MEASURING BANK PERFORMANCE IN THE CURRENT EVOLVING FINANCIAL MARKETPLACE

Dr. Nakhun Thoraneenitiyan*

Abstract

This paper presents a concept of ‘relative efficiency’ as an alternative measure to assess bank performance. Traditional financial statement analysis cannot explain all variations in stock prices. Particularly, in recent years, banks have been forced not only to generate profit, but also to perform efficiently to survive in the industry. Normally, bank financial statements may not clarify non-financial activities such as hidden assets or competitiveness of the bank. This paper illustrates how parametric and non-parametric frontier approaches can measure economic value added in terms of a ‘bank’s managerial efficiency’. The bank efficiency estimate provides incremental information not contained directly in bank financial statements for to the bank’s stakeholders.

*Dr. Nakhun Thoraneenitiyan holds a Ph.D. in Banking and Finance from The University of Queensland, Australia. Currently, he is a lecturer in the Finance and Banking Department, Faculty of Business Administration, Sripatum University, Thailand.
TRADITIONAL MEASURES OF BANK PERFORMANCE

In emerging economies, commercial banks usually dominate the financial system. They mobilise savings from depositors to businesses. In Asian countries, especially members of the Association of Southeast Asian Nations (ASEAN), total assets of commercial banks account for more than 70% of total assets of the financial systems. Furthermore, numerous studies argue that the performance of financial intermediation affects economic growth while others indicate that bank insolvencies can result in systemic crises which have adverse consequences for the economy as a whole.

The performance of banks has been an issue of major interest for various stakeholders such as regulators, customers, investors, and the general public, especially, during the current crisis period. The recent global financial crisis (GFC), which was primarily caused by a credit crisis in the U.S. banking system and extended all over the world, has been attracting the interest of researchers and regulators as to why the financial reports during the period prior to the crisis show very low levels of loan losses even as loan growth is increasing dramatically. This can imply that information from the financial reports alone is insufficient to evaluate the performance of a bank.

Generally, a bank’s performance is usually measured by fundamental analysis, which primarily relies on examining its financial statements. The principal aim of fundamental analysis is to improve the ability to forecast future movements in stock performance, which can then be used to design investment strategies (Avkiran & Morita, 2008). In some cases, however, stock market valuation studies have discovered that variations in stock prices does not reflect the variation in earnings analysed by the fundamental statements (Kothari, 2001), or the explanatory power of earnings levels and changes for market returns has significantly decreased over time (Francis & Schipper, 1999).

Whilst some studies examine whether earnings reflect some of the financial information in stock prices, recent research, however, has shifted towards the use of additional data such as economic value added and efficiency to understand how they affect stock prices and returns. During the last five years, researchers have investigated the relationship between bank efficiency and stock returns both in Western countries and Asia (e.g., Beccalli, Casu, & Girardone, 2006; Guzman & Reverte, 2008; Kirkwood & Nahm, 2006; Pasiouras, Liadaki, & Zopounidis, 2008; Sufian & Majid, 2007; Thoraneenitiyan, 2009). In most cases, the results of these studies indicate a positive relationship between stock returns and efficiency changes. Furthermore, the explanatory power of the model with efficiency scores is higher than that of a model that uses the return on equity (ROE) as a measure of performance (Beccalli et al., 2006).

Another problem when a bank’s performance is assessed by using the financial statements is that the firms are essentially isolated from their industry group and the market as a whole. Any follow up comparison of a company’s ratios against similar firms or industry averages evidently fails to capture the benefits of a simultaneous multi-dimensional benchmarking relative to its peers (Avkiran & Morita, 2008). Thus, given that no bank operates in isolation, ratio analysis is an impro-
vised evaluation of a bank’s performance. DeYong (1997) also suggests that comparing the financial ratios of different banks is not appropriate unless the banks are nearly identical in terms of product mix, bank size, market conditions, and other characteristics that can affect the costs of the banks. Thoraneenitiyan and Avkiran (2009) explore impacts of macro-economic environments and restructuring measures on Asian bank efficiency during the 1997 Asian Financial Crisis and found that different macro-economic conditions result in the differences in recovery level of the banking industries.

Recently, banks have faced greater levels of competition and this has created excess capacity in traditional lines of business and forced them to become more market-oriented. The degree of bank complexity has increased further as they have moved away from being traditional intermediaries to more market-oriented institutions, providing a wider range of non-banking products and services. As a result of such changes, banks nowadays rely more heavily on competitive advantages or hidden assets in their superior operations. During 2003 and 2007, very low interest rates, and rising asset values pushed the U.S. banks to expand their mortgage loans, which were known as risky assets, and that resulted in the collapse of Bear Stearns and Lehman Brothers in 2008, eventually. In other words, statistically based ‘efficient frontier’ approaches would be an alternative for measuring relative performance of a bank, especially during the current market conditions.

THEORY OF BANK EFFICIENCY

Bank efficiency studies can be separated into those that examine scale and scope efficiency and those that examine X-efficiency or frontier efficiency. The scale and scope studies estimate an average practice cost function, which relates bank costs to output levels and input prices (Berger, Hanweck, & Humphrey, 1987; Berger & Humphrey, 1992b; Evanoff & Israilevich, 1991; McAllister & McManus, 1993; Mester, 1993, 1996). These studies implicitly assume that there is no X-inefficiency or differences in managerial ability to control costs for any given scale or scope of production, and that the banks are using the same production function technology.

However, the conventional studies on scale and scope economies are beset by a number of problems. For example, the commonly used translog cost function specification gives a poor approximation when applied to banks of all sizes (McAllister & McManus, 1993). Another potential difficulty in the scale efficiency literature is that most studies do not use a frontier estimation method. Scale and scope efficiency, theoretically, apply only to the efficient frontier, and the use of data from banks not on the frontier could confound scale efficiencies with differences in X-efficiency. In addition, Berger, Hunter, and Timme (1993) assert that the most important origin of cost problems in the banking industry is X-inefficiency.

MEASURING RELATIVE PERFORMANCE USING THE EFFICIENT FRONTIER

In the 1990s, the research focus shifted to X-efficiency or frontier efficiency, which estimates deviations in performance from
that of best practice firms on the efficient frontier, holding constant a number of exogenous market factors such as the prices faced in local markets (Allen and Rai, 1996; Berger & Mester, 1997; English, Grosskopf, Hayes, & Yaisawarng, 1993; Mester, 1996). In the bank efficiency literature, the concept of cost efficiency is the most commonly specified criterion (Weill, 2004). The cost efficiency of a bank is determined by two conditions, technical efficiency and allocative efficiency. Technical efficiency (TE) refers to the ability to maximise output levels at given levels of input (output orientation), or the ability to minimise input levels at given levels of output (input orientation). Allocative efficiency (AE) is the ability to select the optimal mix of inputs in light of prices in order to produce given levels of output. The lack of either technical or allocative efficiency leads to a deviation from cost minimisation and creates inefficiencies (Mester, 1997). However, since true cost functions are not known theoretically, inefficiencies must be measured relative to an efficient cost frontier that is estimated from data.

Figure 1 adapted from Coelli et al. (1998, pp.134-135) and Avkiran (2006, pp.87-88) illustrates Farrell’s (1957) ideas using a simple example involving banks, which use two inputs \( (x_1, x_2) \) to produce a single output \( (y) \), under the assumption of constant returns to scale. The curve \( TT' \) represents the different combinations of the inputs that can produce a given level of output for a technically efficient bank. If a given bank uses quantities of inputs, defined by the point \( P \), to produce a unit of output, the technical efficiency \( (TE) \) is measured by the ratio \( 0Q/0P \) while inefficiency of that bank could be represented by the distance \( PQ \), which is the amount by which all inputs could be proportionally reduced in relation to the curve \( TT' \) without reduction in output.

Assuming the input costs ratio are known, \( CC' \) represents the different combinations of inputs \( x_1 \) and \( x_2 \) that can be purchased with the given budget. The cost efficiency \( (CE) \) of a bank at point \( P \) is then defined as the ratio \( 0R/0P \), with \( RP \) representing cost inefficiency. However, the problem with point \( R \) is, while it is cost effi-
cient, it is not technically efficient. Point $F$ indicates the technically cost efficient point that the bank can reach if it can substitute or reallocate inputs. This point represents the mix of inputs $x_1$ and $x_2$ where production costs are composed of savings realised from the efficient conversion of inputs into outputs, and the effective mix of inputs regarding their cost and the organisation’s budget; this point is so-called the point of allocative efficiency. The allocative efficiency ($AE$) is the ratio that captures how far away a technically efficient bank is from the cost efficient frontier. Thus, the measure of the allocative efficiency for bank P becomes the ratio of $0R$ to $0Q$.

Three main parametric frontier approaches have been widely used in the literature: the Stochastic Frontier Analysis (SFA), the Distribution Free Analysis (DFA), and the Thick Frontier Analysis (TFA). The SFA and the DFA assume a functional form for the production or cost frontier, but separate the inefficiencies from random errors in a different way. On the other hand, the Thick Frontier Analysis (TFA) specifies a functional form and assumes that deviations from predicted performance values within the highest and lowest performance quartiles of observations represent random error, while deviations in predicted performance between the highest and lowest quartiles represent inefficiencies (Berger & Humphrey, 1991; Shaffer, 1993).

Non-parametric frontier approaches, such as Data Envelopment Analysis (DEA) by Charnes, Cooper and Rhodes (1978), and Free Disposal Hull (FDH) by Desprins, Simar, and Tulken (1984), which is a special case of the DEA model, impose no structure on the specification of the best-practice frontier. DEA is a linear programming technique where the set of best-practice or frontier observations are those for which no other decision making unit (DMU) or linear combination of units has as much or more of every output or as little or less of every input. The DEA frontier is formed as the piecewise linear combinations that connect the set of these best-practice observations, yielding a convex production possibilities set. Therefore, DEA does not require the explicit specification of the underlying production relationship. There are two widely used DEA models. The first, developed by Charnes et al. (1978) assumes constant returns to scale (CRS). The second, developed by Banker, Charnes, and Cooper (1984) assumes variable returns to scale (VRS). These models are respectively known as the CCR and the BCC models.

However, the traditional DEA models (the CCR and the BCC models) that measure technical efficiency in a scalar measure ($\theta^*$) accounts for the proportionate change (radial) in input/output values, but neglects the existence of non-radial inefficiencies\(^1\), a bank may have the full efficiency score of 1 following the BCC model, although it has input excesses. Figure 2 illustrates the example of radial and non-radial measures of efficiency using two inputs, $x_1$ and $x_2$, and one output $y$.

Consider the six banks in Figure 2 (adapted from Zhu, 2003). Following a ra-

\(^1\)Non-radial inefficiency is commonly known as ‘slack’. Slacks refer to excesses in inputs (input slacks), or shortfalls in outputs (output slacks) (Tone, 2001).
dial DEA model (e.g., the BCC model), banks A, B, C, D and E are efficient, and bank F is inefficient. However, when non-radial inefficiencies or slacks are accounted for, banks A and E are recognised as inefficient since they have non-zero slack on the input $x_1$, and the input $x_2$, respectively (i.e., bank A can reduce the use of input $x_1$ to point B, while bank E can reduce the use of input $x_2$ to point D). In addition, to obtain the efficiency score for bank F, the radial DEA model selects a convex combination of banks C and D as the efficient target, whereas a non-radial DEA model selects bank B as the efficient target.

Although Charnes, Cooper, Golany, Seiford, and Stulz (1985) developed the additive model of DEA, which deals directly with input excesses and output shortfalls, the model has no scalar measure like $\theta^*$ in the CCR model, and makes it difficult to interpret the results. To overcome this issue, Tone (2001) proposes a slacks-based measure (SBM), which identifies non-radial inefficiency in the scalar measure known as $p$.

In summary, both parametric and non-parametric frontier approaches have advantages as well as disadvantages. Although the parametric approaches are more common and have the advantage of separating noise from inefficiencies, they have a major drawback in requiring an explicit functional form for the technology and specific distributional assumptions for the error term. Since the true production technologies are essentially unknown, the problem of model specification might cause confusion in isolating the inefficiency. On the other hand, non-parametric approaches have a major disadvantage in that they do not capture random errors, in which case the inefficiency estimated might be overstated. Recent efficiency studies have integrated advantages of the two approaches in various ways, especially, to deal with the impact of environmental fac-

---

**Figure 2: Radial and Non-Radial Efficiency (Input Orientation)**

![Figure 2: Radial and Non-Radial Efficiency (Input Orientation)](image-url)
tors those may confound the efficiency estimated by the traditional models.

BANK BEHAVIOURS IN EFFICIENCY MODELS

Modelling the production and cost functions of banks, which have multiple services and products, raises a long-standing debate on the definition of the inputs and outputs of banks. There are two main schools of thought on bank behaviour, which have been widely used in the banking literature. The first is the intermediation approach, which views a bank as a mediator of funds between depositors and investors (Sealey & Lindley, 1977). Following this concept, deposits, labour and physical capital are regarded as inputs being converted into loans. The second is the production approach, which emphasises the role of banks as providers of services for account holders. With this view, banks are regarded as using inputs such as labour and capital to generate deposits and loans. While the production approach is probably better able to evaluate the efficiency of bank branches, the intermediation approach may be better for the evaluation of banks in their entirety (Cavallo & Rossi, 2002).

The intermediation approach is preferred to in examining bank efficiency during the transition period. One explanation is that a key aim of bank policies during the transition period is to improve the banking system’s capacity to provide financial intermediation between savers and borrowers (Dziobek & Pazarbasioglu, 1997), and the intermediation approach normally includes interest expense, which is a large proportion of any bank’s total costs (Berger & Humphrey, 1991; Elyasiani & Mehdian, 1990). This is supported by many studies (e.g., Gilbert & Wilson, 1998; Isik & Hassan, 2002; Kraft & Tirtiroglu, 1998), which examine bank efficiency during a period of regulatory changes in transition countries.

There are also other approaches to modelling bank behaviour. For example, under the value-added approach, high value creating activities such as making loans and taking deposits are classified as outputs, whereas labour, physical capital, and purchased funds are regarded as inputs (Necmi K Avkiran, in press; Wheelock & Wilson, 1995a). Lastly, the user-cost approach regards an asset as an output if the financial returns are higher than the cost of funds. This is similar to a liability item which is classified as an output if the financial costs are less than the opportunity cost (Berger & Humphrey, 1992a). The review of bank behaviours in efficiency models are well documented in Fethi and Pasiouras (2010).

MEASURING MANAGERIAL PERFORMANCE AND IMPACT OF ENVIRONMENTAL FACTORS

The Banking industry is very sensitive to macro-economic conditions (Berger & Humphrey, 1992a; Thoraneenitiyan & Avkiran, 2009; Wheelock & Wilson, 1995b). An economic shock such as a financial or banking crisis can affect the efficiency of a bank (Berger, Bonime, Covitz, & Hancock, 2000). A local economic downturn and corporations with problem loans can be potential reasons for cost inef-
ficient banks (Berger & DeYoung, 1997). Since loans are one of the major outputs of a bank, a rise in problem loans during a crisis period might lead to bank X-inefficiency. High levels of problem loans cause bank costs to rise (e.g., monitoring costs, negotiating costs, and selling off of those problem loans). Thus, those nonperforming loans tend to decrease the cost efficiency of banks.

On the other hand, greater macro-economic stability and competition in banking from foreign entry, as well as the development of supportive institutions, promote cost efficiency (Fries & Taci, 2005). Progress in banking reform has a non-linear association with cost efficiency. In the initial stages of banking reform, cost efficiency increases significantly, but it then declines as reforms advance. Banking systems with higher ratios of capital to total assets and banks with lower loan losses also tend to have lower costs. This may be associated with lower risks in banking sectors. It is possible for some of the banks in the same sample to operate in substantially different environments or be influenced differently when exposed to the same external factors. In a real life analysis, failure to account for environmental factors is bound to confound managerial efficiency and leads to unreliable economic decisions.

There have been recent advances in respect to how researchers incorporate the potential impact of environmental, economic and regulatory factors on bank efficiencies (e.g., Dietsch & Lozano-Vivas, 2000; Drake, Hall, & Simper, 2006; Fries & Taci, 2005; Lozano-Vivas, Pastor, & Pastor, 2002; Thoraneenitiyan & Avkiran, 2009). In the set of parametric studies, the external variables (which are added as control variables to the functional form equation) are assumed to have a direct effect on the production (cost) structure. Hence, each bank is assumed to face a different production (cost) frontier. In the non-parametric studies, the external factors are typically introduced as non-discretionary inputs and/or outputs, having a direct effect on the efficient production frontier.

There are two main approaches to incorporating uncontrollable or environmental factors in a non-parametric approach such as DEA: the single-stage approach, and the multiple-stage approach. In the single-stage approach, the environmental variable can be included directly in DEA as it becomes a constraint in a linear programming model (see Banker & Morey, 1986a; Banker & Morey, 1986b). However, this approach has a problem in that when an environmental variable is included, the comparison set can be reduced, resulting in more firms emerging as efficient and thus, reducing the discriminatory power of the analysis. Also, only one environmental variable at a time can be considered by this method, and the direction of the influence of this variable upon efficiency needs to be known a priori (Coelli et al., 1998).

The multiple-stage approach can be designed in a number of ways. The common two-stage method is to run DEA with traditional inputs and outputs, and then in the second stage, regress the DEA scores obtained from the first stage on the set of environmental variables. Use of second-stage regressions is superior to the single-stage in many ways. First, regressing the efficiency scores obtained in the first stage on the set of environment variables can solve the problem of to what extent environmental factors
affect bank efficiency. This method enables the researcher to consider many environmental variables simultaneously without decreasing discriminating power of fully efficient units (Pastor, 2002). Second, there is no need to make prior assumptions on the orientation of the influence of each environmental variable. Third, this approach allows researchers to conduct hypothesis tests to see if the variables have a significant effect on efficiencies (Coelli et al., 1998).

Fried, Schmidt, and Yaisawarng (1999) introduced a multiple-stage analysis to measure the efficiency of nursing homes. The analysis starts with DEA by using the BCC model, and then uses Tobit regression in the second stage to explain the impact of the operating environment on unit performance. In the third stage, the data used in the first stage are adjusted before a repeat of DEA. Recently, Drake et al. (2006) followed Fried et al.’s (1999) procedure by using the BCC model and the SBM model approaches in the first stage to capture radial and non-radial input inefficiencies, then separately regressed them on a set of exogenous factors by using Tobit regression in the second stage. In stage 3, adjusted inputs were used to re-estimate DEA scores by both DEA models. However, both Fried et al. (1999) and Drake et al. (2006) methods have a drawback in that they are unable to account for statistical noise or measurement error in the data.

The shortcoming of using Tobit regression in the second stage, which cannot separate statistical noise from inefficiency, is addressed by Fried, Lovell, Schmidt, and Yaisawarng (2002) where radial input inefficiency and statistical noise are accounted for in a three-stage analysis. They apply an input-oriented BCC model to obtain efficiency scores and radial input inefficiencies in the first stage. Then in stage 2, they use SFA to regress the radial inefficiencies on environmental variables, and a composite error term that comprises statistical noise and managerial inefficiency. Finally, the DEA of stage 1 is repeated by replacing observed input data with the data adjusted for the influence of environmental effects and statistical noise. Thus, the scores from the final DEA analysis reflect managerial efficiency only.

However, Avkiran and Rowlands (2008) argue that the use of the BCC model, which is partially units-invariant (Lovell & Pastor, 1995), cannot capture non-radial inefficiency of banks. They note that to be consistent in interpretation of DEA and SFA estimates, a fully units-invariant DEA model is needed. Moreover, since the BCC model has to be oriented, the procedure follows Fried et al. (2002) and can focus only on input or output slacks depending on the model orientation. The authors, therefore, present the integrated SBM model, which can capture non-radial inefficiency, and SFA, which can separate impact of external factor and statistical noise from managerial efficiency.

Recent studies illustrate supportive evidence that the integrated DEA/SFA model can measure ‘pure’ managerial efficiency of a bank without impact of environmental factor and statistical noise in the data. Thoranneenitiyan and Avkiran (2009) apply the three-stage approach by Avkiran and Rowlands (2008) that can account for environmental effects and statistical noise to measure relationships between bank restructuring and bank technical efficiency in
Asian developing countries during the period of 1997-2001. The result shows that after adjusting for variation in the operating environment and for the influence of statistical noise, mean efficiency scores change dramatically and exhibit less dispersion. This supports the proposition that some banks that received relatively low (high) initial performance evaluations did so, partly, due to their relatively unfavourable (favourable) operating environments caused by restructuring programs and other country-specific external factors, as well as statistical noise. Avkiran and Thoraneenitiyan (2010) also use the same integrated approach to examine the impact of exogenous factors on bank performance in the UAE banking system. The overall impact of the exogenous factors on average efficiency scores is small. However, after removing the impact of exogenous factors, a discriminatory power for the efficiency model is stronger.

CONCLUSION

This article illustrates frontier approaches as alternatives to measure a bank’s performance. The author argues the drawbacks of the traditional measures of bank performance using financial statements. These include inability to explain variation in stock prices, incapability to assess hidden assets such as competitiveness, and the failure to execute simultaneous multidimensional benchmarking. This paper, then, shows an alternative measure of bank performance in terms of ‘relative efficiency’ to capture competitiveness of bank operations. A bank is to be rated as fully efficient when it performs as the best practice firm on the efficient frontier, holding constant a number of exogenous market factors. Bank efficiency can be estimated using at least two broad approaches: parametric and non-parametric frontier approaches. Recent seminal works have integrated advantages of the parametric and the non-parametric approaches to capture and separate the impact of environment from managerial efficiency. This provides unbiased information for policy makers, regulators, and investors to assess the managerial efficiency compared to the best practice in the industry, without good or bad luck from external factors. Finally, this paper does not oppose the use of financial statement analysis to assess bank performance, but suggests that relative efficiency estimates provide incremental information not contained directly in bank financial statements to the market.

References


at the Australasian Finance and Banking Conference, Sydney.
Charnes, A., Cooper, W. W., Golany, B.,


Thoraneenitiyan, N., & Avkiran, N. K. (2009). Measuring the impact of re-