Issues and Approaches in Efficiency Analysis of Agricultural Banks

Paul N. Ellinger and David L. Neff

Proceedings of a Seminar sponsored by
North Central Regional Project NC-207
“Regulatory, Efficiency and Management Issues Affecting Rural Financial Markets”
Minneapolis/St.Paul, MN
September 26-29, 1992

Food and Resource Economics Department
Institute of Food and Agricultural Sciences
University of Florida

September 1993

Copyright 1992 by author. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Issues and Approaches in Efficiency Analysis of Agricultural Banks

Paul N. Ellinger and David L. Neff

Introduction

In a period of increasing competitive pressure in the financial industry, efficiency is a major concern for bank regulators, policy makers, bank managers and consumers. The savings and loan fiasco, current recessionary environment, credit quality conditions and bank insurance fund problems have each created considerable uncertainty about the profitability and long-term viability of commercial banks. Furthermore, the increased competitive pressure among agricultural lenders has resulted in smaller profit margins. The Farm Credit System has reorganized their credit delivery system while nontraditional lenders, such as farm input suppliers, also have increased their emphasis on providing credit to rural customers. In addition, the deregulation of interstate banking laws has permitted bank expansion into profitable areas and further increased the competitive forces in the industry. The ability and the role of agricultural banks to operate efficiently and profitably in this environment of structural change and increased competition needs to be evaluated.

The ownership structure of banking in the United States has undergone major changes in the last three decades. Approximately 6,500 U.S. banks have merged, consolidated, or been acquired since World War II. The total dollar value of the merger and acquisition activities in the 1980s was $77 billion and in the 1990s it already exceeds $30 billion. The consequences of the adjustments within the banking industry are a major national and political concern. The banking industry comprises almost 20 percent of the U.S. Finance, Insurance and Service sector of the national income accounts. Furthermore, banking's role as the nation's primary financial intermediary and conduit for monetary policy requires the government to maintain a close supervisory role over its activities. However, over the past decade, the government has relaxed many banking regulations. Consequently, degrees of competition in the financial markets have increased and the measurement and evaluation of commercial banks' risk and efficiency levels have received increased attention by policy makers.

Since large issues of policy could be guided by inferences based on the empirical results of studies of bank efficiency, the research design of these studies is critical. The limitations, issues and implications of the research methods should be fully understood. This analysis has two objectives, the first is to discuss the major issues associated with efficiency measurement of agricultural banks. The second objective is to empirically evaluate the efficiency of a sample of agricultural banks by comparing different approaches to three of the major issues discussed. A common body of data is used across different methodological

*Paul N. Ellinger is an assistant professor in the Department of Agricultural Economics, Texas A&M University. David L. Neff is an assistant professor in the Department of Agricultural Economics and Rural Sociology, University of Arkansas.
approaches to illustrate that results, and thus, inferences can differ substantially based on the
definition of the problem and empirical methodology used to analyze the problem.

**Major Issues in Examining Bank Efficiency**

Seven major issues associated with bank efficiency estimation are discussed. These issues
include bank data sources, bank cost definition, bank output definition, the empirical
technique, selection of functional form, the bank entity to examine and the time period
analyzed.

**Commercial Bank Data Sources**

Two principal sources of data are used to estimate efficiencies and economies at commercial
banks. The first is the Board of Governors of the Federal Reserve System Consolidate
Report of Condition and Income (Call Reports) and the second is the Federal Reserve
Functional Cost Analysis Program.

Operating cost data from the Call Reports are aggregated at the bank level into only three
categories: 1) salaries and benefits, 2) expenses of premises and fixed assets and 3) other
noninterest expenses. Complete information about the dollar levels of bank deposits and
loans are provided by quarter for all insured banks. However, no information about the
number of loans or deposits is available. The Functional Cost Analysis Program is a
voluntary program administered by the Federal Reserve. The data contain detailed cost
estimates of distinct bank functions. However, the banks participating in the program may
not be a representative sample of all commercial banks (Clark; Kolari and Zardkoohi) and
time series information on individual banks is not available.

A problem with previous studies that utilize the Call Report data involves the application of
annual accounting data to empirical measures of efficiency. Typically, estimates of
efficiency are based on input and output measures from annual reports. Costs are estimated
using flow measures from the entire estimation period. Output measures, however, are often
based upon stock values measured at a specific point in time. In addition, input prices are
typically measured using a flow measure in the numerator and stock values in the
denominator. Many small, rural banks have a substantial degree of seasonality associated
with their loan and deposit portfolios (stock values) and therefore, the use of annual data may
not provide an adequate representation of input prices and output measures.

**Bank Cost Definition**

The major issue in bank cost definition concerns the inclusion of interest expense as a cost
item. The inclusion of interest expense is an important issue since banks differ in their
sources of funds. Banks which rely relatively more on transaction deposits will typically
have higher operating costs and lower interest costs than banks which rely relatively more on
purchased funds. Hence, the funding mix is a significant component in the tradeoff between
operating and interest costs. Since small banks tend to use transaction and small time-
deposits while large banks tend to use a higher proportion of purchased funds, measures of
bank efficiency could be biased in models which do not include interest expense in the
definition of total bank costs.

Another issue in bank cost definition involves the inclusion of materials and other expenses.
Many studies (i.e. Aly, et al.) include only labor and capital expenses in the operating cost
definition. However, this definition ignores other operating costs of the bank. These items
can be a significant component of non-interest operating costs. For example, in 1990 the
average proportion of other operating costs to total non-interest operating expenses for all
commercial banks was 37.5%.

Bank Output Definition

Considerable controversy exists regarding the definition of bank output. Until recently,
most bank efficiency analyses followed either the intermediation or production approaches to
bank output definition. Currently, three definitions of bank output are typically used. These
alternative approaches are the intermediation or asset approach, user cost and value-added
definitions (Berger and Humphrey, 1992).³

In the intermediation or asset approach, banks are viewed as financial intermediaries between
depositors and borrowers. Loans and other assets are regarded as bank outputs while
deposits and other liability funds are inputs. Banks are perceived as collecting deposits and
purchasing funds and then intermediating these monies into loans and other assets. Total
costs include all operating and interest costs of the bank.

The user cost approach classifies a bank's output or input based on its net contribution to
bank revenue. Assets with rates of return that exceed the opportunity cost of funds, or
liabilities with cost of debt less than opportunity costs, are considered outputs. Basically, all
other items are considered to be bank inputs. A major problem with this approach involves
the estimation of returns on specific assets and costs of liability funds. Implicit costs and
revenues are difficult to estimate using available accounting data. In general, the user
approach is used in shadow price models (i.e. Fixler and Zieschang).

The third approach used in estimating bank outputs has recently been termed the value added
approach. This approach assumes that all asset and liability categories may have some output
attributes. The determination of outputs in a model are based on those having the largest
impact on value added. The major difference from the intermediation approach is that
demand, time and savings deposits are typically included as outputs since they have a
significant impact on value added.
Empirical Technique

Evanoff and Israilevich provide a thorough review of past studies of bank efficiency. In order to estimate efficiency, a standard must first be defined. For example, the statement that a bank produces 95 percent of maximum output given its input usage assumes that there is a measure for 100 percent, or maximum, output. Therefore, most efficiency analyses develop a frontier production or cost function. The standard estimation of an average production or cost function is usually obtained by statistical techniques that have both positive and negative residuals. However, to measure efficiency, the maximum output or minimum cost is relevant and therefore a frontier must be estimated.

The measurement of bank efficiency using frontier functions can be classified into five basic categories: nonparametric; parametric programming; deterministic parametric; stochastic parametric; and thick frontier. In the nonparametric approach (Färe, Grosskopf and Lovell), linear programming is used to construct a piecewise-linear best practice frontier for each bank. No functional form for the frontier is imposed on the data and therefore the problems associated with imposing the wrong functional form are avoided. However, a problem with this approach is that it is deterministic, that is, all deviations from the frontier are interpreted as inefficiencies.

In the parametric programming approach (Timmer), a functional form for the frontier is imposed with a one-sided error term. A programming technique is then used to estimate the function in such a way as to minimize the one-sided errors. The advantage of the parametric approach is its ability to express the frontier technology in a parametric form and thus obtain an explicit production relation. However, the choice of a specific functional form may be difficult and the parametric programming estimates have no statistical properties (standard errors cannot be obtained through the programming algorithms).

In the deterministic parametric approach (Afriat), a functional form with a one-sided error term is imposed on the data and a distribution for the error term is assumed. Maximum likelihood estimation (MLE) or corrected ordinary least squares (COLS) is used in the estimation. An advantage of this approach is that standard errors can be obtained through the estimation procedure given appropriate distributional assumptions for the one-sided error term. However, the frontier is still deterministic and the selection of the appropriate distribution for the one-sided error may be somewhat arbitrary.

In the stochastic parametric approach (Aigner, Lovell and Schmidt; and Meeusen and van den Broeck), a functional form and a two-part error term is utilized. The composed error term includes a symmetric component which accounts for statistical noise and a one-sided component which captures inefficiency. Again, MLE or COLS is used to estimate the frontier (Førsund, et al.) given appropriate distributional assumptions for the error components. The stochastic model represents a significant improvement in the sense that random shocks outside the decision maker’s control are accounted for in the measurement of efficiency. However, it makes the estimation of the individual firm efficiency measures
more difficult. Because the two error components are added together, some additional assumptions must be made to separate them.5

The thick frontier method (Berger and Humphrey, 1991) uses firms in the lowest average cost quartile to construct a thick frontier. Firms in the highest average cost quartile are regarded as inefficient banks and the differences in the error terms within these quartiles are considered to be random errors. The cost differences between the quartiles are used to estimate efficiencies. However, the ad hoc nature of the estimation procedure provides little information for specific firms.

Selection of Functional Form

The parametric approaches discussed above impose a functional form on the data. Early studies of bank efficiency utilized a Cobb-Douglas form (Benston; Bell and Murphy). However, the use of this log-linear specification does not allow U-shaped long-run cost functions, as indicated in later studies (Gilligan, Smirlock and Marshall; Benston, Hanweck and Humphrey). Three separate, but related, functional forms are now more commonly used to measure bank costs. These are: 1) translog; 2) generalized translog; and 3) minflex Laurent. Lawrence rejects the Box-Cox transformation of the generalized translog, but indicates that translog equations provide an adequate fit of the data. Furthermore, Le Compte and Smith do not provide strong evidence that the minflex Laurent translog is a more adequate functional form than the translog. A criticism of the translog, however, is that the behavior of the remainder term away from the point of approximation may produce biased estimates (Barnett and Lee; Gallant). Guilkey and Lovell use a Monte Carlo experiment to provide evidence that the translog may produce upwardly biased estimates.

Nevertheless, the translog has been used extensively in banking cost studies.6 Its flexible functional form provides a second-order approximation to any arbitrary function and contains both the Cobb-Douglas and Constant Elasticity of Substitution (CES) functions as special cases. Few a priori restrictions are placed on the production technology and it does not restrict scale economies to be the same for all data points.

Entity to Evaluate

Bank expansion regulations are quite different among states. Many states allow unlimited branching while others restrict branching to local geographical boundaries. Expansion in the limited branching states has primarily occurred through holding company acquisitions. Thus, three potential entities can be evaluated in bank efficiency studies -- the branch, bank or holding company. Most previous studies have measured efficiency at the bank level. This is primarily due to the data availability of the Call Reports. Cost data on individual branches is typically unavailable, while data on bank holding companies is not as widely used as Call Report data.
As discussed earlier, most of the consolidation of the banking industry has been due to holding company acquisitions and thus, production efficiency gains may occur at the holding company level. Executives of the holding company may allocate and redirect various operating costs among their affiliates. However, the transactions that occur among the holding company and their affiliated banks may simply be accounting conventions or allocations. These changes in costs may be mistakenly interpreted as changes in inefficiencies at individual institutions. An efficiency evaluation at the consolidated holding company level could potentially alleviate this problem because some of the interbank holding company transactions would be negated.

Time Period Chosen

Because of differences in bank funding mixes, efficiency comparisons among institutions are likely to change in low versus high interest rate environments. The slope of a bank’s cost curve changes as interest rates change. The costs of banks that utilize a higher proportion of purchased funds will be affected more rapidly than those banks which rely relatively more on transaction deposits. The total costs for banks with a higher proportion of purchased funds also will be higher in periods of high interest rates. Typically, large banks rely more heavily on purchased funds while smaller banks tend to rely more on transaction deposits. Hence, the level of inefficiency of the larger banks may change over time based on the level of interest rates. Therefore, the interpretation of factors impacting bank efficiency in a low interest rate environment may be quite different than in a high interest rate environment.

Empirical Analysis of Bank Efficiency

The previous discussion outlined some of the major issues involved in estimating bank efficiency. In order to examine some of these issues, an empirical investigation of agricultural bank efficiency is conducted. Specifically, this section addresses three of the issues -- bank cost definition, bank output definition and empirical technique. These three issues are typically the most widely discussed and sensitivity of these primary issues on efficiency needs to be established.

This section is composed of four parts. First, the data and its sources are discussed. Second, four alternative bank model specifications are defined. These models differ in their specification of bank inputs (costs) and outputs. Third, the two most common empirical techniques for examining bank efficiency, the stochastic parametric and nonparametric methods, are presented and estimated. Finally, the results of these analyses are compared and contrasted.

Data

The data are obtained from quarterly Call Reports from March 1987 through December 1990. The criteria used to select the sample of agricultural banks are:
1) at least 50 percent of the deposits of branches of the bank are NOT located in a metropolitan service area (MSA),
2) the bank has an agricultural loan ratio of at least 10 percent in each quarter from March 1987 through December 1990,
3) the bank has at least $1 million of agricultural loans in each quarter from March 1987 through December 1990.

These criteria insure that the sample banks service the agricultural sector and have done so for at least the past three years. From the 3,490 banks that meet the criteria, 500 banks are randomly selected. The data used in the analyses are based on information from each quarter of 1990. The outputs and inputs prices are based on average values for each of the quarters and thus, the seasonality of the stock values should be adequately represented.

Table 1 provides descriptive statistics for the sample of banks used in measuring efficiency. The average bank size in assets for the sample is approximately $45 million and 70 percent of the banks have total asset values less than $50 million. The average agricultural loan ratio for the sample is 0.41. On average, the banks have a significant share of the market as indicated by an average deposit market share of 22.65 percent. The average value for the Herfindahl-Hirschmann concentration index is 0.2708. The concentration and market share measures indicate that most of the markets are highly concentrated with a few banks controlling a large proportion of the deposits.

Six bank output and four bank input and input price measures are selected for the analysis. These measures are:

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Inputs</th>
<th>Input Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$: Transaction deposits employee</td>
<td>$x_1$: Number of employees</td>
<td>$w_{1i}$: Labor expenses /</td>
</tr>
<tr>
<td>$y_2$: Nontransaction deposits</td>
<td>$x_2$: Occupancy expense</td>
<td>$w_{2i}$: $x_2$ / deposits</td>
</tr>
<tr>
<td>$y_3$: Real estate loans</td>
<td>$x_3$: Other noninterest operating expenses</td>
<td>$w_{3i}$: $x_3$ / deposits</td>
</tr>
<tr>
<td>$y_4$: Nonreal estate wholesale loans</td>
<td>$x_4$: Interest expense on deposits</td>
<td>$w_{4i}$: $x_4$ / deposits</td>
</tr>
<tr>
<td>$y_5$: Nonreal estate retail loans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_6$: Other bank output</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Transaction deposits include all demand deposits, NOW accounts and all other transaction accounts held as deposit liabilities at the bank. Examples of nontransaction deposits include savings deposits, certificates of deposit and all other time-deposits. Real estate loans are all loans secured by agricultural, personal or commercial real estate. Wholesale loans consist of commercial and industrial loans, security loans, loans to other depository institutions and agricultural loans. Retail loans are comprised of the dollar volume of personal, credit card and other loans. Other bank output is a proxy that captures off-balance sheet activity. Off-balance sheet activity has increased substantially over the past few years. These activities include such things as loan sales, letters of credit and securitization operations. The measure equals all noninterest income less service charges received on deposit accounts and gains or losses on securities and foreign exchange accounts. Labor expenses include salaries and benefits. Occupancy expenses are included to represent the cost of capital and include all
noninterest expenses related to the use of premises, equipment, furniture and fixtures. Other noninterest operating expenses includes the cost of materials and all other operating expenses not included in labor and occupancy expenses. Interest expenses include interest paid on all deposit liabilities. Occupancy expenses, other noninterest operating expenses and interest expenses are converted from expenses to prices by dividing by total deposits (Mester).

**Model Specifications**

In order to examine the effect of alternative bank input and output definitions on the measurement of bank efficiency, total cost is modeled as a function of outputs and input prices in four separate ways. The models differ both in the number and type of bank inputs and outputs examined. The four models are:

<table>
<thead>
<tr>
<th>Model</th>
<th>Outputs</th>
<th>Input Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$y_1, y_2, y_3, y_4, y_5$</td>
<td>$w_1, w_2, w_3$</td>
</tr>
<tr>
<td>II</td>
<td>$y_1, y_2, y_3, y_4, y_5$</td>
<td>$w_1, w_2, w_3, w_4$</td>
</tr>
<tr>
<td>III</td>
<td>$y_3, y_4, y_5$</td>
<td>$w_1, w_2, w_3, w_4$</td>
</tr>
<tr>
<td>IV</td>
<td>$y_3, y_4, y_5, y_6$</td>
<td>$w_1, w_2, w_3, w_4$</td>
</tr>
</tbody>
</table>

The models are constructed in order to allow comparisons of alternative approaches to measuring bank inputs and outputs. Models I and II define bank outputs and inputs in terms of the value added approach. Here, transaction and nontransaction deposits are included as bank outputs as value added services. Models I and II differ in that an additional input, interest expenses on deposits, has been included in model II. This allows for the examination of the primary concern regarding bank cost definition, that is, the inclusion of interest expense as a cost item.

Models III and IV define bank outputs and inputs using the intermediation approach. Bank outputs are real estate and nonreal estate loans. Models III and IV differ in that "other bank output" is included in the last model, to determine the effect that off-balance sheet activity has on bank costs.

**Empirical Techniques**

Two methods are used to evaluate bank efficiency, the stochastic cost frontier and the nonparametric approach. The approach to measurement of the bank's underlying production/cost structure is significantly different between the two approaches. The parametric approach assumes that the agricultural bank's cost frontier can be modeled using a translog cost function. Furthermore, the approach employs a composed error structure. Part
of the error term is assumed to account for random occurrences in the cost relationship and
the second part of the error term captures (measures) inefficiency. Hence, particular banks
may be "above" the cost frontier either because of random shocks or inefficiency.

The nonparametric approach avoids the need to specify a particular functional form of the
bank cost relationship. A piecewise-linear frontier is developed for each sample bank
through the use of an intensity vector. The intensity vector is used to construct reference
(frontier) banks. In addition, the nonparametric approach is deterministic, that is, all error
associated with a bank being above the cost frontier is assumed to be caused by inefficiency.
No allowance is made for random occurrences in the underlying cost relationship.

The stochastic cost frontier is estimated using the following translog cost function for z banks
using inputs \( x = (x_1, \ldots, x_n) \) available at prices \( w = (w_1, \ldots, w_m) \) and producing outputs
\( y = (y_1, \ldots, y_m) \):

\[
\ln(TC) = \alpha_0 + \sum_{j=1}^{n} \alpha_j \ln x_j + \sum_{j=1}^{n} \beta_j w_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{j=1}^{n} \alpha_{jj} \ln x_j \ln x_j + \frac{1}{2} \sum_{j=1}^{m} \sum_{j=1}^{m} \gamma_{jj} \ln w_j \ln w_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{j=1}^{n} \delta_{ij} \ln x_i \ln w_j + B_i
\]

where \( TC = w'x \) and the composed error, \( B_i = b_i + \mu_i \). The first part of the composed
error term, \( b_i \), is assumed to follow a half-normal distribution with mode zero and variance
\( \sigma_i^2 (b_i \sim N(0, \sigma_i^2) ) \) and accounts for inefficiency. The second part of the error term, \( \mu_i \),
is assumed to be distributed normally with mean zero and variance \( \sigma^2 (\mu_i \sim N(0, \sigma^2)) \) and
accounts for random disturbances. Homogeneity and symmetry requirements are
incorporated directly into the translog function using the following constraints:

\[
\sum_{j=1}^{n} \beta_j = 1, \quad \sum_{i=1}^{n} \gamma_{ij} = 0, \quad \sum_{j=1}^{m} \delta_{ij} = 0 \quad (\text{Homogeneity})
\]

\[
\alpha_i = \alpha_{ji} \quad \text{and} \quad \gamma_{ij} = \gamma_{ji} \quad (\text{Symmetry})
\]

The Jondrow, et al. decomposition is used to estimate individual bank inefficiency.

Overall inefficiency is estimated using the nonparametric approach by solving a linear
program for each observation (z) of the form:

\[
\min_{x_t} \sum_{j=1}^{n} w_j x_j,
\]

subject to:

\[
y_{it} \leq \sum_{i=1}^{Z} u_z y_{it} \quad i = 1, \ldots, m,
\]

\[
x_{jt} \geq \sum_{i=1}^{Z} u_z x_{it} \quad j = 1, \ldots, n,
\]

Where \( u_z (z=1, \ldots, Z) \) is the intensity vector. The cost minimizing solution vector for input
\[ u_z \geq 0, \quad z = 1, \ldots, Z; \quad \sum_{z=1}^{Z} u_z = 1. \]

price \( w_z \) and output vector \( y_z \) is \( x_z^* \). The ratio \( \frac{w_z^{'}}{w_z} \cdot \frac{x_z^{'}}{x_z} \) measures cost efficiency with \( CE_z \) measuring overall cost inefficiency, or the percentage by which bank costs are increased due to inefficiency.

Inefficiency measures for both the stochastic parametric and nonparametric methods range from zero to infinity. An inefficiency index of zero implies that the bank is operating at minimum costs. An inefficiency index of one implies that bank costs are twice as great as minimum costs.

**Empirical Results**

The primary objective of the empirical analysis is to observe differences that may occur among the different approaches to efficiency measurement and bank input and output definitions. Three methods are used to characterize the inefficiency measurements. First, univariate statistics and histograms are used to describe distributional characteristics. Second, correlation analysis is used to compare the inefficiency measurements. High correlation among the estimates would imply that the different empirical approaches are consistent in their ranking of relatively more or less efficient banks. A linear regression analysis is the final method used to compare the approaches. Previous studies (Aly, et al.; Ferrier and Lovell) have shown that inefficiency estimates are related to specific bank and market characteristics. If these relationships are consistent among the alternative approaches examined, strong support for the policy inferences made in these analyses would be provided. Also, it would suggest that the differences among the approaches do not substantially impact the interpretation of bank and market characteristics that effect the inefficiency of financial institutions.

**Distributional Characteristics**

The summary statistics of the estimated inefficiency measures for the separate models are shown in Table 2. The average inefficiency measures for the nonparametric measures are substantially greater than the stochastic model. The mean inefficiency measures for nonparametric models I through IV are 0.8665, 0.5703, 1.0444 and 0.7971, respectively. However, the mean inefficiency measures for the stochastic models I through IV are 0.0873, 0.0313, 0.3465 and 0.2809, respectively. Model III exhibits the highest inefficiency measures for both empirical methods while the mean estimates for model II are the lowest. The average inefficiency measures for the nonparametric method are very high, particularly for models I, III and IV. It seems somewhat unlikely that these banks could have survived with actual costs in excess of 80-100 percent of minimum costs.
The histograms for each of the models are shown in Figure 1 and serve to emphasize the differences that exist among the approaches. The range of inefficiency measurements is much larger with the nonparametric method than with the stochastic model. This is especially evident with models I and II. The distribution of inefficiency is widely dispersed across the 0 to 1.3 range for the nonparametric model I and across the 0 to 0.9 range for model II. However, with the stochastic technique, the measures for models I and II are highly concentrated at levels less than 0.10.

Models III and IV exhibit more dispersion than models I and II for both empirical methods. However, the difference in dispersion is considerably more evident with the stochastic model. The mode of the stochastic distributions for models III and IV is 0.10 to 0.15, but a substantially larger proportion of the distribution exceeds 0.20 than with models I and II. Furthermore, 114 of the 500 banks in the sample exceeded 1.50 for nonparametric model III.

The results indicate substantial differences in the distribution of inefficiency between models and empirical methods. The nonparametric models exhibit significantly more dispersion than the stochastic models. This may in part be due to the fact that the nonparametric method does not account for statistical noise. Hence, the random occurrences outside of the control of the bank, along with possible measurement errors, may increase (or possibly decrease) individual bank inefficiency estimates.

The inefficiency distributions change only slightly when interest expense is added as an input (model I vs. model II). The average measure decreases and the distributions are skewed somewhat more to the right. This small change could in part be due to the predominance of small banks in the sample. Because of this, the proportions of interest and noninterest bearing deposits may be similar across banks. There are much larger changes in the inefficiency distributions when deposits are not included as bank outputs (models III and IV). These models, which represent the intermediation approach to bank output measurement, are more widely dispersed than models I and II, which measure bank output using the value added approach. Hence, bank output specification impacts both the level and distribution of inefficiency measurements significantly. When models III and IV are compared, the addition of "other bank output" (model IV) tends to cause inefficiency measures to decrease somewhat. Hence, efficiency analyses which ignore off-balance sheet activities may tend to over-estimate bank inefficiency.

**Correlation Analysis**

Correlation coefficients between the different models and empirical techniques are presented in Table 3. Correlations between models I and II and between models III and IV for both efficiency methods are very high (greater than 90 percent in each case). However, other correlation estimates provide additional evidence which indicates considerable disparity among the models. All of the correlations between the stochastic models that include deposits as an output (models I and II) and the stochastic models that include deposits as an output (models III and IV) are very low and not significantly different from zero.
Furthermore, these correlations are only slightly higher for the nonparametric models. This result provides strong evidence that the value added and intermediation approaches, which are shown to result in dramatically different estimates of inefficiency, also rank banks by their inefficiency levels quite differently. Hence, analyses which use different approaches to measuring bank output may result in very different results when bank inefficiency measures are summarized by bank size, location or other characteristics.

**Regression Analysis**

Regression analysis is commonly used in bank efficiency analyses to determine whether inefficiency measurements are related to bank attributes and market characteristics (e.g. Aly, et al.). Testing the relationships across all models and empirical techniques provides additional evidence of similarities and differences.

The bank attributes used in the regression models include bank size in assets, holding company affiliation, degree of agricultural involvement and number of branches. The effect of bank size on efficiency is commonly addressed in efficiency studies. Bank size is measured with two variables -- the log of total assets and the number of branches. Affiliation with a multi-bank holding company also is likely to impact costs (Ellinger and Barry). Therefore, a dummy variable with a value of 1 for banks affiliated with a multi-bank holding company and 0 otherwise is included in the model. Aly, et al. observe that banks with a higher concentration of loans in a specific loan category exhibit higher efficiency. However, a drawback of this specialization is the potentially higher risk associated with the lack of diversification. To observe this potential trade-off, the agricultural loan concentration measure is included in the analysis.

The banking environment and market are characterized by the Herfindahl-Hirschmann index and the bank’s market share. Marketing costs may differ based on the concentration of the market and the relative market share that an individual bank controls. The competitiveness and concentration of the market also may influence customer responses to changes in bank services and prices.

The OLS results of the regression analysis are reported in Table 4. Although low, the explanatory power of the models is similar to those of Aly, et al. Holding company affiliation is the only independent variable that maintains the same sign across all models and both efficiency methods. The negative relationship indicates that banks affiliated with a multi-bank holding company have lower inefficiency ratios than nonaffiliates. The negative relationship is highly significant in each model except the stochastic models I and II.

Log assets, number of branches and the agricultural loan ratio each exhibit both positive and negative relationships that are significant at the 90 percent confidence level. The independent variable log assets switches signs (from negative to positive) between the models that include deposits as outputs (models I and II) and models without deposits as outputs (models III and IV).
The number of branches is positively related to bank inefficiency in models I and II and is highly significant in each of the stochastic models. However, this explanatory variable is much less significant in the nonparametric models. The agricultural loan ratio has a negative and significant relationship in each of the nonparametric models, implying that a higher percentage of agricultural loans decreases bank inefficiency. However, the only stochastic model exhibiting a significant relationship with the agricultural loan ratio is model IV, which results in a positive relationship. The only significant relationships with the market variables occurred with stochastic models I and II.

Although some of the differences in the regression results may be due to model misspecification and multicollinearity, the regression analyses provide additional evidence of the disparity of results that can occur across different bank models and efficiency methods. Studies which use similar analyses to make inferences regarding inefficiencies without investigating the sensitivity of model specification or empirical technique may result in improper policy recommendations.

Conclusions

This study addresses the major issues relating to inefficiency measurement of financial institutions. There are a number of issues that need to be considered when specifying a bank cost relationship and selecting an appropriate empirical methodology to measure inefficiency. This study demonstrates that results, and subsequently, inferences can differ quite substantially among different bank models and empirical efficiency methods. The largest differences occur between empirical methods and bank output definition alternatives.

In general, the nonparametric models result in larger and more disperse measures of bank inefficiency than the stochastic models. Furthermore, inefficiency ratios are not highly correlated between the techniques. Hence, individual banks which are classified in one analysis as highly inefficient relative to other banks may not be comparably ranked when another procedure is used. The inefficiency measurements also differed substantially based on the inclusion (value added approach) or omission (intermediation approach) of deposits as an output measure. On average, the inefficiency measures exhibited larger and more dispersed values in the intermediation models.

The differences in the level, distribution and correlation of inefficiency estimates between model definition and empirical method have important policy considerations. If policy decisions are based upon results from a particular estimation procedure, resources which are targeted towards efficiency-enhancing programs may be misallocated or wasted. Care must be taken to ensure that inefficiency measures and environmental relationships are robust across model definition and methodology. Further research is needed to explain why correlations between inefficiency measures are not higher between models and methods and why the inconsistent results of the bank regression relationships occurred.
Endnotes

1. For a discussion of banking regulation, see Spong.
2. For detailed discussions, see Berger and Humphrey, 1992; Ferrier and Lovell; Humphrey; and Clark.
3. The value-added approach is basically a derivation of the production approach. The primary difference is that the production approach typically uses numbers of accounts while the value-added approach uses dollar amounts.
4. For reviews of efficiency methods see Evanoff and Israilevich; Førsund, Lovell and Schmidt; or Schmidt.
5. For details, see Jondrow, et al.
6. For a review of translog bank-cost models, see Lawrence.
7. The Herfindahl-Hirschmann Index is a market concentration measure used by the U.S. Justice Department in determining excessive concentration in banking markets. The index is:

\[ \text{Herfindahl-Hirschmann Index} = \sum_{i=1}^{k} A_i^2, \]

where there are \( k \) banks serving the market and \( A_i \) represents the percentage of market-area deposits controlled by the \( i \)th bank.
References


Figure 1. Histograms of inefficiency values by model specifications and empirical technique.
Table 1. Description of the banks in the sample, 1990.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y1) - Transaction deposits ($000)</td>
<td>9,436</td>
<td>8,498</td>
<td>865</td>
<td>82,255</td>
</tr>
<tr>
<td>(Y2) - Nontransaction deposits ($000)</td>
<td>28,453</td>
<td>26,132</td>
<td>2,340</td>
<td>294,336</td>
</tr>
<tr>
<td>(Y3) - Real estate loans ($000)</td>
<td>9,246</td>
<td>14,420</td>
<td>237</td>
<td>242,774</td>
</tr>
<tr>
<td>(Y4) - Nonreal estate wholesale loans ($000)</td>
<td>8,446</td>
<td>7,802</td>
<td>76</td>
<td>79,400</td>
</tr>
<tr>
<td>(Y5) - Nonreal estate retail loans ($000)</td>
<td>3,489</td>
<td>4,977</td>
<td>92</td>
<td>69,883</td>
</tr>
<tr>
<td>(Y6) - Other bank output ($000)</td>
<td>251</td>
<td>289</td>
<td>8</td>
<td>2,473</td>
</tr>
<tr>
<td>(X1) - Number of employees</td>
<td>22</td>
<td>19</td>
<td>4</td>
<td>164</td>
</tr>
<tr>
<td>(X2) - Expenses of fixed assets &amp; premises ($000)</td>
<td>144</td>
<td>149</td>
<td>5</td>
<td>1,502</td>
</tr>
<tr>
<td>(X3) - Other noninterest oper. expense($000)</td>
<td>422</td>
<td>455</td>
<td>1</td>
<td>6,877</td>
</tr>
<tr>
<td>(X4) - Interest expense on deposits ($000)</td>
<td>2,315</td>
<td>2,103</td>
<td>193</td>
<td>22,648</td>
</tr>
<tr>
<td>(W1) - Labor expense/employee ($000)</td>
<td>28.9566</td>
<td>7.2995</td>
<td>4.5234</td>
<td>68.1667</td>
</tr>
<tr>
<td>(W2) - X2/total deposits</td>
<td>0.0039</td>
<td>0.0018</td>
<td>0.0003</td>
<td>0.0139</td>
</tr>
<tr>
<td>(W3) - X3/total deposits</td>
<td>0.0117</td>
<td>0.3005</td>
<td>0.0000</td>
<td>0.0534</td>
</tr>
<tr>
<td>(W4) - X4/total deposits</td>
<td>0.0608</td>
<td>0.0055</td>
<td>0.0148</td>
<td>0.0751</td>
</tr>
<tr>
<td>Total assets ($000)</td>
<td>44,503</td>
<td>40,121</td>
<td>4,195</td>
<td>423,819</td>
</tr>
<tr>
<td>Agricultural loan ratio</td>
<td>0.4109</td>
<td>0.1990</td>
<td>0.1026</td>
<td>0.8998</td>
</tr>
<tr>
<td>Agricultural loans ($000)</td>
<td>7,223</td>
<td>5,972</td>
<td>1,148</td>
<td>40,178</td>
</tr>
<tr>
<td>Loan to deposit ratio</td>
<td>0.5465</td>
<td>0.1444</td>
<td>0.1067</td>
<td>0.9620</td>
</tr>
<tr>
<td>Equity to asset ratio</td>
<td>0.0945</td>
<td>0.0285</td>
<td>0.0129</td>
<td>0.2095</td>
</tr>
<tr>
<td>Number of branches</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Market share</td>
<td>0.2265</td>
<td>0.0194</td>
<td>0.0122</td>
<td>1.0000</td>
</tr>
<tr>
<td>Herfindahl Hirschman Index</td>
<td>0.2708</td>
<td>0.1639</td>
<td>0.0482</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Size distribution of banks in the sample

<table>
<thead>
<tr>
<th>Asset Size ($000)</th>
<th>Number of Banks</th>
<th>Percent of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than $10M</td>
<td>26</td>
<td>5.2%</td>
</tr>
<tr>
<td>$11M through $25M</td>
<td>151</td>
<td>30.2%</td>
</tr>
<tr>
<td>$26M through $50M</td>
<td>172</td>
<td>34.4%</td>
</tr>
<tr>
<td>$51M through $100M</td>
<td>118</td>
<td>23.6%</td>
</tr>
<tr>
<td>$101M through $250M</td>
<td>31</td>
<td>6.2%</td>
</tr>
<tr>
<td>greater than $250M</td>
<td>2</td>
<td>0.4%</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Table 2. Overall inefficiency measures by model specification and empirical techniques, 1990.

<table>
<thead>
<tr>
<th>Model</th>
<th>Technique</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Median Value</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model I</td>
<td>Nonparametric</td>
<td>0.8665</td>
<td>0.5629</td>
<td>0.8026</td>
<td>0.0000</td>
<td>4.1896</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>0.0873</td>
<td>0.0497</td>
<td>0.0750</td>
<td>0.0235</td>
<td>0.3666</td>
</tr>
<tr>
<td>Model II</td>
<td>Nonparametric</td>
<td>0.5703</td>
<td>0.3535</td>
<td>0.5466</td>
<td>0.0000</td>
<td>2.5599</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>0.0313</td>
<td>0.0189</td>
<td>0.0264</td>
<td>0.0074</td>
<td>0.1573</td>
</tr>
<tr>
<td>Model III</td>
<td>Nonparametric</td>
<td>1.0444</td>
<td>0.6612</td>
<td>0.9520</td>
<td>0.0000</td>
<td>4.4395</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>0.3465</td>
<td>0.3025</td>
<td>0.2536</td>
<td>0.0351</td>
<td>2.0882</td>
</tr>
<tr>
<td>Model IV</td>
<td>Nonparametric</td>
<td>0.7971</td>
<td>0.5138</td>
<td>0.7249</td>
<td>0.0000</td>
<td>2.9706</td>
</tr>
<tr>
<td></td>
<td>Stochastic</td>
<td>0.2809</td>
<td>0.2252</td>
<td>0.2085</td>
<td>0.0341</td>
<td>1.5191</td>
</tr>
</tbody>
</table>