamma_4 z_{4i} + \gamma_{5i} z_{5i} \qquad (3.5.8.)$$

where, the z_i's are firm specific non-input variables (some of which may be categorical or dummy variables) that potentially influences the technical efficiency of the fishing team or individual fisherman as the case may be.

Inefficiency Effects Variables in Case of Team Catchers

In case of multiple or team catchers, z_{2i} is the experience (EXP) of the fishing team members as measured by the average number of years spent by the catchers in fishing; z_{3i} is education (EDU) of the catchers as measured by the average number of years of formal schooling and z_{4i} captures non-fishing income (NFI) at the fishing team level from agriculture and allied activities during slack season. The additional variable z_{5i} is inapplicable for the team level analysis.

Inefficiency Effects Variables in Case of Individual Catchers

In case of the inefficiency effects model (3.5.8.), as applied to the frontier model on single catchers, z_{2i} is the experience (EXP) of the catcher/fisherman as measured by the average number of years spent by the catcher in fishing; z_{3i} is the non-fishing income (NFI) of the catcher/fisherman from agriculture and allied activities during slack season; z_{4i} (SAND) is type of sanitation system dummy (assumes 1 for safe or improved sanitation/toilet facility in the catcher's households, and 0 otherwise) and z_{5i} (HTD) is housing type dummy of the catcher (value 1 if the catcher resides in a pucca or semi-pucca house, and 0 otherwise).

In other words some dimensions of human development and quality of life indicators are being captured by means of these non-input inefficiency effects variables. Non-fishing income of the catcher (NFI) is supposed to raise per capita annual household income and consumption thereby raising living standards. Type of sanitation in the household is an effective indicator of physical standards of living as well as that of health and the extent of hygienic practices in the household. The same is true for access to safe drinking water but for the present access of safe drinking water is ignored. On the other hand the type of housing accommodation of the catcher or fisherman is yet another indicator of physical standards of living and the life-style. People residing in totally kutcha houses are expected to have lower standards of living compared to people living in pucca or even semi pucca houses.

Fishing in open access water bodies such as large lakes and beels is largely dependent on skill of the catcher(s) and arguably skill is heavily dependent on experience. The experience of the team members is vital for increasing technical efficiency of the team. In other words the present study hypothises that other things remaining the same, the team having more experience on an average is expected to be technically more efficient. Thus as a non-input variable, experience is supposed to have a positive impact on technical efficiency or a negative impact on technical inefficiency. In the inefficiency effects model the LHS of equation (3.5.8) represents inefficiency effects and hence comes with a negative sign (see equations 3.5.2 and 3.5.3). Thus negative coefficients of the z_i terms imply negative influence on technical inefficiency and hence positive influence on technical efficiency.

The inclusion of non-fishing income of the catcher as an inefficiency effects variable needs clarification. In beel areas such as the one chosen for the present study (the Sone Beel) water levels go down significantly during the dry season or the winter months allowing for agricultural operations on the beel areas. Paddy cultivation during the dry season is the most common agricultural practice in the Sone Beel during winter months. Irrigation is not rain dependent as water is plentiful in some low lying pockets which form small ponds or natural water-sheds. Interestingly all fishermen do not have access to such land and hence either work as day labourers or do some unskilled jobs in the region when not engaged in fishing. However a significant section of the catchers do engage in paddy cultivation during the winter months on a small scale. In sum the fishermen have a source of livelihood (or say income) in the form of a non-fishing income during the non-fishing season or the dry season when the water body shrinks immensely making fishing very difficult, if not impossible. The present study hypothese NFI as an inefficiency effects variable that can potentially influence technical efficiency of the fishing team as well as that of the individual catcher. NFI or non-fishing income may have a dampening effect on technical efficiency of fish catch. Catchers with relatively high NFI may be dedicated to some economically gainful activity other than fishing either during slack season or even during a part of the peak season. In such situations attention and dedication towards fishing is likely to be compromised. Thus the dedicated fishermen or catchers are expected to be relatively unskilled in non-fishing activities and hence their NFIs are likely to be low. This seems to suggest that lower the NFI, higher is the dependence of the catcher on fishing for his livelihood. Thus for catchers with low NFIs, dedication towards fishing is expected to be high. This study anticipates a negative relationship between NFI and technical efficiency of fish catch.

Arguably formal education has little to do with technical efficiency of fish catch in Sone Beel, especially during the peak monsoon months when fishing is subject to open access. Efficiency in fishing does not depend on the extent of formal education but rather on the experience. Interestingly higher the years of schooling, greater is the possibility of dependence on non-fishing sources for livelihood. Catchers in the region are poor (i.e., belonging to BPL category, discussed in the findings in Chapter 4), with very little or no formal education or schooling. In the present study education (EDU) as measured by the number of years of formal schooling acts as a proxy for awareness and overall knowledge about one's surroundings, nothing more. Education is deliberately kept as a non-input variable in the inefficiency effects model to capture the impact of overall awareness on catch efficiency.

Finally sanitation type dummy and the housing type dummy could not incorporated in the team level efficiency analysis as these dummies would assume different values for different members of a team making it difficult to model or quantify at the team level. For the individual catchers however there is no such complication.

Hypothesis Testing in the Frontier Model

Testing the null hypothesis no technical inefficiency is important. The null hypothesis of no technical inefficiency can be tested by applying the Likelihood Ratio Test. The likelihood ratio test is based on the likelihood ratio statistic (LR) defined as,

$$LR = -2\ln[L(H_0)/L(H_1)] \qquad (3.5.9)$$

Where, $L(H_0)$ and $L(H_1)$ are the optimum values of the likelihood function under the null hypothesis (no technical inefficiency or OLS) and alternative hypothesis (presence of technical inefficiency under the Aigner *et al.* 1977, Normal– half-Normal error specification) respectively. But since the hypothesized value of λ (which equals σ_u/σ_v) lies on the boundary of the parameter space it is difficult to interpret the test statistic. It can be shown that the *LR* statistic in (3.5.9) follows a mixed χ^2 distribution that asymptotically approaches χ^2 distribution with degrees of freedom equal to the number of restrictions imposed in the model (Coelli, 1995). Similar is the test of the hypothesis that inefficiency effects are totally absent in the model. To test that null hypothesis of no inefficiency in the data, which is equivalent to setting $\lambda = 0$ the Kodd and Palm (1986) critical values for relevant degrees of freedom are used. All estimations are done using the software package *FRONTIER 4.1* for WINDOWS (Coelli, 1996).

3.7 Survey Methods and Data

Data for the present study is completely primary in nature and is based on information collected between July and September 2013, from Sone Beel fishing boat-landing sites. Fishing in the Sone Beel is officially managed by the Sone Beel Fishermen Cooperative Society (established in 1975). A new set up for the management of market transactions related to fish catch that includes auctioning and bidding (called *Machher Arath*, i.e., the whole sale trading and transactions place) was formed in July 2012. Under this newly formed institution, fish catchers and sellers sell their daily catch indirectly through a formal bidding system. The number of fish auctions observed in the landing sites are 3 to 4 and this number fluctuates depending on the season. However, only 2 – 3 auctions are found to be active and functional on a regular basis. It was further observed that this system of fish bidding helps fishermen to get better prices for their daily catch. However, catchers are charged with five percent of the value of their daily catch as fee on account of participation in the organized bidding under the *Machher Arath*.

The necessary information was collected from selected single boat using fishermen from the Sone Beel fish-landing sites employing the direct interview method. A well structured pretested survey schedule was used that focused specifically on sale and quantity of catch, labour hours spent, fishing equipments, socio-economic features etc. The face-to-face interviews were conducted in collaboration with four functionally literate volunteers (selected to carry out field survey) from the fishing community. Two enumerators having satisfactory working experience in the field (graduates) helped the volunteers along with the local members (involving with Sone Beel Fishermen Cooperative Society) in the survey work. These volunteers have regular contacts with fishing households dwelling around the Sone Beel. Around 50 to 60 fishing teams with their boat usually land in the fish-landing sites between 6 – 7 AM during the peak fishing season. The crew members engaged in selling and grading of fish in the auctions, varies from two to four persons. In view of the unorganized nature of the transactions activity, data collection was challenging, especially when fishermen were uninterested in facing the interview. Lack of willingness to cooperate was perhaps due to the excessive workload and physical stress and strain associated with catch and sale of fish during the peak working hours during (6 – 8 AM). Expectedly, the fishermen are extremely busy over their respective transactions during peak hours and are hardly in a position to face interviews. Time and place had to be suitably chosen so as to undertake an uninterrupted interview with the single boat using fishing team members. Strictly speaking, under such circumstances, random sampling (and even systematic sampling) is difficult if not impossible.

As per secondary data collected from Sone Beel Fisherman's Cooperative Society office, the total number of registered fishermen under the society is 4934. These people belong to traditional fishing community. Three distinct types of fishing teams are commonly observed in the Sone Beel – (i) paired boat with 6 to 8 catchers, (ii) Single boat with 2 to 3 catcher (net users) and (iii) single boat single catchers (a few with net, and others with various traditional equipments). In the present study all distinct types of fishing teams are separately taken into consideration for measuring technical efficiency.

No official statistical records on the number of fishing boats currently engaged in fishing in the Sone Beel are available. According to a' priory information (based on unofficial sources), there could be total of number of 50-60 paired boat fishing teams (each team comprising of 6 to 8 catchers), out of which 16 paired boat teams are included in the present study. As per Panchayat level and other block level sources the total number of catchers using single boats could be around 600. Out of these, 149 single boat using teams with 2 to 3 catchers and 16 paired boats are selected for the study. Thus number of net using fishing teams with multiple members equal 165.

Single catchers using single boats comprise yet another category of fishermen in the Son Beel. 160 such catchers are included for study in the present sample. However not all catchers out of these are net users. Out of these 60 single catchers in the present sample do not have access to net and use traditional fishing tools and equipments such as fish traps like cylindrical drum traps, vertical slit traps (locally known as *dori* and *kathi*). Net using single catchers (equaling 100), and *dori* and *kathi* using single catchers (equaling 60) have been separately analyzed in the study.

Thus in a nutshell the types of catchers along with the subsample sizes in the present study are as follows: (i) single boat using team catchers with 2 to 3 members (149), (ii) paired boat using team catchers with 6 to 8 members (16), (iii) single boat using single catchers with net (100) and (iv) single boat using single catchers without net (60). Thus total sample size of teams in the present study is 149 + 16 + 100 + 60, i.e. a total of 325 teams.

Because of the flow nature of the population in the boat landing sites, strictly random sampling could not be conducted. For the present study a convenient large sample of 165(including single boat using fishing teams comprising of 2 to 3 members and for paired boat teams comprising of 6 to 8) members is chosen. And 160 of single boat single catchers (i.e., 100 single boat with net users and 60 single boat without net users). The overall number of boats in the present study was covered as 325 which is around 54 percent of the total fishing boats in the study area. The sample size is quite large relative to the size of the population and thus small sample bias may be ruled out. The sample size was fixed using the following formula when population size is not exactly known. $n = Z^2.s^2/d^2$ where *n* is the minimum sample size to be chosen, Z is the value of the standard normal distribution function at 0.05 level, *S* is the population standard deviation of the variable, and *d* is acceptable standard error of the mean of the variable of interest – value of fish catch in this case. Since exact size of the population is not known, S is fixed through a pilot survey. Specifically, $s = \acute{s}\,.\sqrt{\acute{n}/(\acute{n}-1)}$ where n' is the sample size for the pilot survey, and s' is the standard deviation of value of catch computed from the pilot survey. Clearly the smaller is *d* the larger is the minimum sample size needed for statistically robust estimation and inference.

Labour effort is heterogeneous across fishing teams as because 87 fishing teams in the sample (out of 149 fishing teams), have 3 catchers while the rest, i.e., 62 fishing teams have 2 catchers per team. In other words there are 87 times 3 plus 62 times 2, or a total of 385 catchers in all in the sample of 149 fishing teams. For pair boat teams (out of 16 fishing teams), there are a total of 110 catchers in all in the sample of pair boat teams. All fishermen were interviewed for necessary information on key production function related variables such as value of catch, labour hours spent, and fishing tools and equipments like nets and boat. Moreover data on certain non-input factors as, experience in fishing, years of formal schooling and income from sources other than fishing were collected through the personal interview. Data collection was not possible during the fish auction hours due to large and spontaneous public gathering and outcry. The surveyors had to tailor interview time and place according to the convenience of the catchers. The availability of the fishermen during busy working hours was the other key concern.

In short the interviews were conducted just after sales of fishes when the catchers were relatively free and away from crowded gatherings. Interviewing became a lot easier when participants could relax and feel comfortable. Interviews were carried out in usual places of gathering and hang-outs such as tea stalls adjacent to the transaction sites. The respondent of each team was mainly the skipper or boat owner who provided precise information regarding fishing practices of his team. Open ended discussions centered around vessel use, fishing duration, quantity of daily catch, types of fish, income, and even on their respective household conditions. Discussions also focused on overall constraints faced by fishers and the ecological condition of fishing sites. The duration of the interviews with each group was approximately twenty minutes. However for single catcher teams the data was collected from the fishing households because socio economic variables (for example number of family members, education, housing structure and sanitary facility, etc.) are important to fulfill the desired objectives and this can only be possible from household survey.