SBP Research Bulletin
Volume 10, Number 1, 2014

ARTICLES

Cost Inefficiency in the Pakistan Banking Sector 2002-2009

Kent Matthews[1]

Abstract: *This paper uses the Simar-Wilson bootstrap technology to estimate cost inefficiency in the Pakistan banking sector for the period 2002-2009. Several models of outputs including bad output are considered alongside a common set of inputs. Cost inefficiency is decomposed into its Technical and Allocative inefficiency components. Panel regression methods are used to model the drivers of inefficiency. In general the findings suggest that inefficiency is declining over time and that there is strong conditional convergence to peer group clusters based on branch levels, ownership and specialism. It is found that in general banks with more branches have higher cost inefficiency, the one foreign bank operates on lower cost inefficiency and Islamic banks have higher allocative inefficiency which is offset by a lower technical inefficiency.*

JEL codes: D23, G21, G28

Keywords: Cost inefficiency, Pakistan banking. DEA. Bootstrapping

1. Introduction

Bank efficiency is not the sort of subject that sets the pulse racing or the heart beating faster. On the surface the subject sounds dry and could put even the most ardent accountant or banker to sleep. Certainly, efficiency is not the hottest topic in banking in the current economic climate. Improved risk management, stronger capital adequacy, regulation, and banker's remuneration stand higher in the scale of importance following the public opprobrium of bankers in the West. However, I hope to show that not only is this subject worth exploration in its own right but there is much insight that can be gleaned from its study that tells us something about the banking market in question.

[1] Cardiff Business School Cardiff University Cardiff CF10 3EU Wales, UK. matthewsk@cardiff.ac.uk. This paper was the one circulated for the Zaid Husain Memorial Lecture on 23 October 2010 in Islamabad, Pakistan. I am grateful without implication to Tiantian Zhang and Momna Saeed for able research assistance in the production of this paper, to the editor and two anonymous referees for helpful comments.

The efficiency of banks, relates to the efficiency of the banking market which in turn relates to the efficiency of the intermediation process and the efficiency by which monetary policy passes through to bank lending. Studies have also used measures of bank efficiency to explain the positive correlation (where it exists) between concentration and profitability as a challenge to the argument that concentrated markets generate anti-competitive behaviour[2]. Other studies find a link between competitiveness and bank efficiency[3]. For emerging markets, the issue of bank efficiency has particular importance given the trend in deregulation and economic reform of recent decades.

This paper reviews the different ways to measure bank efficiency and highlights the results of research on bank efficiency in Asian emerging economies. In particular it will outline the extent of research thus far conducted on the efficiency of banks in Pakistan and comment on how to build and improve upon them.

The next section will examine the concept of economic efficiency. Section 3 speculates on the possible reasons for the existence of inefficiency. Section 4 reviews the literature of bank efficiency in the Asian emerging economies. Section 5 presents some measures of cost efficiency for Pakistan. Section 6 concludes.

2. Efficiency

The concept of economic efficiency comes easily to the economist. Given an economic objective and information on relative prices, an individual optimum is defined as a profit maximising objective given input and output prices, or cost minimisation given factor inputs and input prices. Under certain optimistic or restrictive assumptions, economic efficiency for the unit is generalised into an equilibrium that can be construed as a socially efficient equilibrium (Koopmans, 1951; Lange, 1942).

From Farrell (1957), economic efficiency can be separated into technical efficiency and allocative efficiency.[4] The formal definition of technical efficiency according to Koopmans (1951) is a case where an increase in any output requires a reduction in at least one other output or an increase in at least one other input. Similarly a reduction in any input requires an increase in at least one other input or a reduction in at least one output. From the definition of technical efficiency comes technical inefficiency which is a position where a producer could produce the same output with less of at least one input or use the same inputs to produce more of at least one output. The Koopman (1951) definition of efficiency can be generalised as a multi-output, multi-input production technology expressed by a transformation function:

$$T(x,q)=0 \qquad (1)$$

[2] For a discussion of the relationship between market structure and market power see Cabral (2000)

[3] See for instance Al-Muharrami and Matthews (2009) for an application to GCC economies.

[4] An excellent introduction to production economics and a review of the concepts of efficiency used here can be found in Kumbhakar and Lovell (2000) and Coelli et al. (2005)

where $q = (q_1, q_2, \dots q_M)'$ is an Mx1 vector of outputs and $x = (x_1, x_2, \dots x_N)'$ is an Nx1 vector of inputs. Shephard (1953) defines the input distance function:

$$d(x,q) = \max[\lambda : (x|\lambda) \in L(q)] \tag{2}$$

which translated into layman's language means that the distance between the minimum combination of $[x']$ and actual $[x']$ used to produce $[q']$ is at a minimum. A measure of technical efficiency is the ratio of the minimum to actual input or:

$$TE(q,x) = d(x,q)^{-1} \tag{3}$$

Now suppose the economic objective is to minimise the costs of producing $[q]$ facing an input price vector $w = (w_1, w_2, \dots w_N)'$. The cost minimisation problem is:

$$c(w,q) = \min_x w'x \text{ such that } T(x,q) = 0 \tag{4}$$

A measure of cost efficiency CE is provided by the ratio:

$$CE(q,w,x) = \frac{c(w,q)}{w'x} \tag{5}$$

Once cost efficiency and technical efficiency is obtained a measure of input allocative efficiency in the sense of Farrell (1957) is obtained as:

$$AE(q,w,x) = \frac{CE(q,w,x)}{TE(q,x)} \tag{6}$$

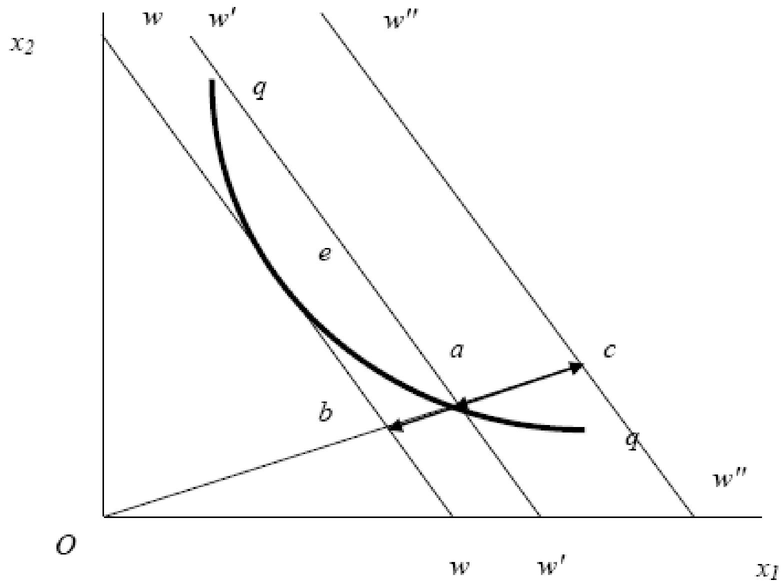
A diagram explains this concept a lot easier. Figure 1 shows an isoquant *qq* producing a single output with factor inputs x_1 and x_2 and isocost *ww*, which traces the ratio of factor prices. The efficient cost minimising position is shown at *e* where *ww* is tangential to *qq*. However, employing a factor combination shown by point *c*, which is to the right of the isoquant *qq* indicates that the unit is technically inefficient. Allocative inefficiency is generated by the employment of the factor mix that is inconsistent with the cost minimising factor mix. Technical efficiency is measured by the ratio *Oa*/*Oc* and technical inefficiency (TIE) is given by *ac*/*Oc*.

$$TIE = 1 - \frac{Oa}{Oc} = \frac{ac}{Oc}$$

The cost to the firm is shown by *w''w''* which is parallel to *ww* and passes through point *c*. Cost efficiency (CE) is measured by *Ob*/*Oc* and AE is obtained residually as

Ob/Oa. It follows that cost inefficiency (CIE) is described by *bc/Oc* and allocative inefficiency (AEI) is *ab/Oa*.

Figure 1: Technical Efficiency and Allocative Efficiency



Once the theoretical concepts have been established the next stage for the researcher is to measure inefficiency and this is where many of the problems begin. The research on banking efficiency has taken one of two approaches – the parametric and the non-parametric. The parametric approach requires the specification and estimation of a cost function or production function. The problem with the parametric approach is that errors arise due to misspecification of the function and the underlying stochastic process.

The non-parametric approach uses linear programming techniques to envelope observed points of the ratio of weighted outputs to weighted inputs. This latter method has been termed Date Envelopment Analysis (DEA). Unlike the parametric approach the DEA approach does not require a specification of an econometric model. It simply uses the observed data to define an efficient frontier as the envelopment of 'best practice'. The drawback of the DEA method is that the deviation from the efficient frontier represents 'inefficiency' and not a combination of inefficiency, measurement error or random error. What this means is that the estimates of cost inefficiency obtained from the DEA approach is not amenable to statistical inference. For example, if cost inefficiency of a firm or unit is said to be 10 percent less than best practice, in what sense is this number significant from a statistical viewpoint? Thankfully, recent innovations in the literature of estimation of efficiency have been able to answer this question.

Simar and Wilson (1998, 2000a, 2000b, 2008) argue that the deterministic DEA can produce estimates that suffer from 'finite sample bias' and propose a bootstrap procedure for non-parametric frontier models. Bootstrapping is based on the notion that if the data can be viewed as a random sample from an underlying population under a specific model (data generating process - DGP), then the process of continuous random draws from the sample under the model generates also random draws from the population. The random raw can be viewed as a pseudo-sample and as a group of new benchmarks to compute the efficiency score for a given point.[5]

In the case of parametric models the principal method of estimation of cost efficiency is the method of stochastic frontier analysis (SFA).[6] Cost efficiency is obtained as the ratio of the estimated cost function which will represent the best practice frontier and the actual cost of a specific firm. Hence in the case of a bank $\{i\}$ that produces $\{k\}$ outputs using $\{j\}$ inputs, cost efficiency is:

$$CE_i = \frac{c(w_{i,j}, q_{i,k})}{c_i} \qquad (7)$$

Equation (7) defines cost efficiency as the ratio of minimum cost attainable to observed expenditure $\{c_i\}$. So $CE_i \leq 1$ and $CE_i = 1$ holds only if bank $\{i\}$ is 100 percent efficient.

Research effort and energy has concentrated on the specification of the function $c(w_{i,j}, q_{i.k})$ [7] and the overall error term associated with the econometric model $\{\varepsilon_i\}$. The function to be econometrically estimated would in implicit form be described as:

$$\begin{aligned} &\ln\left[c(w_{i,j}, q_{i,k})\right] + \varepsilon_i \\ &\varepsilon_i = v_i + u_i \end{aligned} \qquad (8)$$

The properties of the components of $\{\varepsilon_i\}$ are that v_i is normally distributed with zero mean and fixed variance and:

---

[5] Recent applications of the bootstrap approach to banking have been in Casu and Molyneaux (2005), Dong and Featherstone (2006) and Matthews, Guo and Zhang (2007).

[6] In an Appendix to this paper, Momna Saeed of the SBP presents the results of an up to date SFA model of cost efficiency for Pakistan banks.

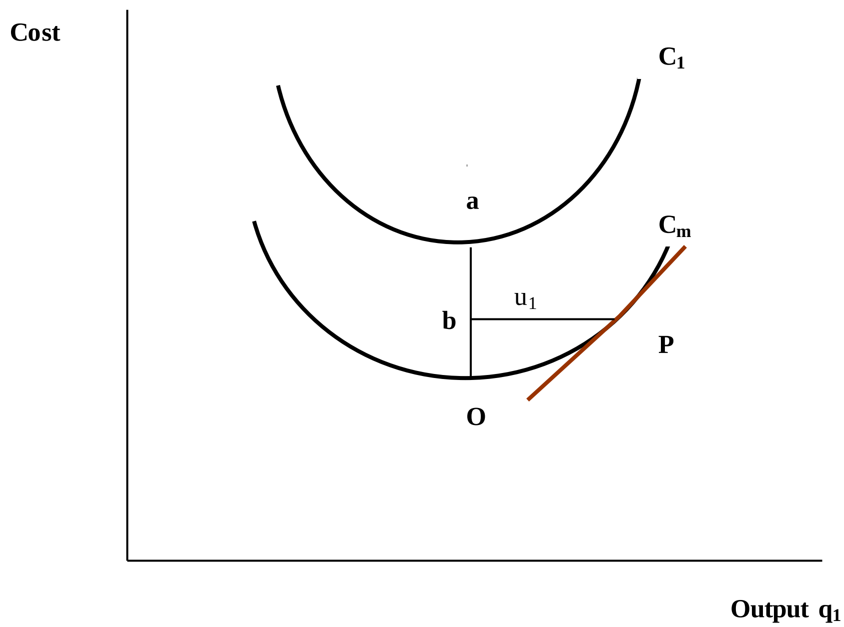
[7] Typical specifications in the literature are the Cobb-Douglas function, constant elasticity of substitution, and trans-log function. Kumbhakar and Lovell (2000) discuss these and also other specifications in Chapter 4 of their book.

$$u_i = h_i u_i^*$$
$$h_i = f(z_i) \qquad (9)$$

The term {μ$_i$} captures the extent of inefficiency and the function $f(z_i)$ describes the process that determines the inefficiency in terms of some specific 'environmental' variables $\{z_i\}$. Again it is easier to explain these concepts with the aid of diagram.

Figure 2 shows the cost function for a single output $\{q_1\}$. The benchmark or minimum cost function is described by $\{c_m\}$. The cost function for bank {1} is described by $\{c_1\}$.

Figure 2: Cost Function and Cost Efficiency



The minimum cost position is shown at point O on the 'best practice' cost frontier. The cost inefficiency of bank 1 is the vertical distance between O and point 'a' on the cost function associated with bank 1. The problem with the conventional SFA technique is that it estimates the overall cost inefficiency but does not decompose the estimate into its technical inefficiency and allocative inefficiency components. In principle Kumbhakar and Lovell (2000) describe the problem as estimating the function:

$$\ln\left[c(w_{i,j}, q_{i,k})\right] + \varepsilon_i$$
$$\varepsilon_i = v_i + u_{T,i} + u_{A,i} \qquad (10)$$

Where $u_T \geq 0$ represents the cost of input oriented technical inefficiency and the error component $u_A \geq 0$ represents the cost of input allocative inefficiency. If the latter error component can be identified then the point 'P' on the minimum cost function $\{c_m\}$ describes the above minimum cost generated by allocative inefficiency. Cost inefficiency 'Oa' can be decomposed into technical inefficiency 'ba' and allocative inefficiency 'Ob'. In reality the estimation of allocative inefficiency from SFA requires overly restrictive assumptions and a simple method has as yet eluded the literature.

3. Why Inefficiency?

Studies of bank efficiency show that inefficiency tends to exist over long periods. Why does inefficiency exist and why does the market not provide a solution through a takeover mechanism whereby inefficient banks are taken over by efficient ones? It is possible that the banking market, particularly in emerging markets, are protected from hostile takeover by government and official agencies. The knowledge that official protection exists could generate 'satisficing' managerial objectives that are not consistent with profit maximising behaviour. However, enforced merger is a strategy that central banks have employed in the aftermath of banking crises – see for example Daley et al. (2008) in the case of Jamaica.

Cost inefficiency relative to 'best practice' is usually blamed on bad management and poor motivation. Following Leibenstein (1966) this efficiency gap is termed 'X-inefficiency'. Studies of bank efficiency have used the terms technical efficiency and X-efficiency interchangeably as if they were the same thing. While similar in concept they are not necessarily the same. The concept of technical efficiency derives its basis in the neoclassical theory of the firm and assumes profit maximising behaviour. A firm or a bank may be technically inefficient for technical reasons such as low training or low human capital levels of managers and workers, or the use of inferior or out-of-date technology. The diffusion of new technology is not instantaneous and some firms or banks may lag behind others in the acquisition and utilisation of new technology. With further training and updating of capital, the firm or bank can expect to move towards the efficient frontier described by the isoquant in Figure 1. X-inefficiency is not caused by the variability of skills or the time variability of technology diffusion but by the use and organisation of such skills and technology.

In an earlier generation of studies of US banks, Berger, Hunter and Timme (1993) argue that X-inefficiency constitutes 20 percent or more of bank costs. Poor motivation and weak pressure resulting in under utilization of factors of production, are part of what Leibenstein (1975) describes as 'organisational entropy'. X-inefficiency arises as a result of low pressure for performance. Some institutions would be protected by government regulation that would reduce the external pressure of competition. But even with a higher degree of pressure from the environment, firms may have organisational deficiencies so that management signals and incentives are lost in the hierarchy of the organisation.

An alternative interpretation of X-inefficiency is ‘rent seeking’ in the sense of Buchanan (1980) and Tullock (1967, 1980). Rent seeking in its basic form is the appropriation of surplus in the process of production or exchange without any real contribution to the process of either. Where there are government regulations on enterprise, barriers to entry and other anti-competitive rules, officials have the opportunity to extract rents through the mechanism of bribery and corruption. Therefore the term rent seeking has been generally associated with extortion, bribery and corruption.

However, a hidden but much more pervasive type of rent seeking is the extraction of larger budgets for bureaucracies and what results in the non-pecuniary rewards to workers in government-owned enterprises (Tullock, 1967 and McKenzie and Tullock 1981). The prestige of the senior bureaucrats is enhanced if the size of the workforce is expanded to be larger than necessary to meet production targets. Similarly, offices are more grandiose, holidays are longer, and benefits are greater and so on.

Bogetoft and Hougaard (2003) suggest that the existence of X-inefficiency in production is the outcome of a rational decision-making process that represents on-the-job compensation to managers. Whereas X-inefficiency is viewed by Leibenstein (1966, 1978) as non-maximising behaviour, Stigler (1976) argues that its existence is symptomatic of firms maximising their individual utility functions. Faced with a target level of output, a give set of inputs and factor prices, the bureaucrat minimises costs subject to a Williamson (1963) type of utility function that includes within it arguments on the level of output and a subset of factor inputs. In other words for the *i*th bank, given the ‘k’ factor inputs, the bureaucrat minimises costs to meet a utility function which contains the ‘j’ outputs and a subset ‘n’ of factor inputs, given standard neoclassical technology.

4. Bank Efficiency in the Asian Emerging Economies

The theory of measuring efficiency is straightforward if one knows precisely the inputs and output a firm produces. While this is quite clear in the case of a manufacturing firm it is not straightforward in the case of a bank. The literature distinguishes between two main approaches – the intermediation approach and the production approach. The intermediation approach recognises that the main function of a bank is a financial intermediary that takes in deposits and transforms them into loans and other earning assets. According to this approach the inputs will be deposits plus borrowed funds along with the traditional factors of production (labour and fixed assets) and the outputs will be loans and other earning assets (Sealey and Lindley, 1977). Total costs according to this approach will be what are traditionally recognised as operational expenses plus total interest costs.

In contrast, the production approach recognises that a bank is a producer of a range of financial services. These services are to deposit holders and borrowers alike and include not just intermediation services but a host of other financial services that would be charged to the non-interest earning account. Under this approach the number of deposit and loan accounts plus the number of financial transactions logged over a period of time would be taken as the appropriate definition of output and the inputs will be purely labour and fixed assets (as a measure of capital in neoclassical production theory). Total costs would only cover operational costs and interest costs are excluded (Ferrier et al. 1993). The literature on bank efficiency has tended to produce results using the intermediation approach, largely because balance sheet and income account data is more readily available that what would be required for the production approach.

At one time most studies of bank efficiency were of the developed economies. Indeed, Berger and Humphrey (1997) survey 130 studies that have employed frontier analysis in 21 countries. Of these studies, only 8 were of developing and Asian countries (including 2 in Japan). Studies on US financial institutions were the most common, accounting for 66 out of 116 single country studies. However, recent years have seen a plethora of studies of emerging economies using frontier analysis to evaluate the effects on efficiency of deregulation and reform in the banking market.

The last big area of reform in the Chinese economy remains the banking system. Banking reform has progressed in the Chinese fashion of cautious deregulation, which has attracted numerous scholars both within and without China to study bank efficiency.[8] Using a stochastic frontier analysis Fu and Heffernan (2007) find average cost inefficiency of 40 percent - 60 percent over the period 1985-2002. They also found that government-owned banks in China are less cost efficient than other banks (confirmed also by Yao et al. 2007 and Zhiang et al. 2009). While confirming the estimates obtained by SFA Chen et al. (2005) using DEA found the reverse – on average that state-owned banks were more cost efficient than non-state-owned banks. However, the latter findings are questionable as the conventional DEA is subject to finite sample bias.[9] The Chen et al. (2005) findings go against the consensus which, is that state owned banks are less efficient than other banks. Matthews et al. (2007) use a bootstrap methodology with DEA and confirm the findings of Fu and Heffernan (2007) including the finding that the average cost efficiency of the government owned banks was less than that of the other national banks.

Studies for the Indian banking market confirm the general finding that reform improves bank efficiency but in contrast to the general findings for China, studies using DEA tend to find that government owned banks exhibit a higher level of cost and technical efficiency than other banks (Ray and Das, 2010; Bhattacharya et al.

[8] For a list of studies using parametric and non-parametric methods published in Mandarin see Zhang (2010)

[9] Also using DEA, Laurenceson and Zhao (2008) find a high level of cost efficiency in Chinese banks in the post-WTO period. However once again the estimates obtained by conventional DEA are questionable.

1997; Sathye, 2003; Atullah and Le, 2006). Shanmugam and Das (2004) use SFA to confirm that state-owned banks are more efficient than private banks but also found that foreign banks are more efficient than the average.

The IMF restructuring policy of weak banks for Indonesia, Korea, Philippines and Thailand is tested with DEA-based efficiency measures by Ariff and Can (2009). Their findings suggest that the efficiency of restructured banks is no greater than the pre-IMF intervention period. However, it is not clear that efficiency analysis alone can provide insight into policies that have a long-term gestation. Efficiency in the pre- and post-Asian economic crisis of 1997 is studied by Margono et al. (2010) for Indonesia. The key finding for Indonesia is that cost efficiency improved in the post - crisis period but increased at a lower rate than in the pre-crisis period suggesting that banks adopted a more cautious approach to expanding balance sheets and with it output after the crisis.

Studies of bank efficiency in Pakistan have also been concerned with the effects of reform and deregulation. Studies by Iimi (2002) and Hardy and Patti (2005) use parametric methods to assess the effects of structural deregulation during the 1990s. Using a distribution free approach, Ansari (2006) finds that cost inefficiency varied between 13 percent and 51 percent across individual banks over the period 1991-2002. The effect of changes in corporate governance on efficiency was examined by Ahmed (2006) who confirms the general finding that financial sector reforms improved banking sector performance and that privatized banks performed the best.

A number of scholars have used DEA to estimate bank efficiency in Pakistan with the objective of assessing the effects of reform. Qayyum et al. (2007) found that the efficiency of banks privatized during the reform process improved. Examining an earlier period Burki and Niazi (2010) found a decline in efficiency in the 1993-96 period followed by an improvement thereon to 2000. The most recent study by Akhtar (2010) concluded that the average level of bank efficiency was low and that foreign banks had a higher level of efficiency than domestic banks.[10] The latter finding is also supported by Usman et al. (2010).

The efficiency of Islamic banks across a number of countries that operate Islamic banking systems was examined by Hassan (2006) who found that on average Islamic banks had a lower level of cost and technical efficiency than non-Islamic banks. This is also confirmed by the research of Shahid et al. (2010) for Pakistan but only in the specific case of variable returns to scale.

The findings for bank efficiency in the emerging economies can be separated into firm and tentative compartments. The firm finding is that deregulation and reform tend to improve bank efficiency. The tentative findings are that first, ownership and

[10] This is in contrast to a previous study, where Akhtar (2002) found little difference in the technical and allocative efficiency estimates between foreign banks and domestic banks in 1998. Rizvi (2001) found that foreign banks underperformed domestic banks in terms of technical efficiency for the period 1993-98.

governance tend to produce mixed results depending on the economy examined. In some countries state-owned banks are less efficient than private banks. Second, foreign banks tend to have a higher level of efficiency than domestic banks. Third, specialist banks such as Islamic banks tend to have a lower level of efficiency than conventional banks.

A number of criticisms can be applied to the findings of researchers using the SFA and DEA methodologies to measure bank inefficiency. With regard to SFA estimation of cost functions, the use of panel data provides a fixed ranking of banks which does not change from year to year. The implication is that a bank cannot improve its position relative to other banks no matter what it does – clearly an absurd and unpalatable conclusion. This is not a problem for researchers who use DEA because in principle this method can be applied yearly provided there are sufficient data. However, there are a number of issues with the use of DEA raise doubts about its general application. First, robustness of efficiency estimates using DEA is an important issue in many studies. Different inputs and outputs produce different results as in Chansarn (2008) for Thailand. Efficiency estimates are either relatively too volatile with some sets of inputs and outputs and performance rankings which are not stable. The lack of robustness in DEA estimates highlights the importance of using a wide range of inputs and outputs and bootstrapping the results to evaluate the difference between sets of estimates for statistical significance.

Second, researchers typically use a two-stage estimation method to explain the efficiency estimates obtained from DEA using Tobit-type estimation techniques. However, it is often that these studies are flawed by employing endogenous variables as explanatory terms which raise doubts about the interpretation of the findings. Furthermore, Simar and Wilson (2007) argue that such an approach is flawed since the DEA estimates are biased and serially correlated in a complicated and unknown way. They propose a double-bootstrap method which applies the bootstrap to the DEA estimate while regressing environmental and bank specific variables to the efficiency scores.

In what follows, I present estimates of cost efficiency for Pakistan banks using a number of different models of inputs and outputs. Furthermore these estimates are obtained using the bootstrap technology of Simar and Wilson (2000a, 2000b). Third, HAC estimation is used to estimate the speed of reduction of inefficiency and to identify drivers of the inefficiency reduction.

5. Cost Inefficiency in the Pakistan Banking Sector

In this section, I present the input and output variables used in estimating bank efficiency in Pakistan. There is a reasonable consensus in the literature as to what constitutes inputs in banking efficiency studies but less of a consensus regarding outputs. With inputs, the favoured model is the intermediation method of Sealey and Lindley (1977). The argument is that besides the conventional inputs of capital (fixed assets) and labour, in the case of banks, deposits are an input for the production of loans.[11] In the case of outputs the literature distinguishes between stocks and flows. So outputs under the intermediation method would Loans and the stock of Other Earning Assets, but as Drake (2003) reports, researchers have included non-interest earnings as an output to reflect a growing area of bank activity. Others have concentrated output on the flows of interest and non-interest earnings (Matthews, 2013).[12]

Five models are estimated and the estimated cost efficiency scores are converted into scores of cost inefficiency. These in turn are decomposed into its technical inefficiency and allocative inefficiency components.[13] The models have the same inputs but differ in the outputs. Inputs are Labour, Capital (fixed assets) and Deposits and their respective prices, unit cost of labour, unit cost of fixed assets and unit cost of deposits.

Table 1: Output Vectors for Efficiency Estimation

Model	Variables
1	Loans minus NPLs, other earning assets, non-interest income, NPLs as bad output
2	Loans minus NPLs-, other earning assets, NPLs as bad output
3	Non-interest income, interest income
4	Total loans, other earning assets, non-interest income
5	Total loans, other earning assets

Outputs are distinguished by the treatment of non-performing loans as a bad output and non-interest income as an additional output.[14] Table 1 describes the output data for each model. Data was obtained from Bureau Van Dijk Bankscope supplemented by SBP sources and covered the period 2002-2009. The data is an unbalanced sample of banks, constituting at the maximum 23 private banks, 4 government-owned banks, 5 Islamic banks of which 4 are domestic private banks and 1 is a foreign-owned private bank. In total the sample constitutes 174 bank-years. Table 2 shows the bias-corrected estimates of cost inefficiency and its components technical inefficiency and allocative inefficiency.

It can be seen that except for model 3 where the output consists of interest revenue and non-interest revenue, the estimates are broadly similar. Treating non-performing loans as a bad output do not appear to have had much of an effect on the estimates of overall cost inefficiency. The larger picture is that cost inefficiency is dominated by technical inefficiency (or X-inefficiency as it is sometimes referred to).

[11] See also Drake (2003)

[12] For a recent survey see Fethi and Pasiouras (2010)

[13] With the conventional DEA, allocative efficiency AE = (CE/TE) but with the bootstrap DEA, AE cannot be obtained in this way because the statistical property that $E(x/y) \neq E(x)/E(y)$. We therefore define cost inefficiency CIE = 1 – CE and technical inefficiency TIE = 1- TE, and allocative inefficiency as AIE = CIE – TIE.

[14] The conventional method of dealing with a bad output in DEA is to invert it as an output vector so that maximising the inverted value, minimises the bad output. See Thanasoulis et al. (2008)

Table 2: Bias-corrected estimates of inefficiency

Model	Cost Inefficiency	Technical Inefficiency	Allocative Inefficiency
1	36.1% (21.0%)	22.5% (21.0%)	13.6% (22.7%)
2	34.8% (19.6%)	20.4% (18.5%)	14.4% (21.3%)
3	40.0% (18.5%)	32.7% (18.9%)	7.3% (23.2%)
4	37.9% (19.9%)	26.9% (21.7%)	11.0% (28.5%)
5	34.3% (18.5%)	21.4% (18.7%)	12.9% (24.3%)

Average of all banks. Standard deviation in parenthesis (1000 bootstraps) 2002-2009; 174 bank-years per model

Table 3 provides the breakdown of cost inefficiency by ownership and type of bank for the pooled results of the five models.[15] A test for the difference between the average inefficiency of banks within a particular category type from those that are not, is conducted using the pooled results of the outputs of the five models. Since the distribution of the inefficiency scores may not be standard normal, a non-parametric test is applied (Mann-Whitney). The Rank-Sum test provides the probability that the particular bank category is part of the population. In other words it tests if the particular bank category can be separated from the rest. A significant value suggests that the category can be separated.

The results suggest that there is sufficient independent variation in each bank category type to separate the distribution of technical efficiency from the rest of the population. Also the only bank category that has a distribution of allocative inefficiency that can be separated from the rest is foreign banks. However, these statistics are only indicative as the means and distribution could be signaling other relevant but unidentified factors.

Table 3: Cost inefficiency by bank type

Bank Type	Bank-Years	Cost Inefficiency	Technical Inefficiency	Allocative Inefficiency
All banks	870	36.60%	24.80%	11.80%
Public owned	115	38.4% (-0.66)	21.8% (2.97)**	16.6% (-1.20)
Private	755	37.2% (-2.07)**	25.6% (5.23)***	11.6% (-0.09)
Foreign	35	18.6% (5.13)***	16.4% (4.96)***	2.2% (2.26)**
Islamic	100	29.5% (3.35)***	16.5% (5.93)***	13.0% (-0.70)

* 10% significance, ** 5% significance, *** 1% significance

Average values. ‘z’ values for rank-sum test

[15] Most banks in Pakistan conduct Islamic and conventional business however, this paper categorises Islamic banks as those that exclusively conduct Sharia compliant banking.

One way of identifying these other factors is to estimate the change in cost inefficiency and the change in allocative inefficiency between periods in terms of a set of determining variables. Using the concept of conditional beta-convergence from the growth convergence literature (Barro 1991), a measure of the speed of convergence to a common level of inefficiency is obtained by regressing the change in the level of inefficiency on the lag of inefficiency and environmental and bank specific variables to allow for convergence to different levels of inefficiency. However it is shown by Simar and Wilson (2007) that the estimated inefficiencies may be serially correlated. They propose a double bootstrap procedure to adjust for the bias caused by the inherent correlation among the estimated inefficiencies. The problem of potential bias is further compounded by the existence of the lagged inefficiency score. Developing a valid bootstrap procedure for estimating conditional beta-convergence is computationally intractable. However in an attempt to deal with the potential serial correlation, estimates of the rate of decline of inefficiency are obtained using a panel GLS Heteroskedastic-Autocorrelation Consistent (HAC) estimator.

The dependant variable is the yearly change in the specific type of inefficiency. The coefficient on the lag level determines the speed of reduction of inefficiency to a cluster determined by the steady state values of the driving variables. Table 4 presents the results for cost inefficiency, allocative inefficiency and technical efficiency respectively using the HAC estimator.

The second but last row is the Breusch-Pagan test for heteroskedasticity and the final row is the Woolridge (2002) F test for autocorrelation in panel data (see Drukker, 2003). The last two rows of Table 4 indicate that the use of the HAC estimator was appropriate. Autocorrelation in the panel could not be rejected at the 1 percent level for the regressions of both types of inefficiency, and heteroskedasticity could not be rejected at the 5 percent significance level.

The results of Table 4 have several important implications. First the negative coefficient on the lag of the level of inefficiency indicates a dynamic convergence to a cluster defined by the steady-state values of the driving variables. Here the variable YEAR is a trend term which shows that the secular decline in cost inefficiency is being driven by a decline in allocative inefficiency. Banks that have a large number of branches (BRANCH) exhibit a higher level of inefficiency than those with a small number. Also included in the analysis are zero-one dummy variables to identify the different types of models used.

Columns 3 and 4 of Table 4 show the determinants of allocative inefficiency and the speed of convergence to a cluster. Allowing the trend decline in allocative inefficiency, banks with a larger number of branches have a higher level of allocative inefficiency but a lower level of technical inefficiency than banks that have a smaller number. This suggests that branches create an allocative inefficiency through an excess of branches but those branches are technically more efficient than banks with fewer branches. Other specific factors indicate that foreign banks have a lower level of allocative inefficiency and technical inefficiency; (FOR – zero-one dummy) and Islamic banks have a higher level of allocative inefficiency (ISLAM – zero-one dummy) but this is counterbalanced by lower technical inefficiency so that the overall effect on cost inefficiency is neutral. The competitive state of the banking market is picked up by the Herfindahl Hirschman Index [16] (HHI) which shows that even allowing for the secular decline in allocative inefficiency, the improved competitiveness of Pakistan banking has generated a lower level of inefficiency.

Table 4: HAC Estimates (2002-2009)

Variable	ΔCIE	ΔAIE	ΔTIE
Intercept	66.74 (.000)***	68.15 (.000)***	-1.861 (.599)
CIE(t-1)	-.757 (.000)***	-	-
AIE (t-1)	-	-.824 (.000)***	-
TIE(t-1)	-	-	-.740 (.000)***
YEAR	-.033 (.000)***	-.034 (.000)***	.001 (.571)
BRANCH	.00003 (.004)***	.000078 (.000)***	-.00003 (.000)***
FOR	-.190 (.000)***	-.219 (.000)***	-.077 (.011)**
HHI	.00003 (.163)	.000065 (.002)***	.000009 (.537)
ISLAM	.041 (.192)	.117 (.000)***	-.085 (.000)***
MODEL1	.0527 (.002)***	.0061 (.390)	.0345 (.001)***
MODEL2	.0424 (.008)***	.0167 (.280)	.0148 (.146)
MODEL3	.0436 (.005)***	-.0538 (.000)***	.0811 (.000)***
MODEL4	.0196 (.231)	-.0131 (.039)**	.0282 (.007)***
Chi-Sq(1)	6.43**	4.29**	34.19***
F(1, 119)	71.60***	43.29***	64.06***

‘p’ values in parenthesis * 10% significance, ** 5% significance, *** 1% significance

6. Conclusion

Bank efficiency is clearly a topic worthy of consideration and it is particularly worthy of study in the case of emerging markets. In economies where capital and debt markets are as yet undeveloped, the principal conduit for economy wide investment and saving is through the banking system. The efficiency of the banks is an indicator of the efficiency of financial intermediation. Furthermore, the banking sector of the emerging economies is facing stronger competition due to the globalisation of the

[16] Measured as the sum of the square of market shares of individual banks in terms of assets. The banking market is said to be competitive if HHI is less than 1000, somewhat concentrated if HHI is 1000-1800 and very concentrated if HHI is greater than 1800 (Rhoades 1993). By this measure the Pakistan banking market was somewhat concentrated until 2006 when HHI fell below 1000.

financial system. While the trend in deregulation and global competition will be muted for the next few years as a result of the financial crisis, the pace will pick up once the world economy is stabilised.

Individual banks will be interested to know if they represent benchmarks for others to emulate or laggards that need improving. The general finding is that cost inefficiency is declining over time and converging on clusters defined by a small set of driving variables. Cost inefficiency is directly related to the number of branches a bank operates. With the Western banks, the improvement of cost efficiency has been through the downsizing of the branch network and expansion of online and telephone banking facilities. This may not be an appropriate policy for Pakistan, where social considerations may dictate the existence of unprofitable branches. This then becomes a constraint and not a choice variable, which suggests that any allocative inefficiency is actually rational and not due to poor management.

The conceptual problem of the meaning of efficiency is not an issue. The main issue is what constitutes a bank's input and output and what is the most appropriate measure of cost efficiency. To the State Bank of Pakistan this is a practical matter and they require a practical solution. There is no alternative to using different methods to triangulate to a consensus. However, it is important that the deficiencies of the methods used should be carefully articulated so that the policy maker is aware of the strengths and the weaknesses of the information on which policy is constructed.

References

Ahmad U (2006). “Impact of Financial Sector Reforms on Efficiency and Productivity of Domestic Commercial Banks in Pakistan”, *MPhil Thesis, FUUAST, Islamabad.*

Akhtar M. H. (2002). “X-Efficiency Analysis of Commercial Banks in Pakistan: A Preliminary Investigation”, *The Pakistan Development Review*, 41:4, 567-580

Akhtar, M. H. (2010). “X-Efficiency Analysis of Pakistani Commercial”, *International Management Review*, 6:1, 12-24.

Al-Muharrami S. and Matthews K. (2009). “Market Power versus Efficient-Structure in Arab GCC banking”, *Applied Financial Economics*, 19, 1487-1496

Ansari M. S. (2006). “An Empirical Investigation of Cost Efficiency in the Banking Sector of Pakistan”, *SBP Working Paper*, No. 12, June 2006.

Atullah A. and Le H. (2006). “Economic Reforms and bank Efficiency in Developing Countries: The Case of the Indian banking Industry”, *Applied Financial Economics*, 16, 653-663.

Ariff M, Can L. (2009). “IMF Bank-Restructuring Efficiency Outcomes: Evidence from East Asia”, *Journal of Financial Services Research*, 35:2, 167-187.

Barro R. J. (1991). “Economic Growth in a Cross Section of Countries”, *Quarterly Journal of Economics*, 106, 407-443.

Berger A. N., Hunter W. C. and Timme S. G. (1993). “The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present and Future”, *Journal of Banking and Finance*, 17, 221-249.

Berger, A.N., and Humphrey, D.B. (1997). “Efficiency of Financial Institutions: International Survey and Directions for Future Research”, *European Journal of Operational Research*, 98, 175-212.

Bogetoft P. and Hougaard (2003), “Rational Efficiencies”, *Journal of Productivity Analysis*, 20, 243 – 291.

Bhattacharyya A, Lovell C. and Sahay P. (1997). “The Impact of liberalization on the Productive Efficiency of Indian Commercial Banks”, *European Journal of Operational Research*, 98, 332-345.

Buchanan J. (1980). “Rent Seeking and Profit Seeking”, in Toward a Theory of the Rent Seeking Society, (eds) J Buchanan, R Tollison, and G Tullock, *College Station: Texas A&M University Press.*

Burki Abid A. and Niazi, G. S. K. (2010). “Impact of Financial Reforms on Efficiency of State-Owned, Private and Foreign Banks in Pakistan”, *Applied Economics*, 42:24, 3147-3164.

Cabral L. M. B. (2000). “Introduction to Industrial Organization”, *Cambridge Mass: MIT Press.*

Casu B. and Molyneux (2005). “A Comparative Study of Efficiency in European Banking", *School of Accounting, Banking & Finance, University of Wales Bangor, Discussion Paper.*

Chen X, Skully M. and Brown K. (2005). “Banking Efficiency in China: Applications of DEA to Pre-and Post-Deregulation Eras: 1993-2000”, *China Economic Review*, 16, 2229-245.

Daley J, Matthews K, and Whitfield K. (2008). “Too-big-to-Fail: bank Failure and Banking Policy in Jamaica”, *International Financial markets, Institutions and Money*, 18, 290-303.

Dong F. and Featherstone A. (2006). “Technical and Scale Efficiencies for Chinese Rural Credit Cooperatives: A Bootstrapping Approach in Data Envelope Analysis”, *Journal of Chinese Economic and Business Studies*, 4:1, 57-75.

Drake L. (2003). “Costs and Efficiency in Banking: A Survey of the Evidence from the US, the UK and Japan”, in Mullineaux A. W. and Murinde V. (Eds.) *Handbook of International Banking, Edward Elgar: Cheltenham.*

Drukker D. M. (2003). “Testing for Serial Correlation in Linear Panel-Data Models”, *Stata Journal*, 3, 168-177.

Farrell M. J. (1957). “The Measurement of Productive Efficiency”, *Journal of the Royal Statistical Society*, Series A, 120:3, 253-281.

Ferrier G. D, Grosskopf S., Haynes K. J. and Yaisawarng G. (1993). “Economies of Diversification in the Banking Industry”, *Journal of Monetary Economics*, 31, 229-245.

Fethi M. D. and Pasiouras F. (2010). “Assessing Bank Efficiency and Performance with Operational Research and Artificial Intelligence Techniques: A Survey”, *European Journal of Operational Research*, 204, 189-198.

Fu X. and Heffernen S. (2007). “Cost X-Efficiency in China’s Banking Sector”, *China Economic Review*, 18, 35-53.

Hassan M. K. (2006). “The X-Efficiency in Islamic Banks”, *Islamic Economic Studies*, 13, 2, 49-78.

Iimi, A. (2002). “Efficiency in the Pakistani Banking Industry: Empirical Evidence after the Structural Reform in the Late 1990s” *JBICI Working Papers No 8. Tokyo: JBIC Institute*.

Jiang C, Yao S, Zhang Z. (2009). “The Effects of Governance Changes on Bank Efficiency in China: A Stochastic Distance Function Approach”, *China Economic Review*, 20, 717-731.

Koopmans T. C. (1951). "Efficient Allocation of Resources", *Econometrica*, XIX, 10, 455-465

Kumbhakar S. C. and Knox-Lovell C. A. (2000). *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press

Lange O. (1942). "The Foundations of Welfare Economics", *Econometrica*, 10, 215-228.

Leibenstein H. (1966). "Allocative Efficiency vs. X-Efficiency", *American Economic Review*, 56:3, 392-415.

Leibenstein H. (1975). "Aspects of the X-Efficiency Theory of the Firm", *The Bell Journal of Economics*, 6:2, 580-606.

Margono H, Sharma S. C, Melvin P. D. (2010). "Cost Efficiency, Economies of Scale, Technological Progress and Productivity in Indonesian Banks", *Journal of Asian Economics*, 21:1, 53-65.

Matthews K. (2013). "Risk Management and Managerial Efficiency in Chinese Banks: A Network DEA Framework', Omega, 41, 207-215.

Matthews K, Guo J. and Zhang N. (2007). "Rational Inefficiency and Non-performing Loans in Chinese Banking: A Non-parametric Bootstrapping Approach', *China Finance Review*, 1, 3, 1-21.

McKenzie R. B. and Tullock G. (1975). "The New World of Economics", Homewood, Illinois: Richard D Irwin Inc.

Qayyum A. (2007). 'Financial Sector Reforms and the Efficiency of Banking in Pakistan', South Asian Network of Economic Research Institutes (SANEI).

Ray S. and Das A. (2010). "Distribution of Cost and Profit Efficiency: Evidence from Indian Banking', *European Journal of Operational Research*, 210, 290-317.

Rhoades S. (1993), "The Herfindahl Hirschman Index", *Federal Reserve Bulletin*, 79, 188-189.

Rizvi S F. A. (2001). "Post-liberalisation Efficiency and Productivity of the Banking Sector in Pakistan", *The Pakistan Development Review*, 40:4, 605-632.

Saythe M. (2003). "Efficiency of Banks in a Developing Economy: The Case of India", *European Journal of Operational Research*, 148:3, 662-671.

Sealey C. W. and Lindley J. T. (1977). "Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions", *Journal of Finance*, 32:4, 1251-1266.

Shahid H, Rehman R, Niazi G. S. K, Raoof A. (2010), "Efficiencies Comparison of Islamic and Conventional Banks of Pakistan", *International Research Journal of Finance and Economics*, 49, 25-44.

Shephard R. W. (1953). "Cost and Production Functions", Princeton: *Princeton University Press.*

Simar L. and Wilson P. W. (1998). “Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Non-Parametric Frontier Models”, *Management Science*, 44, 49-61.

Simar L. and Wilson P. W. (2000a). “Statistical Inference in Non-Parametric Frontier Models: The State of the Art”, *Journal of Productivity Analysis*, 13, 49-78.

Simar L. and Wilson P. W. (2000b). “A General Methodology for Bootstrapping in Non-Parametric Frontier Models”, *Journal of Applied Statistics*, 27, 779-802

Simar L. and Wilson P. W. (2007), “Estimation and Inference in Two-Stage Semi-Parametric Models of Production Processes”, *Journal of Econometrics*, 136, 31-64

Simar L. and Wilson P. W. (2008). “Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives”, in H O Fried, C A Knox-Lovell, S Schmidt, The Measurement of Productive Efficiency and Productivity Growth, Oxford: *Oxford University Press*

Stigler G. (1976). “The Xistence of X-efficiency”, *American Economic Review*, 66, 213-216.

Thanassoulis E, Portela M, and Despic O. (2008). “Data Envelope Analysis: The Mathematical Programming Approach to Efficiency Analysis”, in The Measurement of Productive Efficiency and Productivity Growth (Eds) H Fried, C A Knox-Lovell and S Schmidt, Oxford: *Oxford University Press.*

Tullock G. (1967). “The Welfare Costs of Tariffs, Monopolies, and Theft”, *Western Economic Journal*, 5, 224-232.

Tullock G. (1980). “Efficient Rent Seeking”, in Toward a Theory of the Rent Seeking Society, (eds) J Buchanan, R Tollison, and G Tullock, College Station: *Texas A&M University Press.*

Usman M, Wang Z, Mahmood F, Shahid H. (2010). “Scale Efficiencies in Banking Sector of Pakistan”, *International Journal of Business and Management*, 5:3, 104-116.

Wooldridge J. (2002). “Econometric Analysis of Cross-Section and Panel Data”, Cambridge MA: *MIT Press.*

Yao S, Jiang C, Feng G, Willenbockel D. (2007). “WTO Challenges and Efficiency of Chinese Banks”, *Applied Economics*, 39, 629-643.

Williamson O. E. (1963). “Managerial Discretion and Business Behaviour”, *American Economic Review*, 53:5, 1032-1057.

Zhang Xu (2010). “Three Empirical Essays on Efficiency and Productivity in Chinese Banking Industry”, *Unpublished PhD Thesis, Cardiff University*, October.