Chapter III

REVIEW OF LITERATURE

The present chapter seeks to review the literature in the context of the central theme of this dissertation, the X-efficiency. In particular, our thrust is on the concept of the term, the methodologies for its measurement and also relevant empirical findings in the literature. These are, however, subject-matters in three subsequent sections that follow. In the section on empirical findings, we concentrate mainly on Indian experiences. Section IV concludes.

Section I: X-efficiency theory and criticism.

Dissatisfied with the neo-classical paradigm, especially its assumption of profit maximization, as apparent in Braithwaite[111], Simon[112] and Cyert and March[113], Leibenstein propounds the idea of X - efficiency in 1966.[114] His idea bridges the gap between neo-classical hypothesis and real-life experiences.

Leibenstein's X-efficiency theory deviates from the traditional theory in the treatment of human resources. The human inputs cannot be treated in the same way as the non-human inputs insofar as a firm cannot purchase human inputs, but human time. Moreover, the human time that the firm buys does not enter into the production process; it is the human effort that matters.[115] The human effort in turn depends upon the motivation. In a multi-person firm, the level of motivation varies, and so the effort level, across the individuals since the motivation for work determines the effort that a worker provides in the work place. Now, the motivation of an individual determines

---

[111] Braithwaite, *Theory of Games*
[112] Simon, 'Theories of Decision Making'
[113] Cyert and March, *A Behavioural Theory of the Firm*
[114] Leibenstein, 'Allocative Efficiency Vs "X-Efficiency"', pp.392-415
[115] Leibenstein, 'Organizational or Frictional Equilibria', p.601

how he interprets the job. This is so because, in a real world scenario, (i) all inputs are not traded in the factor market, (ii) there is no fixed correspondence between inputs and outputs as the production function is not exactly known, and (iii) there exist the problems of asymmetrical information leading to the principal agent problem. 'There exist a distinction between agents and principals'[116] and the 'agents may' take 'trade decisions which are not to the mutual benefit of the principal.'[117] Given that motivation determines the level of effort, motivation is a crucial issue in the question of efficiency. Leibenstein, however, argues that any deviation from an optimal effort level can be attributed to: (i) incomplete labour contract, (ii) non-market factor inputs, and/or (iii) incomplete specification of the production function, as the production function guiding the organisations production activity is not known with surety because of the 'experimental element involved'.[118] The deviation from optimal output level can also be explained by the following factors, (i) rationalisation of costs, (ii) innovation, and (iii) introduction of new commodities, of these factors, cost reduction is attributed to increase in X-efficiency.[119]

To explain these factors, we note that each individual who joins an organisation brings into the organization not only his/her idiosyncratic motivation but also his/her utility-effort range. The lowest limit of the range corresponds to the point where the individual prefers to work rather than remaining jobless. Its highest limit corresponds to the point beyond which he does not put any effort. The area between these two points is the individual's inert area. This inert area is constituted by a set of effort points whose associated levels of utility are not the same. The action that is required to move from a lower utility level to a higher utility level involves costs

[116] Leibenstein, 'Aspects of the X-efficiency theory of the firm', p.588
[117] Ibid, p.588
[118] Leibenstein, 'Allocative Efficiency Vs "X-efficiency"', p.407
[119] Ibid, p.408

which may not be compensated by the gain in utility, and thus the inertia prevails. An organization is in a state of equilibrium only when the managers, as well as the workers, are in a state of inertia.

In the theory of X-efficiency, however, an entrepreneur is a person who promotes changes over resistance. The X-efficient entrepreneur or manager is a self-propelled input-gap filler and completer. This is so because he fills prevailing gaps in the input/output market, and covers up the lacuna in the production function.[120] Leibenstein writes that the inert area is a consequence of the discretion that the workers display, and this ultimately leads to 'entropy'[121]. The word entropy is used in the sense 'of a tendency towards disorganization.'[122] He points out, 'If the vertical and horizontal constraints are much weaker than anticipated' then the individual will shift those effort points within his/her effort position that s/he prefers and 'which is less likely to be connected with the objective of the firm.'[123] It is here that an entrepreneur with his motivating skill comes into the picture.

A questionable assumption of the neoclassical economics that Leibenstein seeks to address is the assumption of rationality. This very assumption of rationality, outright ignores the existence of X-efficiency. In the neoclassical economics, the centre of analysis is a firm/ an industry. *Per contra*, in the theory of X-efficiency, the unit of study is the individual, who is governed by his/her motivation, inert area, as well as APQT set (that is, the set of Activity, Pace, Quality and Time). This bundle determines the 'Activity' that the individual chooses; the 'Pace' at which he carries out those activities; the 'Quality' of the activity he performs and the 'Time' spent on

---

[120] Leibenstein, 'Organisational or Frictional Equilibria', p.612
[121] Leibenstein, 'Aspects of the X-efficiency theory of the firm', p.600
[122] Leibenstein, 'A Branch of economics is missing', p.487
[123] Ibid, p.487

performing the activities. This APQT bundle is partially determined by an individual's work potential set of desires, attitude and the sense of responsibility about the activities of others surrounding him. Individual managers as well as workers have their own APQT set. If the effort levels go down, it leads to entropy, which is a swan song for the organization.[124]

In contrast to complete rationality of neo-classical economic theory, the X-efficiency theory assumes selective rationality for individuals. It signifies that an individual selects the extent of his/her deviation from the maximizing behaviour, which in fact depends on his/her personality. Personality is in turn determined by his/her responsibility standard and unconstrained behaviour - a concept akin to the Freudian concept of id, ego and superego.[125] The responsive behaviour and unconstrained behaviour determine together the internal pressure that an individual faces. Coupled with external pressures, it then determines an individual's deviation from his/her maximizing behaviour.[126]

Taking clues from the management discretion theory, the X-efficiency theory recognises the divergence in objectives between managers and owners of a firm,[127] and argues that it is the cause of X-inefficiency. To keep the objectives of the managers in line with the objectives of the owners, Crew *et al*[128] introduce the concept of policing, which enables the X-efficiency theory to ground on the profit-maximising behaviour. Their logic is that though there is a divergence in the objectives between owners and managers, the latter are compelled to opt for the best

[124] Leibenstein, 'Competition and X-efficiency', p.767
[125] Leibenstein in his article 'Aspects of the X-efficiency theory of the firm' p.583 writes that 'individuals behave on the one hand (a) as they like to and on the other hand, (b) as they feel they must' and 'they strike a compromise between' the two. This is corollary to Freuds' concept of id ego and superego. See Freud '*The Ego and the Id'*.
[126] Leibenstein, 'A Branch of economics is missing', p.485
[127] Crew *et al*, ' X-Theory versus Management Discretion Theory'
[128] Ibid, pp.173-177

use of inputs, and thus to ensure the maximisation of profit. Their theory is labelled as the X-theory.

But Blois[129] contradicts the position, and using the concept of inert area, he explains why firms do not maximise profit. *Prima facie*, owners should not welcome any deviation from profit maximisation, since profit is distributed among them as dividend. But, he argues, the owners may not pressurise the management to go for profit maximisation if the cost of movement from the equilibrium position to the profit-maximising equilibrium is higher than the benefit accruing to the majority of the share-holders. In that case, the owners of the firms will not seek to reduce the X-inefficiency which has crept in due to the differences in the objectives of the owners and managers.[130]

The neo-classical assumption of rationality that generates the objective of profit maximisation is challenged on the question of decision-making unit. This is so because, according to Leibenstein, the neo-classical microeconomics is not micro enough. The unit of analysis in the neo-classical economics is a firm or a household which takes up decisions. But a firm or a household is made up of individuals, and hence the focus of analysis should shift from the firm or household to the individual motivation.[131] In this sense, the X-inefficiency is in-built in the framework of the neoclassical microeconomics. Only when the focus is confined to firms as the decision-making unit, the X-inefficiency is eliminated from the neoclassical theory.[132]

[129] Blois, 'A note on X-efficiency and profit maximization'
[130] Ibid pp. 310-311.
[131] Leibenstein, 'Comment on the nature of X-efficiency', p.327
[132] Leibenstein, 'Competition and X-inefficiency: Reply' p.766

Comanor and Leibenstein[133] relate the vices of monopoly with the X-inefficiency. They point out that competition reduces costs by inculcating 'disciplinary pressure' on all firms in an industry. It eliminates the high-cost producers, and disciplines managements and employees to optimally use the inputs. They thus believe that 'a shift from monopoly to competition has two possible effects: (1) the elimination of monopoly rents, and (2) the reduction of unit costs.'[134]

Stigler,[135] however, re-emphasises the neo-classical assumptions of rationality and profit maximization, and points out that analysing the X-efficiency by the economic behaviour is a 'short gun marriage', which is 'not fertile'.[136] According to Stigler, individuals do not maximise output; rather they maximise utility where output is only a component. If an individual increases output by a higher effort level, this increase should not be regarded as an increase in efficiency. The effort that is required to enforce the contracts cannot be regarded as X-inefficiency as Stigler considers 'New techniques of contract enforcement may be productive as other improvements of technology'[137] rather than increase in X-efficiency.

Stigler also questions the existence of X-efficiency pointing out that it can be assimilated into 'the traditional theory of allocative inefficiency.'[138] He, in particular, challenges 'the property of treating changes in motivation as a source of changes in output',[139] and argues that '[t]he effects of these variations in output are all attributed to specific inputs, and in the present case chiefly to the differences in entrepreneurial

---

[133] Comanor and Leibenstein, 'Allocative efficiency, X- efficiency '
[134] Ibid, p. 304
[135] Stigler, ' The Xistence of X-Efficiency'.
[136] Ibid, p.214
[137] Ibid, p.214
[138] Ibid, p. 213
[139] Ibid, p.213

capacity.'[140] Moreover, he believes, the removal of motivational deficiencies that exist in monopoly does not improve the level of efficiency; it only changes the composition of output. He stresses that 'in every motivational case, the question is: what is output? Surely no person ever seeks to maximize the output of any one thing: even if the single proprietor, unassisted by hired labor, does not seek to maximize the output of corn: he seeks to maximize utility, and surely other products including leisure and health as well as corn enter into his utility function. When more of one goal is achieved at the cost of less of another goal, the increase in output due to (say) increased effort is not an increase in "efficiency;" it is a change in output,'[141]

Di Lorenzo extends Stigler's criticism on the strength of arguments which he borrows from the literatures on property rights and agency costs. His main argument is that 'managers of private monopoly firms, acting as rational utility maximizing agents, will not pursue profit maximization less arduously than will their counterparts in more competitive industries.'[142] Since managerial reward is tied to profitability, the managers of both monopoly firm as well as competitive firm will not be 'motivationally deficient'[143]. If managers fail to maximize profit then 'it will depress stocks'[144] and make it vulnerable to take-over bids. Inefficient, non profit maximizing managers will face managerial competition both from within and outside, and this will discipline the managers to maximize profit.[145]

---

[140] Ibid, p.215
[141] Ibid, p.213
[142] DiLorenzo, ' Corporate management, property rights', p.117
[143] Ibid, p.118
[144] Ibid, p.121
[145] Ibid, p.121

Section II: Method of Measurement

Similar to the debate on the concept of efficiency, the measurement of efficiency attracts an intensive controversy among economists and statisticians alike. The earliest method of productivity measurement is the method of average labour productivity.[146] But there is a serious drawback of the method in that the firm's productivity is measured in terms of the productivity of labour alone, ignoring all other factors of production. This drawback has led to the introduction of the efficiency index method in this literature,[147] which takes into account all factors of production. In view of the adding-up problem for different inputs involving various units, it considers them in value terms. But this method is also blurred in view of appropriate input-prices for individual firms. If all the firms face with the same input-prices, the index method is reduced to comparing production costs among different firms. [148]

Farrell overcomes the problems of the earlier methods by applying the concept of production frontier to measure efficiency. Farrell's method splits up the technical efficiency and the allocative efficiency, which the earlier methods are unable to do. His methodology is based on the assumption of an efficient production function for an industry, which is compared with the production function of individual firms to make a judgement on their respective inefficiency levels. This methodology thus requires to properly define the most efficient production function for an industry. The production function may be defined theoretically or empirically and Farrell uses the empirical production function due to the difficulties posed by the theoretical production function.[149] But there are three major criticisms against this approach. First, it requires

[146] Farrell, 'The Measurement of Productive Efficiency', p.263
[147] Ibid, p.264
[148] Ibid, p.264
[149] Ibid, p. 255

a priori knowledge about the needs of a firm; second, as the theoretical production function has been assuming various complex characters over the last decades; and third, it talks about what can best be achieved rather than what is actually achieved. As such, this approach is less preferred in empirical studies. The alternative method is, however, to construct the most efficient production function by taking into account the best input-output combinations in various lines of production in an industry from different firms. Farrell shows that such an efficient production function is 'represented by an isoquant'[150] which is convex to the origin. The technical inefficiency of a firm is derived by comparing the efficient frontier with the 'hypothetical firm which uses the factors in the same proportion' wherein this 'hypothetical firm is constructed as a weighted average of two observed firms' 'its inputs and outputs is the same weighted average of those observed firms'. [151]

Aigner and Chu[152] improve Farrell's methodology by incorporating an error component in the analysis. They, indeed, consider in line with Farrell - that there are two types of production functions, one for the industry and a set of others for different firms in the industry. The former contains only pure random shocks. The latter, however, reflects the differences in technical and economic inefficiencies across the firms. Aigner and Chu argue that differences in technical efficiency arise owing to differences in the holdings of capital equipment, both in quantity and vintage, as also sources of fund, internal and external. Differences in the economic efficiency, however, stem from the adjustment lags in production during the changes in the market situation.

[150] Ibid, p.255
[151] Ibid, p.256
[152] Aigner and Chu, 'On estimating the Industry Production Function'

Aigner and Chu consider a Cobb-Douglas production function:

$$x_0 = A x_1^{\alpha logR}\, x_2^{\beta \log r} u \qquad (3.1)$$

where x_0 is the output; x_1, x_2 are the inputs; and u is the random shock which is assumed to be one sided; R represents the ratio of the value of equipment to plant and r the ratio of the number of technical personnel to production workers. [153] On the basis of cross-section firm-level data, Equation (3.1) can be estimated to derive the industry production function. They note that 'since r and R vary over firms, the firm production function can be derived from the industry function by appropriately adjusting these proxy variables.'[154] The technological inefficiency for individual firms is then obtained as the difference between the industry production function and the firm production function. To estimate the economic inefficiency, however, Aigner and Chu propose the share equations for inputs in expenditure along with the production function (3.1). Since the firms' deviations lie only on one side, they propose to estimate the function 'within the framework of linear programming',[155] i.e. by minimising the sum of residuals as a linear loss function.

Timmer[156] improves Farrell's method by incorporating a probabilistic production frontier in the place of an average production function. He points out that since only 'extreme observations are used in the estimation of the frontier, it is highly subject to errors in data' and the use of probabilistic production frontier would

---

[153] Aigner and Chu, 'On estimating the Industry Production Function' pp. 827-831
[154] Ibid, p.833
[155] Ibid, p. 832
[156] Timmer, 'Using a Probabilistic Frontier Production Function'.

solvethe 'problem of spurious errors in extreme observation'.[157] The equation (3.1) can be written as

$$\Pr\left(\hat{A}x_1^{\hat{\alpha}}x_2^{\hat{\beta}} \geq x_0\right) \geq P \qquad (3.2)$$

where P is the specified minimum probability.[158]

Aigner, Lovell and Schimidt[159] (ALS) incorporate both positive and negative disturbances in their model, which is specified as

$$y = x\beta + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots.(3.3)$$

where y is nx1 vector of outputs, x is n x k matrix of observation on k fixed regressors and β is k x 1 vector of unknown regression coefficients. The elements of n x 1 disturbance vector ε is determined by

$$\varepsilon_i = \varepsilon_i^*/\sqrt{1-\theta} \text{ if } \varepsilon_i^* > 0$$

$$= \varepsilon_i^*/\sqrt{\theta} \text{ if} \varepsilon_i^* \leq 0$$

where ε_i^* has either a positive or negative truncated normal distributions. [160] There are two components in ε, ε_i^* and θ. The term θ measures random variations due to (i) the firms' inability to implement the best practise technology, and (ii) the problem of measurement in y. The other component,ε_i^*, incorporates the firm-level inefficiency as well as the effects of favourable and unfavourable external factors, which are beyond the control of the firms, and, therefore, purely stochastic. Their stochastic frontier assumes the form of:

[157] Ibid, p.781
[158] Ibid, p.781
[159] Aigner *et al*, 'Formulation and Estimation'.
[160] Ibid, p. 23

$$y_i = f(x_i, \beta) + \varepsilon_i \qquad (3.4)$$

The term ε_i in equation (3.4) represents a composite error comprising of pure white noise and inefficiency. Thus,

$$\varepsilon_i = v_i + u_i \qquad (3.5)$$

where v_i ~ IIDN (0, σ^2) (3.6)

u_i ~ N (0, σ^2) (3.7)

The term v_i is symmetrically distributed, but u_i belongs to the negative domain, $u_i \leq 0$. These non-positive $u_i's$ signify that each firm's output must lie on or below the frontier. The deviations from the frontier, however, represent the firm-level technical or economic inefficiencies. The stochastic component v_i accounts for situation where the output varies as a result of favourable or unfavourable external factors. Thus the production function in equation (3.4) can be written as

$$y_i = f[(x_i, \beta) + v_i - u_i] \qquad (3.8)$$

In this case, the efficiency measurement is given by the ratio $y_i \big/ [f(x_i, \beta) + v_i - u_i]$ instead of $y_i \big/ [f(x_i \beta)]$.[161]

Simultaneously with ALS, Mauseen and van Den Broeck[162] put forward a stochastic frontier method for the measurement of efficiency. Similar to the ALS

[161] Ibid, pp. 22-25

method, they propose an efficiency model with a composed multiplicative disturbance term - the 'product of a "true" error term and an inefficiency measure'. The proposed model is:

$$y_t = \emptyset(x_t)k_t u_t t = 1, \ldots \ldots \ldots T \qquad (3.9)$$

Here, k_t is the inefficiency term distributed in (0,1) interval, and the random shock u_t is distributed over (0,∞).[163]

Battese and Corra[164] also recognise the composite error structure in a firm's production function. They define the production frontier as

$$y_t = x_t\beta + E_t \qquad (3.10)$$

where

$$E_t = U_t + V_t \qquad (3.11)$$

The random error U_t is, by assumption, and 'arise by truncation of the normal distribution with mean zero and positive variance'[165]. The stochastic component V_t is assumed to follow normal distribution with zero mean and positive variance.[166]

These three papers, all published in 1977, enhance the acceptability of the stochastic frontier analysis by virtue of its stochastic component in errors. Its main difficulty – the decomposition of the error term into the random shocks and the inefficiency component – is overcome to some extent in Jondrow *et al*[167] (JLMS) in 1982. The JLMS provides a measure of producer-specific estimates of efficiency, but

---

[162] Mauseen and van den Broeck, 'Efficiency estimation'
[163] Ibid, p. 436
[164] Battese and Corra, 'Estimation of a Production Frontier Model'
[165] Ibid, p.170
[166] Ibid, p.170
[167] Jondrow *et al*, 'On the estimation of technical inefficiency', 1982.

fails to decompose the *error* term into technical and allocative (economic) efficiencies. Following the prevailing practice, the model is estimated using a two-stage method. In the first stage, the efficiency is estimated, and, in the second, stage the estimated efficiencies are regressed on a vector of explanatory variables.[168]

The JLMS method estimates u_{it} as

$$E[u_{it}|\varepsilon_{it}] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(a_{it})}{1-\Phi(a_{it})} - a_{it}\right] \qquad (3.12)$$

where

$$\sigma = [\sigma_v^2 + \sigma_u^2]^{1/2}$$

$$\lambda = \sigma_u/\sigma_v$$

$$a_{it} = \pm\varepsilon_{it}\lambda/\sigma$$

$$\phi(a_{it}) = \text{standard normal density evaluated at } a_{it}$$

$$\Phi(a_{it}) = \text{the standard normal CDF evluated at} a_{it}$$

A single step method is proposed in Battese and Collie.[169] The explanatory variables in the single stage approach are incorporated in the random variable such that

$$y_{it} = \exp(x_{it}\beta + v_{it} - u_{it}) \qquad (3.13)$$

with the u_{it} term representing 'a function of set of explanatory variables,$z_{it}s$ and an unknown vector of coefficient, δ'.[170] Thus,

$$U_{it} = z_{it}\delta + W_{it} \quad (3.14)$$

[168] Kumbhakar and Lovell, *Stochastic Frontier Analysis*, p.10
[169] Battese and Coelli, 'A stochastic Frontier production'
[170] Ibid, p.327

The technical efficiency of the ith firm is defined as:

$$TE_{it} = \exp(-U_{it}) = \exp(-z_{it}\delta - W_{it}) \qquad (3.15)^{171}$$

Along with the development in the measurement of efficiency, the literature also witnesses the use of more sophisticated production functions over years. The earliest used production function is the Cobb- Douglas production function of the form,

$$Q = AK^{\alpha}L^{\beta} \qquad (3.16)$$

where 'Q' represents the quantity of output, K and L are capital and labour inputs respectively, and A, α and β the parameters. In this function, the elasticity of substitution is constant, and hence represents only constant returns to scale. It is incapable to measure the U shaped average cost curve. Also, it cannot accommodate multiple outputs. In order to overcome the former problem, the Constant Elasticity of Substitution (CES) production function[172] is used:

$$Q = A[\alpha K^{-\rho} + (1-\alpha)L^{-\rho}]^{\frac{-1}{\rho}} \qquad (3.17)$$

where ρ is a constant but not unity. The improvement that the CES production function provides over the C-D function is that since elasticity of substitution is constant but not unity, this functional form can depict either increasing or decreasing returns to scale. However, the limitation of the CES and C-D function is that both are monotonic function. The C-D function gives monotonically constant returns to scale, and the CES production function gives monotonically either increasing, decreasing or

[171]Ibid, p. 327
[172] Arrow *et al*, 'Capital-labour substitution'

constant returns to scale. Moreover, both of them can analyse only a single-product firm.

Recent studies have been using flexible functional forms to approximate the production technology by the Taylor or Fourier series. The Transcendental Logarithmic (translog) cost function[173] is a Taylor series expansion to accommodate various outputs separately and do not enforce any particular returns to scale. It accommodates multiple outputs without violating the curvature condition; also, it is flexible in the sense of providing a second order approximation to any well-behaved cost frontier at the mean of data. This functional form is now widely used in empirical studies, including those on the decomposition of cost efficiency.

The translog specification of a cost function is

$$LnC = \alpha_0 + \sum_{i}^{n} \alpha_i Lny_i + \sum_{j}^{m} \beta_j Lnp_j + \frac{1}{2}\sum_{i}^{n} \sigma_i Lny_i^2 + \frac{1}{2}\sum_{j}^{m} \gamma_j Lnp_j^2$$

$$+ \sum_{i}^{n}\sum_{j}^{m} \Omega_{ij} Lny_i Lnp_j + \varepsilon \qquad (3.18)$$

Where y_i (i= 1,2, ..., n) represents the ith output, and p_j (j= 1,2, ..., m) its price. The translog cost function, however, assumes certain regularity conditions like homogeneity and the concavity of cost function, and also certain behavioural assumptions like the objective of cost minimisation and the prevalence of perfect competition in both product and factor markets.

[173] Christensen *et al*, 'Transcendental Logarithmic Utility Functions'

The homogeneity condition is satisfied when

$$\Sigma_j \beta_j = 1 \qquad (3.19)$$

$$\Sigma_j \gamma_j = 0 \qquad (3.20)$$

$$\Sigma_i \Omega_i = 0 \qquad (3.21)$$

The major limitation of the translog function is that it gives only local maxima or local minima. The global optimum solution cannot be generated in this method.

Parallel with the increasing application of econometric tools and techniques, the literature also witnesses the growing use of programming-based methodology. Such a development is the Data Envelopment Analysis (DEA), pioneered by Charnes, Cooper and Rhodes (CCR),[174] which is based on mathematical programming, and, hence, non-parametric. This method was originally developed to measure the efficiency of non-profit organisation.[175] The units under study are referred to as the decision Making Units (DMU). The CCR model measures the efficiency of a particular unit (denoted by the subscript 0) as the ratio of its weighted outputs to its weighted inputs given as

$$h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \qquad (3.22)$$

The objective function of the unit is then

$$\max h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \qquad (3.23)$$

[174] Charnes et al, 'Measuring the efficiency of decision making units'
[175] Ibid, p.429

$Subjectto$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1$$

$$j = 1 \ldots \ldots N$$

$$v_r u_i \geq 0; \quad r = 1 \ldots \ldots \ldots s; \quad i = 1 \ldots \ldots . m$$

where y_{rj} is the output of the jth DMU and x_{ij} its inputs and the v_r, u_i are the weights assigned. The above equation can be transformed into a linear programming problem as follows:

$$MaxZ_0 = \sum_{r=1}^{s} u_r y_{r0}$$

$Subjectto$

$$\sum_{i=1}^{m} v_i x_{i0} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0$$

$$j = 1 \ldots N$$

$$u_r, \ v_i \geq \epsilon$$

where $\epsilon > 0$ is a small non Archimedean quantity.[176]

The dual of the above equation is given as

[176] Banker et al, 'Some Models for Estimating', p.1083

$$Min w_0 - \epsilon \left[\sum_{i=1}^{m} s_i + \sum_{r=1}^{s} s'_r \right] \qquad (3.24)$$

$subjectto$

$$0 = w_0 x_{i0} - \sum_{j=1}^{n} x_{ij} \lambda_j - s_i , \qquad i = 1, \dots \dots m$$

$$y_{r0} = \sum_{j=1}^{n} y_{rj} \lambda_j - s'_r r = 1, \dots \dots . . s$$

$$\lambda_j , s_i, s'_r \geq 0$$

$Z_0 = w_0 = 1$ if and only if the slack values s_i, s'_r are equal to zero.[177]

The model developed by CCR is based on constant returns to scale, and, therefore, it is not an appropriate measure of efficiency when variable returns to scale prevail. Banker, Charnes and Cooper (BCC)[178] improvise it by incorporating variable returns to scale. For a single input (X), single output (y) model of production, the following diagram gives the essence of the model.

[177] Ibid, p.1083
[178] Ibid, p.1078

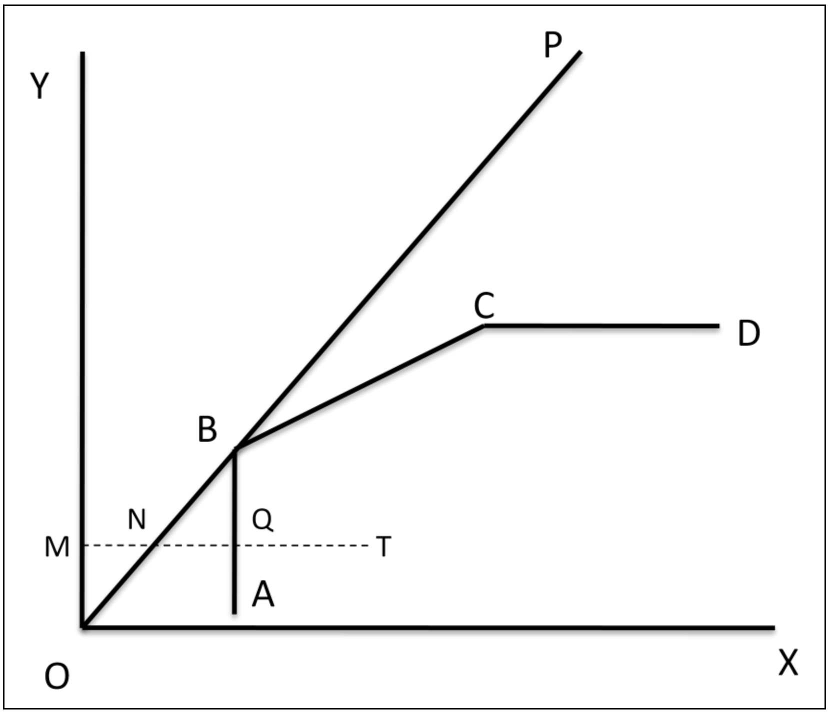

Fig 3.1: Technical and Scale Efficiency

The figure displays the production frontier by ABCD when the technology gives variable returns to scale, and by OBP when the constant returns to scale prevail. For any DMU, the overall technical inefficiency is measured by ${}^{\mathrm{MN}}/_{\mathrm{MT}}$ for the production point at T. Input technical efficiency is measured by the ratio ${}^{\mathrm{MQ}}/_{\mathrm{MT}}$ and the input scale efficiency is measured by the ratio ${}^{\mathrm{MN}}/_{\mathrm{MQ}}$.

The fractional programming problem for BCC is given as

$$\max h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0} - u_0}{\sum_{i=1}^{m} v_i x_{i0}} \qquad (3.25)$$

$subjectto$

$$\frac{\sum_{r=1}^{s} u_r y_{rj} - u_0}{\sum_{i=1}^{m} v_i x_{ij}} \leq 0$$

$$u_r, v_i \geq 0$$

Its linear program form can be written as

$$\max h_0 = \sum_{r=1}^{s} u_r y_{r0} - u_0 \qquad (3.26)$$

$subjectto$

$$\sum_{i=1}^{m} v_i x_{i0} = 1$$

$$-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} - u_0 \leq 0$$

$$j = 1, \dots \dots ., n$$

$$u_r, v_i \geq \epsilon$$

$u_0$is unconstrained in sign.

The dual of the problem can be written as

$$\min h - \epsilon \left[\sum_{i=1}^{m} s_i^+ + \sum_{r=1}^{s} s_r^- \right] \qquad (3.27)$$

$subjectto$

$$hx_{i0} - \sum_{j=1}^{n} x_{ij} \lambda_j - s_i^+ = 0 \qquad i = 1, \dots \dots .., m$$

$$\sum_{j=1}^{n} y_{rj} \lambda_j - s_r^- = y_{r0} r = 1, \dots \dots \dots . s$$

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

$$\lambda_j, s_i^+, s_r^- \geq 0$$

A special case of the DEA is the Free Disposal Hull (FDH) approach. In the case of FDH, the main improvement is that 'the production possibility set is composed of DEA vertices'.[179] The specification, however, generates inefficiency which is greater than that generated by the DEA.

In addition to SFA and DEA, the other approaches in the field of parametric method are the Distribution Free Approach (DFA) and the Thick Frontier Approach (TFA). The DFA was introduced by Berger in 1993.[180] It assumes 'that the efficiency of each firm is stable over time'[181] and the random error overtime tends to zero. Therefore, there is no random error or residual in the final analysis. The TFA was developed by Berger and Humphrey in 1991.[182] This method recognises the residuals like the SFA, but groups the data into quartiles and estimates the parameter separately for each quartile. The measurement of efficiency is then made on the basis of highest and lowest quartiles.

[179] Berger and Humphrey, 'Efficiency of financial institutions', p. 5
[180] Berger, 'Distribution-free estimates of efficiency'
[181] Berger and Humphrey, 'Efficiency of financial institutions', p.7
[182] Berger and Humphrey, 'Dominance of inefficiency'

Section IV: Studies relating to efficiency

Literature in the field of efficiency measurement can be classified into two major groups, the DEA from the non-parametric side and the SFA from the parametric side. Other measures are less frequently used. In this section, we first review the efficiency measurement studies, belonging to both DEA and SFA, outside the field of commercial banking. Then we concentrate on similar studies in banking, but our domain of empirics is outside India. Finally, we take up the studies on Indian Banking.

Danilin *et al* [183] use SFA to estimate the inefficiencies in the command economy of the Soviet Union. They analyse the efficiency for a cross-section of 151 cotton refining enterprises in 1974, using the dynamic homothetic CES production function with capital and labour as inputs. Much against the popular belief of the western thinkers that the command economy generates inefficiency, they find that the cotton refining firms are close to the efficiency frontier. The inter-enterprise efficiency score is also very negligible for them.

Stevenson[184] analyzes the X-efficiency of the electricity generating industryin the United States. He studies the effect of competition on efficiency, using the translog cost function with input prices for labour, capital and fuel, as also the utilisation of capacity and competitive pressures. The difference in competitive pressure is found giving rise to different levels of technical efficiency as well as

[183] Danilin *et al*, 'Measuring enterprise efficiency'
[184] Stevenson, 'X-inefficiency and inter firm rivalry'.

dynamic efficiency. The article interchangeably uses the terms technical efficiency and X-efficiency.

Kumbhakar and Heshmati[185] adopt the SFA to analyse the technical efficiency of Swedish dairy farms, decomposing the residuals at multi stages, rather than in a single step. Using fodder and other materials, land, labour and capital as inputs, and also the age of farmers and the time trend as other explanatory variables, they reveal that the persistent technical inefficiency is higher than compared to the residual inefficiency.

The SFA is also used in Liberman and Dhawan[186] for Japanese and the US auto producers using the value added per employee as a proxy for labour productivity, capital stock per employee for Investment per worker, the firm size is proxied by the number of employee, Plant size is represented by the volume per plant, work-in-progress (WIP) represents the shop floor manufacturing capacity, vertical integration by the ratio of value added and sales, final product quality and cumulative output as dependent variables. The study underscores the existence of scale efficiency, and notices that higher efficiency is achieved by firms at higher production levels. The productivity differences among firms are evidently related to the type of organization and the scale of operation.

Further application of SFA is evident in Esho and Sharpe,[187] which measures the X-efficiency of Australia's Permanent Building Societies (PBS) using the cost of fund and wage index. But, the X-efficiency is taken to mean technical as well as

[185] Kumbhakar and Heshmati, 'Efficiency measurement'
[186] Liberman and Dhawan, 'Assessing the resource base'
[187] Esho and Sharpe, 'X-efficiency of Australian'

allocative efficiency. The evidence of X-inefficiency is noticed among Australia's PBSs, which they attribute, by logic, to managers' inefficiency.

The X-efficiency of Taiwan hotels is the area of interest in Chen,[188] where we find three input prices – the price of labour, the price of food and beverage, and the price of materials - total revenue as the output variable, and the room occupancy rate and the value of catering per unit space as control variables. The paper, however, represents the technology by the Cobb-Douglas production function in the SFA framework. His empirical exercise estimates the hotels' efficiency at 80 per cent, and suggests that management exerts a strong effect on the performance of the hotels. Other findings of the study are: (i) the efficiency of the hotel chains is higher than that of the independent hotels, and (ii) the scale of operation and the location of a hotel do not change the efficiency level.

Among other studies in this field, we should mention Olsen and Henningsen[189] who analyse the technical efficiency of commercial pig farms in Denmark using the translog specification. The study seeks to understand the effects of the size and timing of investment, as also the investment utilisation rate on level of efficiency. They observe a growth in farm efficiency in consequence of a higher level of investment and the experience level of farmers.

In the study of X-efficiency among commercial banks in Hong Kong by Simon Kwan[190], X-efficiency is linked to various aspects of the organization, like organizational structure, executive compensation, market concentration, risk-taking, merger and acquisition, and common stock performance. Using labour, physical

[188] Chen, 'Applying the stochastic frontier approach'
[189] Olsen and Hanningsen , 'Investment Utilisation'
[190] Kwan, 'The X-efficiency of Commercial banks'

capital and borrowed funds as inputs and loans as output in a translog cost function, he shows that the banks became more cost efficient overtime, especially after the Asian Financial crisis (1997). The X-efficiency is found decreasing with bank size, deposits-to-asset ratio, loan-to-asset ratio, provision for loan loss and the growth of loan, but increasing with off-balance sheet activities.

The X-efficiency of Australian Banks is studied by Sathye[191] using the data envelopment approach with three inputs - labour, capital and loanable funds - and two outputs, loans and demand deposits. The empirical analysis suggests that the technical efficiency is lower than the allocative efficiency, and that the former pulls down the overall efficiency of the Australian banks much below the world average. The study has also found that domestic banks are more efficient than foreign banks. There is, however, a lack of clarity in the concepts of various forms of efficiency. The allocative efficiency is regarded as X-efficiency, and the overall efficiency as economic efficiency. In fact, the concept of X-efficiency as considered here is not in line with what Leibenstein conceives.

The efficiency in Japanese Banking is studied by Altunbas *et al.*[192]using the Fourier flexible form in a SFA framework of three inputs (labour, fund and physical capital) and three outputs (loans, securities and off-balance sheet activities). The study indicates that the scale inefficiencies are larger than X-inefficiencies, so that overall efficiency could be augmented by way of policy measures relating to scale efficiency, rather than X-efficiency.

---

[191] Sathye, ' X-efficiency in Australian Banking'
[192] Attanubas *et al*, 'Efficiency and risk'

Kwan and Eisenbeis [193] investigate two important problems about U.S banking firms: (i) how inefficient firms continue in business, and (ii) whether such inefficient firms cause risks to the banking business. They employ the translog cost function using four output variables – (a) the book value of real estate loan, (b) the commercial and industrial loan, (c) the consumer loan and off balance sheet commitments, and (d) contingencies – and three price variables – (a) the unit price of capital, (b) total expenses and (c) the unit price of labour. Their findings are: (i) the large banks are more efficient than the smaller banks, (ii) The inefficient firms have higher stock return variance, (iii) higher idiosyncratic risk is associated with lower capitalization and higher loan losses, and (iv) the X-inefficiency of the firm brings down the stock returns of the firms.

Clark and Siems[194] delve into the off-balance sheet activities (OBS) to evaluate the X-efficiency of the US banking firms for 1990-1999, using the translog functional form in the SFA and DFA frameworks. In view of measurement problems for OBS, it is proxied by a credit equivalent measure ('standby letters of credit to derivative contracts')[195], an asset equivalent measure ('the rate of return on on-balance-sheet assets to capitalize noninterest income')[196] and a non-interest income measure ('the sum of income from fiduciary activities, service charges on deposit accounts, trading fees and gains or losses from foreign transactions, trading account gains or losses, fee income, and all other noninterest income').[197] These different OBS activities significantly influence the bank performances, especially their cost

[193] Kwan and Eisenbeis, 'An analysis of inefficiency in Banking'
[194] Clark and Seims, 'X-efficiency in Banking'
[195] Ibid p.995
[196] Ibid p.995
[197] Ibid, p.996

efficiency. Their profit efficiency is not, however, significantly augmented in the process.

The X-efficiency and the scale efficiency among the US banking firms also attract a study by Stavros Peristiani.[198] By way of the translog cost function and DFA, it reveals that X-efficiency is constant over banks of all sizes. Their X-efficiency, however, declines after the merger, but their scale efficiency improves.

The efficiency of Greek banking is studied in Christopoulos *et al* [199], who adopt a translog function defined on capital, labour and deposits as inputs, and loan investments and liquid assets as outputs. Their study denies the existence of scale efficiency among those banks. Their explanation is that the extended branch network of larger banks escalates their cost of operation. After entering the European Monetary Union (EMU), the Greek banking system was also highly affected by the cross-country competition that put pressure on bank profitability. The entry into the EMU, increased the liquidity of the banks as a result of the abolition of minimum reserves, reduction in bank investment in government securities to finance public deficit. The large banks are found to be less cost efficient than the small and medium size banks.

The Chinese banks are the subject of Chen *et al*,[200] where their performance is analysed with respect to deregulation. Using DEA they measure the cost efficiency, technical efficiency as well as allocative efficiency. Their empirical evidences include, (i) that the technical efficiency scores higher than allocative and cost efficiencies, (ii) that the allocative and technical efficiencies shot up during the reform period, (iii) that the Asian Financial crisis of 1997 greatly enhanced the efficiency of

---

[198] Peristiani, 'Do Mergers improve the scale efficiency'
[199] Chiristopoulos *et al*, 'Efficiency of the Greek banking'.
[200]Chen *et al*, 'Banking efficiency in China'

regional joint-equity banks, and (iv) that the medium sized banks were least efficient among banks of various sizes.

The CAMEL methodology is used by DeYoung[201] to categorise the U.S. banks into two groups, well-managed and poorly managed. The underlying methodology is the Thick Frontier Analysis where the translog cost function is used with three outputs, loan, transaction services and fee-based activities, and three input prices, wage rate, price of physical capital and interest on deposits. He finds that the well-managed banks operate with a relatively less number of branches than a poorly managed bank and that the 'well-managed commercial banks out-perform banks with poor management'. [202]

Spong *et al*[203] examine the effects of management and ownership on the efficiency of banks for the Tenth Federal Reserve District. The cost efficiency is measured by a cost efficiency index, and the profit efficiency by adjusted-returns on average assets. Among various findings in this study, we may note: (i) the banks in holding-companies are inefficient than the independent banks, (ii) the efficient banks are characterised by better attendance rate, more board meetings, and active deliberation of the board members on the workings of the banks, (iii) the efficient banks can effectively control their costs of operation, and (iv) managers' compensations in efficient banks are also greater than those in inefficient banks.

Allen and Rai[204] focus on the efficiency of international banks on the premise that international competition would compel the banks to minimise their cost – an idea associated with the X-efficiency. They employ both the SFA and the DFA to

[201] DeYoung , 'Management Quality and X-inefficiency'
[202] Ibid, p.19
[203] Spong *et al*, 'What makes a Bank efficient'
[204] Allen and Rai, 'Operational efficiency in Banking'

measure cost efficiency using the translog cost function. Their study covers fifteen countries, each with 194 banks, for a period of five years. The study underlines that the countries where universal banking is dominant are less X-inefficient than those where commercial and development banking are separated. Their further findings are that, though the bank size is an insignificant variable to control efficiency among universal banks, the X-inefficiency grows up with the size of the banks.

D'Souza and Lia[205] analyse the effect of diversification on the efficiency of the Canadian banks. The efficiency is measured by constructing a risk-return efficient frontier,[206] and the diversification is proxied by four different types of the Herfindahl-Hirschman Index (HHI), namely, the industrial HHI, the regional HHI, the business line HHI and the financing HHI.[207] They report that diversification decreases bankers' risk, and that the banking efficiency is irresponsive to their sizes.

Altunbas *et al*[208] have estimated the scale, technical and X-efficiency of the banks in the European Union. They employ the Fourier functional form in SFA using total loans, total securities and off-balance sheet activities as outputs and labour, physical capital and deposits as inputs. Cross-country variations in efficiency are found significant, and, among them, the banking sector of Austria, Denmark, Germany and Italy are found most efficient. There is no evidence to suggest that large banks are more efficient. The study, however,suggest thatscale efficiencies could be achieved by expanding bank operations, and that the overall changes in the managerial, technological and other factors have an impact on the X- efficiency of those banks.

[205] D'Souza and Lia, 'Does Diversification Improve the Bank efficiency'
[206] Ibid, p.109
[207] Ibid, p.113
[208] Altunbas *et al*, 'Efficiency in European banking'

Hassan *et al* [209] use the DEA to calculate the allocative, technical, scale and overall efficiency in the US banking. Prices of labour, capital and loanable funds have been taken as inputs and the real estate loans, commercial and industrial loans, consumer loans, all other loans and demand deposits as output. They conclude that the US banks are characterized by low levels of overall efficiency, and that their allocative efficiency is greater than their technical efficiency.

Cebenoyan *et al*.[210] measures the 'agency-related inefficiency problems'[211] or X-inefficiency problems in the thrift industry of Atlanta Federal Home Loan Bank District. They employ the SFA and translog cost function, using construction loans, permanent mortgage loans, mortgage backed securities and other loans as output, and prices of capital, deposits and labour as inputs. They find that the inefficiency scores vary widely amongst the Savings and Loans (S&Ls) and that the inefficiency is not related to the form of ownership.

Mester[212] too analyses the agency related problem in savings and loan industry in the U.S. using translog cost function in SFA. She finds that the stock S&Ls are more efficient than mutual S&Ls and the increase in competition has led to a reduction in X-inefficiency.

In yet another article,[213] Mester analyses the efficiency of the U.S. banks taking into consideration the quality and risk of banks output.Using translog cost function in SFA and the prices of labour, physical capital and borrowed money as the inputs , and loans as the output , she finds that there are significant variations in the efficiencies of various districts.

[209] Hassan *et al*, 'Technical, Scale and Allocative efficiencies'
[210] Cebenoyan *et al*, 'The relative efficiency'
[211] Ibid, p.152
[212] Mester, 'Efficiency in the Savings and Loan Industry'
[213] Mester, 'Measuring efficiency at US Banks'

Huang[214] measures the X-efficiency of Taiwanese banks using the translog shadow profit function, using two outputs (namely, investment and loans), and two inputs (namely, borrowed fund and labour). He finds that the loss in profit due to X-inefficiencies is high for both private and public banks, and that the technical inefficiency is greater than the allocative inefficiency.

The effect of competition on the efficiency and soundness of the banks in Europe and the U.S have been studied by Schaeck and Cihak.[215] Competition is measured by the Lerners index and the soundness of the banks by the Z score. They employ the translog profit function and the cost function in the SFA framework, with loans and other earning assets as the output, labour cost and other costs as input, and fixed assets, loan loss provision and equity capital as the netputs. They find that the efficiency of the banks increases with the higher level of competition, which, in turn, increases the soundness of the banks.

Similar study has been conducted for the Czech banks by Podipera et al.[216] Employing the SFA, cost efficiency is measured using the translog cost function, where the prices of labour, physical capital and borrowed fund are taken as inputs, and loans as the output. In contrast to the results found in the studies on the US and European banks, a negative relationship are found between efficiency and competition in case of the Czech banks.

For the Italian Banks, the relationship between competition and efficiency is studied in Coccorese and Pellecchia.[217] They use the SFA and the translog cost function to measure the efficiency of the banks. Using the price of labour, deposits

---

[214] Huang, 'Estimating the X-efficiency'
[215] Schaeck and Cihak, 'How competition affects efficiency'
[216] Podpiera *et al*, 'Banking competition and efficiency'
[217] Coccorese and Pellecchia, 'Testing the Quite Life Hypothesis'

and capital as inputs and the total assets as output, they find that a higher degree of market power is associated with the cost inefficiency.

Similar results are found for the German banks by Koetter and Vins[218] who employ the translog cost and profit functions using the SFA. But, for the profit efficiency, a positive relationship is witnessed with market power.

Casu and Girardone[219] have also analysed the impact of competition on the efficiency of European Union banks. They use both DEA and SFA for estimating the efficiency. The loans and other earning assets are regarded as outputs in both the models. The inputs in SFA include the price of deposits, the price of labour and the price of capital, whereas, in DEA, the total cost is the only input. They infer that the EU banks become less cost efficient as more consolidation takes place.

The effect of managerial ability on the efficiency of banks in Finland is analysed by Kauko.[220] The cost efficiency of the banks has been measured using the translog cost function in SFA. He finds that the managers' ability in containing cost is significant. The age and qualification of the managers are also found to have impact on the their efficiency levels.

Fare *et al*.[221] have applied DEA to analyse the impact of competition on the efficiency of Spanish banks. Labour, capital and purchased funds are taken as input variables while loans, fixed-income securities, other securities and non-traditional output (Non Interest Income) are regarded as the output variables. The inefficiency of the Spanish banks is found to have been growing over time, and, on the efficiency, the effects of competition vary across the banks.

[218] Koetter and Vins, 'The Quite life hypothesis in Banking'
[219] Casu and Girardone, 'Does competition lead to efficiency'
[220] Kauko, 'Managers and efficiency in Banking'
[221] Fare *et al*, 'Revisiting the quite life hypothesis in banking'

Solis and Maudos[222] evaluate the welfare loss for the market concentration in Mexican banking on the basis of the SFA and the translog cost function. Using the prices of labour and capital as the inputs, and loans and deposits as the output, they observe that the cost efficiency increases with the market concentration.

Chang *et al*. [223] study the efficiency of the U.S owned banks and the foreign owned banks in the U.S. economy. Using the translog cost function in SFA with the prices of labour, physical capital and funds as the inputs, and loans and assets the output, they find that the U.S owned banks are more efficient than their counterparts, i.e. foreign owned banks.

Mitchell and Onvural apply the Fourier functional form to study the cost and scale efficiency of the U.S. banks using SFA. They find that the larger banks are more cost efficient than the small ones, and that there is considerable ray scale economies among them.

In the domain of Indian banking, the studies may be grouped in two categories, DEA and SFA. To the former category belongs the studies of Das *et al*[224], Rammohan and Ray[225], Reddy[226], Ghosh[227], Dash and Charles[228], Varadi et al[229], Kumar and Gulati[230], Mariappan et al[231], Bhatia and Mahendru[232], Jayaraman and Srinivasan[233], Dwivedi and Charyulu[234], and Seshadri et al.[235]

---

222 Solis and Maudos, 'The social cost of bank market power'
223 Chang *et al*, 'Efficiency of multinational banks'
224 Das *et al*, 'Liberalization, Ownership, and Efficiency'
225 Rammohan and Ray, 'Comparing the performance'
226 Reddy, 'Banking Sector Reforms'
227 Ghosh, 'Financial Deregulation and Profit efficiency'
228 Dash and Charles, 'A study of technical efficiency of banks in India'
229 Varadi *et al*, 'Measurement of efficiency of banks in India'
230 Kumar and Gulati, 'An Examination of technical'
231 Mariappan, *et al*, 'A study on performance efficiency'
232 Bhatia and Mahendru, 'Assessment of technical efficiency'
233 Jayaraman and Srinivasan, 'Performance Evaluation of Banks in India

In the field of SFA, the literature is relatively sparse; it includes Sensarma[236] Shanmugan and Das[237], Srivastava[238], Das and Drine[239], Mahesh[240], Rajan et al.[241] Battachrayya et al[242] and Bhattachryya and Pal[243]have used both DEA and SFA while, Bhaumik and Dimova[244] have used ordinary least square regression (without decomposing errors, thus remaining outside SFA).

The input and output sets, however, vary across these studies, depending upon, *inter allia*, whether banks are considered as producers or intermediaries (to be discussed in the following chapter). In DEA, the inputs that have generally been used are equity capital, borrowed funds, number of employee, number of branches, operating expenses, deposits, interest expenses, physical capital (Jayaraman , Dwivedi and Charyulu, Bhatia and Mahendru, Kumar and Gulati, Ghosh, Rammohan and Ray and Bhattacharyaa et al). The output variables include loans, non-interest income, loans and advances, investment and net interest income. In the studies relating to SFA, more frequently used inputs are labour, capital and purchased fund (Srivastava, Das and Drive, Mahesh, Rajan et al, Shanmugam n Das, Ray and Sanyal) while the dominant outputs are loans and advances, deposits and investments. A further field of difference in SFA concerns about the underlying functional form; it varies from the Cobb-Douglas function as in Bhattachryya and Pal, and Shanmugam and Das to the translog specification as in Sensarma, Ray and Sanyal , Mahesh and Srivastav . Fourier flexible form is used by Das and Drine.

---

[234] Dwivedi and Charyulu, Efficiency of Indian banking Industry
[235] Seshadri *et al*, Efficiency of public '
[236] Sensarma, 'Cost and Profit efficiency '
[237] Shanmugam and Das, 'Efficiency of Indian commercial banks'
[238] Srivastava, 'Size, Efficiency and Financial Reforms'
[239] Das and Drine, 'Financial Librelization'
[240]Mahesh, 'Liberalisation and Efficiency'
[241] Rajan *et al*, 'Efficiency and productivity Growth'
[242] Battachrayya *et al*, 'The Impact of Liberalisation'
[243]Bhattachryya and Pal, 'Financial reforms'
[244] Bhaumik and Dimova, 'How important is Ownership'

The results of the studies are not unanimous. Major reasons for such variations are differences in the time-frame, methodology, type of organisation under study, as well as the concept of efficiency that various studies consider. The public sector banks are found profit efficient in Jayaraman and Srinivas and Das et al, but, according to Mahesh, their profit efficiency is less than that of private domestic banks. A number of studies – such as Das and Drine, Mahesh, Varadi et al., Sensarma and Bhattachryya et al – reveal that the public sector banks are the most cost efficient. But contradictory results are found in Dash and Charles. In an intra public-sector- banks analysis, Shanmugam and Das, however, note that the State Bank group scores the highest level of efficiency, compared to other public sector banks in India. While searching for reasons behind low efficiency among the public sector banks, Bhatia and Mahendru identify that their efficiency is pulled down by their non performing assets.

Taking all banks in India together, Mariappan and Lakshmi carry out the efficiency study. They underscore that their efficiency scores belong to the inner domain of the frontier so that there is much scope for gain in efficiency. In a similar study by Dwivedi and Charyulu, we come across the findings that banks' technical efficiency moved upwards till 2008, but the trend has reversed thereafter.

Section V: Conclusions

Given the Review of Literature, let us now clarify the rationale behind the present study. The first justification relates to the period of investigation which ranges from 1994 to 2012. The choice of period provides an opportunity to assess the impact of the financial sector reforms in India. The period between1994-2012 saw the effect of not only the first generation of economic reforms but also the second generation reform initiated in 1998. By the year 2012 we can capture the effects of the two reforms together.

Although the literature points out that the improvements in technical and allocative inefficiency lead to a very miniscule gain in efficiency,[245] majority of the empirical studies focus on the technical and allocative inefficiencies of the banks. However, the focus of this study is the X-efficiency of the banks. We note in this connection that the financial sector reforms have not only introduced quantitative changes in the working of the banks (such as the reductions in CRR and SLR) but have also brought about changes in the management of the banks. From the year 1994 to 2012 the number of commercial banks rose significantly under the relaxed norms of entry into banking business. The banks now not only face competition from the other banks but also from other financial institutions as the financial products overlap. The effects of those changes can be analysed by studying the X-efficiency of the banks rather than their technical aspect of efficiency.

We add that this study takes into account all commercial banks, save a few for the want of data, irrespective of their ownership pattern. While analyzing the data, we group them in three categories, public sector banks, private domestic banks and private foreign banks, so that the effects of ownership pattern on the level of efficiency are examined. Lastly, the methodology that is used in the study i.e. the stochastic frontier model is relatively less used than the DEA method. The study, therefore, seeks to measure X-efficiency among Indian commercial banks from 1994to 2012 using the SFA.

[245] Leibenstein, 'Allocative efficiency vs "X-efficiency"'