Consistency Conditions for Regulatory Analysis of Financial Institutions:
A Comparison of Frontier Efficiency Methods

Paul W. Bauer
Federal Reserve Bank of Cleveland
Cleveland, OH 41101-1387

Allen N. Berger
Board of Governors of the Federal Reserve System
Washington, DC 20551
and
Wharton Financial Institutions Center
Philadelphia, PA 19104 U.S.A.

Gary D. Ferrier
University of Arkansas
Fayetteville, AR 72701-1201

David B. Humphrey
Florida State University
Tallahassee, FL 32306

Abstract

We propose a set of consistency conditions that frontier efficiency measures should meet to be most useful for regulatory analysis or other purposes. The efficiency estimates should be consistent in their efficiency levels, rankings, and identification of best and worst firms, consistent over time and with competitive conditions in the market, and consistent with standard nonfrontier measures of performance. We provide evidence on these conditions by evaluating and comparing efficiency estimates on U.S. bank efficiency from variants of all four of the major approaches -- DEA, SFA, TFA, and DFA -- and find mixed results.

I. Introduction

To make informed policy decisions regarding financial institutions, regulators need to have fairly accurate information about the likely effects of their decisions on the performance of the institutions they regulate/supervise.[1] Specifically, the regulators of commercial banks, thrifts, credit unions, and insurance companies should have some expert knowledge based on rigorous empirical research regarding whether the mergers and acquisitions they are petitioned to approve will result in higher or lower costs, and whether the increases in equity capital ratios they may require will raise costs significantly and reduce the supply of intermediation services. Similarly, regulatory authorities should be aware whether the observed managerial inefficiency they may observe could raise the probability of financial institution failure substantially, and so could be used to reallocate scarce supervisory resources to where they are most needed. In addition, regulators should have quantitative evidence on the performance effects of regulatory restrictions on the interest rates and insurance premiums these institutions are allowed to pay and receive, the prudential restrictions on the risks these firms are allowed to bear, the geographic areas they are allowed to serve, and the types of financial services they are allowed to offer. If regulatory authorities do not have the benefit of quality information based upon quantitative research regarding the performance effects of their actions, then their decisions may have the unintended consequences of raising the costs of providing financial services to the public, reducing the quantity or quality of these services, or increasing systemic risk.

In recent years, the academic research on the performance of financial institutions has increasingly focused on frontier efficiency or X-efficiency, which measures 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. That is, the frontier efficiency of an institution measures how well it performs relative to the predicted performance of the "best" firms in the industry if these best firms were facing its same market conditions. Frontier efficiency is superior for most regulatory and other purposes to the standard financial ratios from accounting statements -- such as return on assets (ROA) or the cost/revenue ratio -- that are commonly employed by regulators, financial institution managers, and industry consultants to assess performance. This is because frontier efficiency measures use programming or statistical techniques to try to remove the effects of differences in input prices and other exogenous market factors affecting the standard

---

[1]For convenience, we simply use the term "regulators" to refer to all lawmakers, supervisory agencies, antitrust authorities, etc. that exercise any regulatory or supervisory authority over financial institutions.

performance ratios in order to obtain better estimates of the underlying performance of the managers.

The financial institution efficiency literature is both large and recent -- a review of 130 studies of financial institution frontier efficiency across 21 countries found that fully 116 were written or published during 1992-1997 (Berger and Humphrey 1997). Frontier inefficiency or X-inefficiency of financial institutions has generally been found to consume a considerable portion of costs on average, to be a much greater source of performance problems than either scale or product mix inefficiencies, and to have a strong empirical association with higher probabilities of financial institution failures over several years following the observation of substantial inefficiency.

Frontier efficiency has been used extensively in regulatory analysis to measure the effects of mergers and acquisitions, capital regulations, deregulation of deposit rates, removal of geographic restrictions on branching and holding company acquisitions, etc. on financial institution performance. The main advantage of frontier efficiency over other indicators of performance is that it is an objectively determined quantitative measure that removes the effects of market prices and other exogenous factors that influence observed performance. This allows the researcher to focus on the quantitative effects on costs, input use, etc. that changes in regulatory policy are likely to engender.

Despite intense research efforts, there is no consensus on the best method or set of methods for measuring frontier efficiency, and the choice of method may affect the policy conclusions that are drawn from the analyses. In the past twenty years, at least four main frontier approaches have been developed to assess firm performance relative to some empirically defined "best-practice" standard. These are the nonparametric linear programming approach, often referred to as data envelopment analysis (DEA), and three parametric econometric approaches -- the stochastic frontier approach (SFA), thick frontier approach (TFA), and distribution-free approach (DFA). These approaches differ in the assumptions they make regarding the shape of the efficient frontier, the existence of random error, and (if random error is allowed) the distributional assumptions imposed on the inefficiencies and random error in order to disentangle one from the other. As discussed below, these approaches also often differ in whether the underlying concept analyzed is technological efficiency versus economic efficiency, although this difference need not occur in practice.

In this paper, we argue that it is not necessary to have a consensus on which is the single best frontier approach for measuring efficiency for the efficiencies to be useful for regulatory analysis. Instead, we

propose a set of consistency conditions that efficiency measures derived from the various approaches should meet to be most useful for regulators or other decision makers. The efficiency estimates derived from the different approaches should be consistent in their efficiency levels, rankings, and identification of best and worst firms, consistent over time and with competitive conditions in the market, and consistent with standard nonfrontier measures of performance. Specifically, the consistency conditions are:

(i)   the efficiency scores generated by the different approaches should have comparable means, standard deviations, and other distributional properties;

(ii) the different approaches should rank the institutions in the approximately the same order;

(iii) the different approaches should identify mostly the same institutions as "best practice" and as "worst practice;"

(iv) all of the useful approaches should demonstrate reasonable stability over time, i.e., tend to consistently identify the same institutions as relatively efficient or inefficient in different years, rather than varying markedly from one year to the next;

(v)   the efficiency scores generated by the different approaches should be reasonably consistent with competitive conditions in the market; and

(vi) the measured efficiencies from all of the useful approaches should be reasonably consistent with standard nonfrontier performance measures, such as return on assets or the cost/revenue ratio.

Consistency conditions (i), (ii), and (iii) may be thought of as measuring the degree to which the different approaches are mutually consistent, and conditions (iv), (v), and (vi) may be thought of as measuring the degree to which the efficiencies generated by the different approaches are consistent with reality or are believable. The former are more helpful in determining whether the different approaches will give the same answers to regulatory policy questions or other queries, and the latter are more helpful in determining whether these answers are likely to be correct.

Specifically for the mutual-consistency conditions, if the approaches generate similar distributions of efficiency as in condition (i), then the projected quantitative effects of regulatory policies on performance would be more likely to be similar across the approaches. If the methods all rank the institutions in about the same order as in (ii), then regulatory authorities would generally get the same answer when evaluating whether institutions that had undergone mergers and other regulatory-influenced events had became more or less efficient as a result. As a weaker condition than (ii), if the approaches at least found mostly the same institutions to be in the highest and lowest efficiency groups, as in condition (iii), then regulatory authorities

could draw reasonable conclusions about which operating policies and procedures or managerial control structures were "best-practice" and "worst-practice," and design their policies accordingly. For example, if it were determined that branch banking or universal banking were best practices that consistently maximized measured efficiency across all the approaches, then regulators might be less inclined to put restrictions on branch expansion or circumscribe banking powers. Importantly, all of the efficiency approaches could be mutually consistent as in conditions (i), (ii), and (iii), but still not be very useful if they are not realistic or believable as in conditions (iv), (v), and (vi).

For the consistent-with-reality or believability conditions, if the efficiency scores are stable over time as in condition (iv) instead of the efficiencies bouncing up and down dramatically from year to year, this would be consistent with the likely pattern of true managerial efficiencies over time. Management usually does not turn over often, and even when this occurs, it is difficult to implement new policies and procedures quickly. Similarly, some of efficiency differences may arise from differences in technology that are embodied in durable plant and equipment that may be difficult and costly to replace in the short term. Thus, only in exceptional cases would it be likely that efficiencies would fluctuate markedly over short periods of time. If condition (iv) were met, then authorities could also be more confident that their policies targeted toward either very inefficient or very efficient firms would still generally identify them correctly after normal policy and implementation lags for regulatory actions. In addition, competitive conditions may help limit the range of believable efficiencies in the market, as in condition (v). For instance, if the entry barriers to the industry or local market are not too steep and the market is reasonably unconcentrated, then condition (v) suggests that most firms that remain in business for a long period of time should be reasonably efficient, since competition should drive most of the very inefficient firms out of the industry. Finally, if the efficiencies generated by the different approaches all were positively related to standard financial ratio measures of performance as in condition (vi), then authorities could be more confident that the measured efficiencies were accurate indicators of actual accomplishment, and not just artifacts of the assumptions of the efficiency approaches. It is expected that accurate efficiency measures would have positive rank-order correlations with the standard nonfrontier performance measures, but the correlations should be far from 1.00 because the standard measures embody not only the efficiencies, but also the effects of differences in input prices and other exogenous variables over which financial institution managers have no little or no control.

There is some prior evidence on these points, but there has never been a comprehensive study of financial institutions that has examined all six of these consistency conditions for regulatory usefulness and applied them to all four of the major frontier approaches. The prior evidence suggests that the efficiency scores from the different approaches often yield quite different distributions of measured efficiencies, contrary to condition (i). For example, a comparison of 118 average annual efficiency values from 66 studies of U.S. banks indicated that nonparametric methods such as DEA yielded a lower mean and higher standard deviation than did the parametric methods such as SFA, TFA, and DFA (Berger and Humphrey 1997, Table 2).[2] However, these studies differed in their choice of efficiency concept (technological efficiency versus economic efficiency), samples of banks, time periods chosen, specifications of inputs and outputs, functional forms, and employment of different techniques within each approach, and so do not provide a good experimental design for evaluating whether the efficiency approaches are consistent. For evaluating consistency, it is necessary to hold these other factors constant and apply multiple efficiency methods to the same data set.

There are a few such studies that applied two or more methods to the same data set, and these are reviewed in Section III below. As will be shown, this evidence is very limited at present, and the results are quite mixed across studies. The studies sometimes find the average efficiencies to be similar and sometimes dissimilar across the approaches, and sometimes consistent and sometimes inconsistent with market competitive conditions, yielding ambiguous evidence regarding conditions (i) and (v). These studies have also yielded mixed evidence on the issues of whether the different efficiency approaches rank the best and worst institutions similarly, as in conditions (ii) and (iii). There is also very little evidence on the stability of efficiency, and whether measured efficiencies rank firms in the same order as standard nonfrontier measures of performance, as in conditions (iv) and (vi).

The main purpose of the current study is to add to this limited information set by providing specific evidence on all of the six conditions for regulatory usefulness by evaluating and comparing new efficiency estimates from all four of the major approaches. To be complete, we employ multiple techniques within each of the four approaches, using single-period and panel methods, for a total of nine efficiency techniques

---

[2]The mean and standard deviation of the nonparametric efficiency scores were 72% and 17%, respectively, as opposed to 84% and 6%, respectively, for the parametric estimates.

evaluated. To be sure that the applications are comparable, all nine techniques use the same efficiency concept (economic efficiency), the same sample of banks, the same time period, the same specifications of inputs and outputs, and (for the parametric methods) the same functional form. To be sure that the results do not depend upon any one particular economic environment of the banking industry or any peculiarities of any one small group of banks, we estimate the average efficiency over time of a panel of 683 banks over a 12-year period during which there were significant changes in the banking industry. This experimental design helps assure that the observed differences in efficiency scores reflect the effects of the differences in the measurement techniques, rather than any of these other factors.

Our examination of the consistency conditions is in the spirit of Charnes, Cooper, and Sueyoshi (1988), who advocated the "methodological cross-checking" of results that have policy importance. Our application is also in concordance with Leamer and Leonard (1983) and Leamer (1994), who emphasized assessing the "fragility" of one's results by reporting the results of diverse or "extreme" models to better understand the implications of one's analysis. We believe that our estimation of nine different models using 12 years of data, and applying all six consistency requirements to the nine models qualifies as "extreme," and if there is "fragility" in the findings across efficiency approaches, we are likely to find it. As well, if there are dimensions of consistency in the frontier efficiency approaches, we should also be able to find some evidence of these consistencies.

The remainder of this paper is organized as follows. Section II outlines the four frontier efficiency approaches, and Section III reviews earlier studies that compared two or more approaches. Section IV describes the data set and model specifications, and Section V applies the consistency conditions to the nine efficiency techniques. Section VI pulls all of this information together and draws some conclusions about the consistency of the various frontier efficiency approaches for use in regulatory analysis, and discusses some avenues for future research.

II. The Four Frontier Efficiency Approaches

As noted above, the four frontier approaches differ in the assumptions made about the shape of the frontier, the treatment of random error, and the distributions assumed for inefficiency and random error. These methods also often differ in whether the underlying concept of efficiency is technological or economic, with the nonparametric DEA studies usually measuring technological efficiency and the parametric SFA,

7

TFA, and DFA studies usually measuring economic efficiency. In this section, we briefly review the methods, focusing on the underlying concepts and assumptions, rather than the technical details of the estimation methods which have already been well-explained in several comprehensive surveys.[3]

We begin by discussing the efficiency concepts, and then assess the four frontier approaches. Technological efficiency, or technical efficiency as it is sometimes called, focuses on levels of inputs relative to levels of outputs. To be technologically efficient, a firm must either minimize its inputs given outputs or maximize its outputs given inputs. Economic efficiency is a broader concept than technological efficiency, in that economic efficiency also involves optimally choosing the levels and mixes of inputs and/or outputs based on reactions to market prices. To be economically efficient, a firm has to choose its input and/or output levels and mixes so as to optimize an economic goal, usually cost minimization or profit maximization. Economic efficiency requires technological efficiency as well as allocative efficiency -- i.e, the optimal inputs and/or outputs are chosen based on both the production technology and the relative prices in the market. It is quite plausible that some firms that are relatively technologically efficient are relatively economically inefficient and vice versa, depending upon the relationship between managers' abilities to use the best technology and their abilities to respond to market signals. Therefore, the use of the two different efficiency concepts may give significantly different rankings of firms, even for a given frontier approach. Technological efficiency scores will also tend to be higher than economic efficiency scores on average, all else equal, because economic efficiency sets a higher standard that includes allocative efficiency.

Technological efficiency requires only input and output data, but economic efficiency also requires price data. Most of the early nonparametric frontier models (e.g., Charnes, Cooper, and Rhodes 1978) as well as some of the early parametric frontier models (e.g., Aigner, Lovell, and Schmidt 1977) focused on technological efficiency. In fact, DEA was developed specifically for measuring technological efficiency in the public and not-for-profit sectors, where prices may not be available or reliable, and the assumption of cost minimizing or profit maximizing behavior may not be appropriate (Charnes, Cooper, and Rhodes 1978).

However, in recent efficiency analyses, there is usually a difference in the efficiency concept

---

[3] See, for example, the surveys by Banker, Charnes, Cooper, Swarts, and Thomas (1989), Bauer (1990), Seiford and Thrall (1990), Ali and Seiford (1993), Greene (1993), Grosskopf (1993), Lovell (1993), or Charnes, Cooper, Lewin, and Seiford (1994).

employed between the nonparametric and parametric approaches. Most nonparametric DEA studies continue to apply technological efficiency to inputs and outputs, although a few studies do use cost-based DEA (e.g., Ferrier and Lovell 1990, Ferrier, Grosskopf, Hayes, and Yaisawarng 1993, Cummins and Zi 1998). In contrast, virtually all recent parametric SFA, TFA, and DFA studies employ prices and examine economic efficiency.[4] This means that in most cases, efficiency scores generated by DEA are not fully comparable to those of SFA, TFA, and DFA.

We argue here for the appropriateness of economic efficiency for use in the regulatory analysis of financial institutions. Price data do exist for financial institutions, and cost minimization and profit maximization are likely important behavioral objectives. Moreover, the economic inefficiencies of financial institutions are better measures for regulators to use in evaluating the costs and benefits to society of various policies than are the technological inefficiencies, which do not put value weights on the inputs wasted or outputs not produced. Using technological efficiency in place of economic efficiency, thus neglecting allocative efficiency, would likely increase the level of average efficiency (condition i), affect the overall rankings of financial institutions and the identification of best practice and worst practice firms (conditions ii and iii), and may reduce the consistency of measured efficiency with the state of competition in banking markets and with standard nonfrontier measures of performance (conditions v and vi), which generally depend on economic reactions to market prices. Therefore, in all the empirical applications in this paper (including DEA), we incorporate price data and employ the concept of economic optimization.

We choose cost minimization over profit maximization because it is a more commonly specified and accepted efficiency concept in the literature, and because there are problems measuring output prices from the bank Call Reports during the first part of our sample (prior to 1984). Ideally, both cost and profit specifications would be employed and compared, but examining our six consistency conditions over nine different efficiency techniques already seem to strain the very limits of space and time. We recommend that future investigations follow this path.

---

[4] The move to economic efficiency by the parametric studies may have been motivated in large part by another concern -- the need to account for multiple outputs. Unlike with DEA, this cannot be accomplished in a single parametric production function, but can be handled in cost or profit functions, which would normally include prices as arguments. However, with recent advances in distance function estimation, multiple outputs can now be handled in production settings.

Data Envelopment Analysis (DEA). Nonparametric approaches to measuring efficiency, represented here by DEA (but also including the Free Disposal Hull or FDH), use linear programming techniques. In the usual radial forms of DEA that are based on technological efficiency, efficient firms are those for which no other firm or linear combination of firms produces as much or more of every output (given inputs) or uses as little or less of every input (given outputs). The DEA efficient frontier is composed of these undominated firms and the piecewise linear segments that connect the set of input/output combinations of these firms, yielding a convex production possibilities set.[5] In the version of DEA we apply here which is based on economic efficiency, efficient firms are those which minimize the cost of producing their observed outputs given the best-practice technology and input prices.[6] An obvious benefit of DEA is that it does not require the explicit specification of a functional form and so imposes very little structure on the shape of the efficient frontier.

A potential problem of "self-identifiers" and "near-self-identifiers" may arise when DEA is applied. Under the usual radial forms of DEA, each firm can only be compared to firms on the frontier or their linear combinations with the same or more of every output (given inputs) or the same or fewer of every input (given outputs). In addition, other constraints are often imposed on DEA problems which require comparability with linear combinations of other firms. Other constraints specified in financial institutions research include quality controls, such as the number of branches or average bank account size, or environmental variables, such as controls for state regulatory environment. These other constraints potentially apply to **both** the radial and cost-based forms of DEA. Having to match other firms in so many dimensions can result in firms being measured as highly efficient solely because no other firms or few other firms (and their linear combinations) have comparable values of inputs, outputs, or other constrained variables.[7] That is, some firms may be self-identified as 100% efficient not because they dominate any other firms, but simply because no other firms or linear combination of firms are comparable in so many dimensions. Similarly, other firms may be

---

[5] DEA presumes that linear substitution is possible between observed input combinations on a piecewise linear frontier while FDH presumes that no substitution is possible.

[6] In applying DEA, we followed procedures outlined in Färe, Grosskopf, and Lovell (1994). Variable returns to scale were permitted through use of a side summation restriction in the linear program.

[7] While new procedures have been devised to test for and limit extraneous specification of inputs and outputs or other constraints (e.g., Lovell and Pastor 1997), their application is not yet common.

measured as 100% efficient or nearly 100% efficient because there are only a few other observations with which they are comparable. The problem of self-identifiers and near-self-identifiers most often arises when there are a small number of observations relative to the number of inputs, outputs, and other constraints, so that a large proportion of the observations are difficult to match in all dimensions. Some empirical evidence from the literature on this point is presented below.

Our DEA application tries to minimize the self-identifier problem in three ways. First, we use input prices in a cost-based DEA methodology. In the usual radial input-based (output-based) DEA applications, input mix (output mix) is held constant, so firms with unusual input (output) mixes may be found to be self-identifiers or near-self-identifiers. We can compare any input mix in our application by combining input prices and quantities and comparing total costs, rather than having to compare firms in every input dimension. Second, we do not impose any extra constraints on our DEA problem, so firms only have to minimize costs relative to other firms or linear combinations of firms producing the same output bundle. By specifying costs and by imposing no extra constraints, we can only have self-identifiers or near-self-identifiers to the extent that the output bundles of some banks cannot be easily replicated by linear combinations of other banks.[8] Third, we use a relatively large number of observations relative to the small number of constraints in our DEA problems, so that most firms will have quite a few linear combinations of other firms that are comparable. Specifically, we solve cost minimizing linear programming problems with data on 683 observations for each single year (DEA-S), specifying 4 outputs and no other constraints. We also combine all 12 years of data into a panel (DEA-P), where the reference set is constant over the entire period, for a total of 8196 observations.[9]

One potential problem with DEA that we do not try to solve is that DEA usually does not allow for random error due to measurement problems associated with using accounting data, good or bad luck that temporarily raises or lowers inputs or outputs, or specification error such as excluded inputs and outputs and imposing the piecewise linear shape on the frontier. Any random errors that do exist may be counted as differences in efficiency by DEA. Presumably, this would result in lower average efficiency, as there will

---

[8]The comparability problems for both inputs and outputs could be solved by using a profit-based DEA approach (e.g., Färe and Whittaker 1996).

[9]In our empirical application of cost-based DEA, none of the banks that were identified as technologically efficient were found to be cost efficient. This suggests that including prices and accounting for allocative inefficiency helps ameliorate the potential self-identifier problem.

be more dispersion in the data, unless there is some unusual statistical association between random error and "true" efficiency. This effect may be quite large, since the random error in a single observation on the efficient frontier will affect the measured efficiency of all of the firms that are compared to any linear combination on the frontier involving this firm.[10]

Stochastic Frontier Approach (SFA). The parametric methods -- SFA, TFA, and DFA -- have a disadvantage relative to the nonparametric methods of having to impose more structure on the shape of the frontier by specifying a functional form for it. As noted above, we choose a cost function specification here. However, an advantage of the parametric methods is that they allow for random error, so these methods are less likely to misidentify measurement error, transitory differences in cost, or specification error as inefficiency. The primary challenge in implementing the parametric methods is determining how best to separate random error from inefficiency, since neither of them are observed. The parametric methods SFA, TFA, and DFA differ in the distributional assumptions imposed to accomplish this disentanglement.

SFA employs a composed error model in which inefficiencies are assumed to follow an asymmetric distribution, usually the half-normal, while random errors are assumed to follow a symmetric distribution, usually the standard normal (Aigner, Lovell, and Schmidt 1977). That is, the error term from the cost function is given by $\epsilon = \mu + \upsilon$, where $\mu \geq 0$ represents inefficiency and follows a half-normal distribution, and υ represents random error and behaves according to a normal distribution. The reasoning is that inefficiencies cannot subtract from costs, and so must be drawn from a truncated distribution, whereas random error can both add and subtract costs, and so may be drawn from a symmetric distribution. Both the inefficiencies μ and the random errors υ are assumed to be orthogonal to the input prices, output quantities, and any other cost function regressors specified. The efficiency of each firm is based on the conditional mean (or mode) of inefficiency term μ, given the residual which is an estimate of the composed error ϵ.

Greene (1990) and others have argued that alternative distributions for inefficiency may be more appropriate than the half-normal, and the application of different distributions sometimes do matter to the average efficiencies found for financial institutions (e.g., Yuengert 1993, Mester 1996, Berger and DeYoung

---

[10] There are some efforts to deal with random error in DEA using bootstrapping to gain statistical inference (e.g., Simar and Wilson 1995, Ferrier and Hirschberg 1997) and chance-constrained programming to reduce the effects of noise (e.g., Land, Lovell, and Thore 1993). See Grosskopf (1996) for a survey of these approaches.

1997). We argue here that **any** distributional assumptions simply imposed without basis in fact are quite arbitrary and could lead to significant error in estimating individual firm efficiencies. For example, the half-normal assumption on the inefficiencies μ imposes that most of the firms are clustered near full efficiency, but there is no theoretical reason why inefficiencies could not be more evenly distributed, or distributed close to symmetrically like the assumed distribution of the random error. In fact, prior studies using the DFA approach (described below) -- which imposes no shape on the distribution of inefficiencies -- suggested that the inefficiencies behaved more like symmetric normal distributions than half-normals (Bauer and Hancock 1993, Berger 1993).[11]

Despite these potential problems with measuring the **levels** of efficiency, one positive aspect of the SFA approach is that it will always **rank** the efficiencies of the firms in the same order as their cost function residuals, no matter which specific distributional assumptions are imposed. That is, firms with lower costs for a given set of input prices, output quantities, and any other cost function regressors will always be ranked as more efficient, since the conditional mean or mode of μ (given the estimate of the residual ϵ) is always increasing in the size of the residual. This property of SFA has intuitive appeal for a measure of performance for regulatory purposes -- a firm is measured as high in the efficiency rankings if it keeps costs relatively low for its given exogenous conditions. This is likely to prove helpful in meeting our consistency conditions, which are primarily based on rank orderings.

In our empirical application of SFA, we use the half-normal - normal assumptions on the inefficiencies and random error, since these are the most common assumptions in the literature, and leave examination of our consistency conditions for other distributions for future research efforts. Similar to DEA-S and DEA-P, we apply SFA to both each single year of data separately (SFA-S) and to the 12-year panel as a whole (SFA-P). SFA-S allows the coefficients of the translog cost function to vary over time, while the SFA-P holds the slope coefficients fixed over time and allows the cost function intercepts to vary over time with changes in technology, regulatory environment, and the macroeconomy.

---

[11]To investigate this issue further, we tested the distributions of efficiency scores from all nine techniques employed here for symmetry using the nonparametric test proposed by D'Agostino, Balanger and D'Agostino (1990). In all cases, the null hypothesis of symmetry was rejected. This is not surprising for the SFA results, since an asymmetric distribution was imposed on these scores, but the other seven rejections suggest that the underlying efficiencies may not be distributed close to symmetrically as was found in the prior studies.

Thick Frontier Approach (TFA). TFA uses the same functional form for the frontier cost function as SFA, but is based on a regression that is estimated using only the ostensibly best performers in the data set -- those in the lowest average cost quartile for their size class.[12] Parameter estimates from this estimation are then used to obtain estimates of best-practice cost for all of the firms in the data set (Berger and Humphrey 1991). Banks in the lowest average cost quartile are assumed to have above-average efficiency and to form a "thick frontier."

As it is usually implemented, TFA assumes that deviations from predicted performance values within the highest and lowest performance quartiles of firms represent only random error, while deviations in predicted performance between the highest and lowest average-cost quartiles represent only inefficiencies (a special case of composed error) plus exogenous differences in the regressors. Measured inefficiencies thus are embedded in the difference in predicted costs between the lowest and highest cost quartiles. This difference may occur in either the intercepts or in the slope parameters.

In most applications, TFA gives an estimate of efficiency differences between the best and worst quartile to indicate the general level of overall efficiency, but does not provide point estimates of efficiency for all individual firms. In our application, we need to obtain efficiency estimates for each bank in each time period so that we can compare these estimates to our other frontier efficiency methods. This requires an adjustment. The thick frontier is estimated from data limited to only the lowest cost quartile of banks for each size class (as is standard). A separate efficiency term for every bank (including banks not in the thick frontier) is calculated using a method very similar to the DFA estimates described below. The estimated residuals for the entire sample are calculated and it is assumed that the inefficiency disturbances are uncorrelated with the regressors, so that a separate intercept for each bank can be recovered as the mean of its residuals. The most efficient 1% of the sample (7 banks) are assumed to be fully efficient and their average residuals are truncated to be at the 1% point of the sample distribution, and the efficiency of each bank is determined from the difference from the frontier in these average residuals. The TFA efficiency

---

[12] Banks are first stratified into 8 asset size classes and their average cost over the entire time period (measured here as total cost per dollar of assets) is computed. Those banks in each size class with the lowest average cost form the subset of the data used to estimate the thick frontier for each year separately or for all years together. This ensures that an equal number of banks of all size classes are included in the estimation.

estimates from the panel data set (TFA-P) are based on one set of parameter estimates over the entire time period (corrected for first-order serial correlation), and the TFA efficiency estimates for each year (TFA-S) estimate the cost function parameters separately for each year.

As was the case for SFA, the **levels** of efficiency generated by TFA are potentially suspect, since they are based on rather arbitrary assumptions -- that the lowest average cost quartile within each size class is an adequate "thick frontier" of efficient firms, etc. Nevertheless, there are again reasons for optimism regarding the **rank orderings** generated by TFA. Since the efficiency orderings are determined by cost function residuals after controlling for input prices, output quantities, and possibly other factors, they have intuitive appeal, and are likely to be very consistent with the SFA estimates and other measures of performance.

Distribution-Free Approach (DFA). DFA specifies a functional form for the cost function as does SFA and TFA, but DFA separates inefficiencies from random error in a different way. It does not impose a specific shape on the distribution of efficiency (as does SFA), nor does it impose that deviations within one group of firms are all random error and deviations between groups are all inefficiencies (as does TFA). Instead, DFA assumes that there is a "core" efficiency or average efficiency for each firm that is constant over time, while random error tends to average out over time (Schmidt and Sickles 1984, Berger 1993). Unlike the other approaches, a panel data set is required, and therefore only panel estimates of efficiency over the entire time interval are available (DFA-P). These estimates may be derived using three different techniques.

The first DFA technique, DFA-P WITHIN, is a fixed-effects model which estimates inefficiency from the value of a firm-specific dummy variable (derived by estimating with all the cost function variables measured as deviations from firm-specific means). Efficiency is estimated using the deviation from the most efficient firm's intercept term. A single set of parameters are obtained so inefficiency is fixed over time. However, since inefficiency is no longer a separately specified element in a composed error term, we do not need an assumption that inefficiency is uncorrelated with the regressors (as in SFA) and we adjust for possible first-order serial correlation.

The second DFA technique, DFA-P GLS, applies generalized least squares to panel data, obtains a single set of parameters, assumes that bank inefficiencies are fixed over time,[13] and that inefficiency is

---

[13] This assumption is not strictly necessary. Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990), and Battese and Coelli (1992) generalized the approach to allow inefficiencies to vary over time, but in a

uncorrelated with the regressors. In our cost function, which is also corrected for first-order serial correlation, a separate intercept for each firm is recovered from the panel estimates as the average residual for that firm over the time period. The firm with the smallest average residual is presumed to be the most efficient firm and the inefficiency of all the other firms is measured relative to this benchmark.

The third DFA technique, DFA-P TRUNCATED, estimates the cost function separately for each year. The efficiency estimates are based on the average residuals for each bank. Since some noise might also be persistent over time, we follow Berger (1993) and truncate the residuals at both the upper and lower 1% of the distribution, thus limiting the effects of extreme average residuals at both ends.

As with the other efficiency approaches, there is concern that the levels of the DFA efficiency estimates may be influenced by the somewhat arbitrary assumptions. The measurement of the "core" efficiency means that efficiency variations over time for an individual firm tend to be averaged out with the random error. DFA also implicitly assumes that inefficiency is the only time-invariant fixed effect. If there are other factors that are persistently affecting a firm's costs that are not included in the regression model, such as being in a high-crime location, this may be counted as inefficiency (although this would affect all the other frontier approaches as well).

Nonetheless, similar to SFA and TFA approaches, DFA is intuitively appealing as a measure of economic performance because it is based on keeping costs low for a given set of outputs and input prices over a long period of time and over many changes in economic conditions. We therefore expect the DFA efficiency ranks to be highly correlated with SFA and TFA ranks and other measures of bank performance.

III. Results from Earlier Efficiency Comparisons

Although there is a large literature on financial institution efficiency, there is not much information available on our consistency conditions because most studies applied a single efficiency approach and these conditions are best analyzed by comparing the application of multiple approaches to a single data set. A few studies did compare multiple techniques, usually applying two efficiency methods to the same data set. The comparisons of bank efficiencies using more than one approach include Ferrier and Lovell (1990), Bauer, Berger, and Humphrey (1993), Hasan and Hunter (1996), Berger and Mester (1997), Eisenbeis, Ferrier, and

structured manner.

Kwan (1997), Resti (1997), and Berger and Hannan (forthcoming).[14] We briefly examine some of the evidence from these studies here.[15]

The studies by Bauer, Berger, and Humphrey (1993), Hasan and Hunter (1996), Berger and Mester (1997), and Berger and Hannan (forthcoming) compared estimates using two or more of the parametric approaches. In most cases, these studies found that average efficiencies were comparable and reasonably consistent with competitive conditions in the banking industry -- supporting consistency conditions (i) and (v) above -- but there were exceptions. Hasan and Hunter (1996) found SFA average efficiency values to be much higher than TFA, .81 versus .67, respectively, and Berger and Hannan (forthcoming) found SFA average efficiencies of .92 to be quite a bit higher than the .70 average for DFA. All of these studies found that the parametric approaches tended to rank the banks similarly and identify the same ones as highly efficient and inefficient -- supporting consistency conditions (ii) and (iii) -- but again there were differences of degree. For example, Berger and Mester (1997) found a rank-order correlation of .988 between SFA and DFA efficiencies, but Bauer, Berger, and Humphrey (1993) found that the two methods identified the same banks in the most and least efficient 25% of the banks 38% of the time and 46% of the time, respectively, not all that much higher than the 25% correspondence that would be expected by chance alone. When consistency conditions (iv) and (vi) were looked at in these and other studies, the limited evidence suggested that the parametric approaches appeared to be yield efficiencies that persisted over several years (e.g., Berger and Humphrey 1991,1992, Eisenbeis, Ferrier, and Kwan 1997, DeYoung 1997), and these efficiencies were related in the expected way (although not always strongly) with standard, nonfrontier measures of performance such as return on assets (e.g., Berger and Humphrey 1991, Berger and Mester 1997, Eisenbeis, Ferrier, and Kwan 1997).

Perhaps more interesting are the comparisons of bank efficiencies between nonparametric and parametric approaches, which are really much more dissimilar from each other than the parametric approaches are from one another. DEA and SFA were compared by Ferrier and Lovell (1990), Eisenbeis, Ferrier, and

---

[14] A few frontier model comparisons have also been made using data for other financial institutions, such as bank branches (Giokas 1991), insurance firms (Fecher, Kessler, Perelman, and Pestieau 1993, Yuengert 1993, Cummins and Zi 1998), mutual funds (Ferrier and Philpot 1994), and Federal Reserve offices (Bauer and Hancock 1993).

[15] Some additional summary details may be found in Berger and Humphrey (1997).

Kwan (1997), and Resti (1997). These studies reported fairly close average efficiencies generated by the two approaches. However, this belies the potential problem that the levels of efficiency under DEA may be sensitive to "self-identifiers" or "near-self-identifiers" when there are too few observations relative to the number of constraints in DEA. There is some empirical evidence that this problem may have occurred. For example, Ferrier and Lovell (1990) found that the average efficiency level rose from 54% to 83% when constraints on number of branches and average account sizes were added to the model, keeping the same number of observations. Since the average efficiency for SFA was 79% and the average efficiency for DEA is somewhere between very low (54%) and relatively high (83%), the question of whether DEA and SFA yield similar distributions of efficiency that are consistent with competitive conditions in the banking industry -- as in consistency conditions (i) and (v) -- remains open.[16] With regard to consistent rankings -- as in conditions (ii) and (iii) -- the results from the literature are contradictory. Resti (1997) found very high rank-order correlations between DEA and SFA of .73 to .89, and Eisenbeis, Ferrier, and Kwan (1997) found fairly high rank correlations ranging between .44 and .59, but Ferrier and Lovell (1990) found rank-order correlation of only .02, which was not significantly different from zero. With respect to consistency conditions (iv) and (vi), the very small amount of evidence suggested consistency over time, and very low, but positive correlation with nonfrontier measures of performance (Eisenbeis, Ferrier, and Kwan 1997).[17]

---

[16] Additional empirical evidence on this question comes from studies of bank branches, where there are often small numbers of observations employed in DEA analyses. For example, several DEA studies of bank branches used 35 or fewer observations and large numbers of inputs and outputs and usually found most branches to be either 100% efficient or very close to it (Sherman and Gold 1985, Parkan 1987, Oral and Yolalan 1990, Vassiglou and Giokas 1990, Giokas 1991, and Pastor 1993), whereas a DFA study found much lower average efficiencies (Berger, Leusner, and Mingo 1997).

[17] There is also some mixed evidence regarding the consistency of estimates within the same technique applied to the same data set, where some of the assumptions or methods are altered. Some of these studies found strong consistency. For example, Berger and Mester (1997) found the DFA efficiency estimates to be robust to most changes in specification, Maudos (1996) found very high rank-order correlations of efficiencies generated using different distributional assumptions on the inefficiencies under SFA (.86 to .99), and Ferrier, Kerstens, and Vanden Eeckaut (1994) obtained correlations between .87 and .99 when applying four different radial and nonradial DEA procedures. However, other studies found less consistency. For example, Berger and DeYoung (1997) found the average efficiencies to differ significantly when the specifications of the inefficiency distribution and the functional form for the cost function under SFA were altered, DeBorger, Ferrier, and Kerstens (forthcoming) compared radial and nonradial technical efficiency using input-based and output-based FDH and found rank correlations to vary substantially between .32 and .96, and DeYoung (1998) found very different identification of the best and worst performance groups under TFA, depending upon whether average costs versus supervisory ratings of management (the M in CAMEL)

Thus, the evidence is quite limited and sometimes contradictory on the extent to which the efficiency approaches pass our consistency conditions for use in regulatory analysis, or which subset of them may pass and which may fail. We therefore proceed with our empirical analysis of the four major methods using nine techniques applied to a large data set of banks over an extended period of time in order add to the evidence.

IV. Data and Specification Issues

Consistent with the spirit of Charnes, Cooper, and Sueyoshi (1988), Leamer and Leonard (1983), and Leamer (1994) -- who argued for using diverse or extreme conditions for evaluating models -- we choose a long and turbulent time period with many regulatory changes and many changes in market conditions for evaluating our consistency conditions. Our data set is composed of 683 U.S. banks over the 12-period 1977-88. All banks have over $100 million in assets, come from branch-banking states, and were in continuous operation over the entire period. As a group, the banks account for over two-thirds of all assets in the U.S. banking system. In addition, since all states now allow branch banking, our results may be taken as fairly representative of the banking system as a whole.[18]

This time period is one of many changes in the U.S. banking industry. During the late 1970s, rapid inflation and financial market innovation in the areas of cash management and money market mutual funds expanded competition in bank corporate and consumer markets from less-regulated financial intermediaries, who were not subject to restrictions on deposit rates. In the early 1980s, deposit interest rates and account types at banks and savings and loans were substantially deregulated and bank charters became more freely issued, leading to higher costs and further competition among financial institutions. As well, 20 of the 51 states (District of Columbia counted as a state) either relaxed or eliminated remaining geographic restrictions on branching within the state, and 43 of the 51 removed barriers to interstate banking through holding company acquisitions during this period (Berger, Kashyap, and Scalise 1995, Table B6). In addition, the entry of foreign banks into the market for nonfarm, nonfinancial corporate debt during this period

---

was used to determine the groups.

[18] If failed banks had been included in the data set, it is likely that our efficiencies would have been lower on average. Failing banks and thrifts typically have lower than average efficiency levels than other banks, although it is a matter of some controversy the extent to which the inefficiencies cause the failures versus high costs created by dealing with problem loans just before failure cause measured efficiency to be downwardly biased (e.g., Berger and Humphrey 1992, Cebenoyan, Cooperman, and Register 1993, Hermalin and Wallace 1994, Barr, Seiford, and Siems 1994, Berger and DeYoung 1997).

dramatically reduced the market share of U.S. banks and likely reduced margins on the loans that domestic

banks continued to make. This time interval also witnessed a substantial amount of technological and

financial innovation, starting with the substitution of ATMs for human tellers in the late 1970s, and the

development and refinement of derivative contracts and other products of financial engineering during the

1980s. As a result of all of these changes, many banks performed well, many performed poorly, and a total   ---

of over 800 banks failed over this period.

Thus, the period 1977-88 for U.S. banks is almost an ideal interval to determine how the different

frontier models identify and measure bank efficiency over a variety of extreme conditions. As well, it is

under such extreme conditions that it is most important for regulators to be able to evaluate the effects of their

policies.

Table 1 shows the main variables employed in the various frontier efficiency estimations. We specify

the same four banking outputs and same four inputs in all of our frontier models, whether estimated for a

series of single years or pooled within a panel data set. The outputs are demand deposits, real estate loans,

commercial and industrial loans, and installment loans, all measured in real dollar terms. Production of

services in these account categories is associated with the vast majority of banking costs. The four banking

inputs specified are labor, physical capital, small denomination time and savings deposits, and purchased

funds. The parametric approaches use only the input prices, whereas our cost-based DEA techniques specify

both input quantities and prices.[19] In all cases, total costs -- operating expenses from the physical inputs of

labor and capital, plus interest costs from the financial inputs of time and savings deposits and purchased

funds -- are included to measure total cost efficiency.

The outputs and inputs chosen are fairly basic and standard, although there is considerable variation

within the literature, and many studies add other bank outputs (e.g., off-balance sheet activities), other inputs

(e.g., financial equity capital), other bank characteristics (e.g., nonperforming loans), and environmental

factors (e.g., state income growth) to the models. The specification of the cost function for the parametric

---

[19]The input prices are not directly observed, and so must be constructed from the available information
by dividing flows of expenditures by stocks. The price of labor equals salaries and benefits divided by the
number of full time equivalent workers. The price of physical capital is expenditures on equipment and
premises divided by the book value of physical assets, and the prices of time deposits and purchased funds
are the interest expenses on these categories divided by the dollars in these accounts. These procedures create
data errors and likely account some of the substantial variation in prices shown in Table 1.

models is also fairly basic and standard, a translog cost model with partially-restricted share equations.[20] As above, we choose the most standard specifications from the literature, and are prevented by space and time constraints from trying all of the interesting variations on our nine separate models in evaluating our six separate consistency conditions.[21] We recommend that future research try to verify or overturn our results with robustness checks using more and perhaps better specifications of the outputs, inputs, and functional forms.

V. Examination of the Consistency Conditions

The data presented in Tables 2 - 6 and Figures 1 and 2 provide direct evidence on our consistency conditions for the nine efficiency techniques, arranged in order of the conditions.

Consistency Condition (i) -- Comparisons of Efficiency Distributions with Each Other. A number of distributional characteristics of the efficiency scores generated by the nine efficiency techniques are reported in Table 2. The mean efficiency from the seven SFA, TFA, and DFA parametric models averaged .83 (with a mode of .84), while mean efficiency averaged only .30 (mode of .21) across the two nonparametric DEA models. The average standard deviation of efficiency estimates from the parametric models (.06) was less than one-half that for the nonparametric models (.14).

The level and time pattern of mean efficiency for each frontier method over our 12-year period are displayed in Figure 1. By assumption, the three DFA-P panel methods estimate only a single "core efficiency" over time, and so yield flat lines by construction, but their levels can still be compared with the other methods. As shown, the parametric methods generally yield relatively high mean efficiencies, between about 80% and 90%, that are reasonably close to one another in terms of level, and do not vary much over time even when data for separate years (S) are used. The only significant exception is TFA-S, which has

---

[20] The total cost function was jointly estimated with n-1 of the cost share equations with the standard cross-equation Shephard's Lemma restrictions on the slope parameters imposed. The intercepts of the share equations were allowed to vary to incorporate allocative inefficiencies. Berger (1993) found that efficiency estimates using no share equations, share equations like this with the intercepts free to vary, and fully restricted share equations gave very similar efficiency results.

[21] Some recent frontier efficiency studies use more globally flexible functional forms, such as the Fourier-flexible specification (e.g., Spong, Sullivan, and DeYoung 1995, Berger, Cummins, and Weiss 1997, Berger and DeYoung 1997, Berger, Leusner, and Mingo 1997, Berger and Mester 1997, DeYoung, Hasan, and Kirchhoff 1998).

mean efficiency of 67.4% and varies quite a bit over time. It appears that using only 25% of the data from a single year to estimate the cost functions adds significant noise to the model, but that this problem is solved by using information from the other years, as indicated by the TFA-P results.

The most striking result from Table 2 and Figure 1 is how much lower the efficiencies from the DEA approaches are. The mean efficiencies from DEA-S and DEA-P are substantially below the efficiencies from all the parametric methods. This inconsistency between the distributions of the DEA and the parametric distributions of efficiency is further illustrated in Figure 2, which shows the cumulative distribution functions for the efficiencies from one panel parametric technique (DFA-P GLS) and one panel nonparametric technique (DEA-P).[22] This shows that the relatively low mean efficiency for the DEA methods is manifested in low efficiencies for the great majority of the banks. The nonparametric method identifies about 90% of the banks as having less than 30% efficiency, while the parametric method suggests a much closer correspondence of efficiency across observations, with almost all of the firms near 90% efficiency.[23]

These data suggest that the parametric methods are generally consistent with one another in terms of the distributions of the efficiencies generated, yielding relatively high efficiencies for the vast majority of firms and the nonparametric methods are generally mutually consistent, yielding relatively low efficiencies for most firms. The determination of which set of methods may be more useful for regulatory analysis or other uses must wait for evaluation of the other consistency conditions, particularly which approaches yield more realistic or believable efficiency estimates.

Consistency Condition (ii) -- Rank-Order Correlations of the Efficiency Distributions. Although estimates of the **levels** of cost efficiency for the parametric and nonparametric frontier methods are quite different across banks, it is still possible that these methods will generate similar **rankings** of banks by their efficiency scores across frontier methods. As discussed above, identifying the rough ordering of which

---

[22]The maximum value for DEA-P efficiency shown in Figure 2 is less than 1.00 because it is the average efficiency over time for each bank, and no bank was on the DEA panel frontier in every period.

[23]It is also notable that the efficiency scores are generally not very strongly affected by the choice of a single-year versus panel method or other differences in technique within each approach, with the possible exception of TFA. The nonparametric Chi-square and Kolmogorov-Smirnov tests can be used to test whether a pair of samples share a common distribution. Both test procedures failed to reject the null hypotheses that each of the pairs DEA-S and DEA-P, SFA-S and SFA-P, TFA-S and TFA-P, and DFA-P WITHIN and DFA-P GLS belong to the same population.

financial institutions are more efficient than others is usually more important for regulatory policy decisions than measuring the level of efficiency, so that regulators can determine whether regulatory-influenced events like mergers result in improved or worsened financial institution firm efficiency. If the methods do not rank institutions similarly, then policy conclusions may be "fragile" and depend on which frontier efficiency approach is employed.

Table 3 contains Spearman rank-order correlation coefficients showing how close the rankings of banks are among each of the nine frontier methods using the full sample of banks. The ranking for each method is based on the average efficiency value for each bank over the entire 12-year period. It would be expected that the rankings among all seven of the parametric methods would be relatively high, since all of these methods essentially rank the banks by teasing efficiencies from random error in the residuals from similarly specified cost functions. Indeed, the average rank-order correlation among these seven methods is .756, and all of these correlations are statistically significant at the 1% level. We would also expect a relatively high rank correlation among the two nonparametric methods, since they also generate efficiencies from essentially the same model. Again, this expectation is justified, as the rank-order correlation between them is a statistically significant .895.

However, the data suggests that the DEA and the parametric techniques give only very weakly consistent rankings with each other. The average rank-order correlations between the parametric and nonparametric methods is only .098. Ten of the fourteen correlations are positive and statistically significant, two are negative and statistically significant, and two are not statistically significantly different from zero.[24] Thus, the DEA and the parametric models cannot be relied upon to generally rank the banks in the same order, and so may give conflicting results when evaluating important regulatory questions.

Consistency Condition (iii) -- Identification of Best-Practice and Worst-Practice Firms. As discussed above, even if the methods do not always rank the financial institutions similarly, they may still be useful for some regulatory purposes if they are consistent in identifying which are most efficient and least efficient institutions. The upper triangle of the matrix shown in Table 4 reports for each pair of frontier

---

[24]The nonparametric Kruskal-Wallis test can be used to test whether multiple samples are drawn from the same population. Given the lack of consistency between the parametric and nonparametric techniques, it is not surprising that a Kruskal-Wallis test rejected the null hypothesis that the nine sets of efficiency scores all were drawn from the same distribution.

efficiency techniques, the proportion of banks that are identified by one technique as having efficiency scores in the top 25% that are also identified in the top quarter by the other technique. For example, of the banks identified as in the best-practice 25% by DEA-S, 35.7% of these same banks were also identified as being in the top quarter by SFA-S. This number also describes the proportion of the best quarter of firms as identified by SFA-S that are also in the top 25% by DEA-S, since the number of banks in the top 25% is always the same (171 of 683 banks). Random chance alone would yield an expected value of a 25.0% correspondence, and the value of .357 shown in the table is not statistically significantly different from .250. The same analysis with respect to the lowest efficiency 25% of banks -- the "worst-practice" -- is shown in the lower triangle of the table.

Table 4 tells essentially the same story as the rank-order correlations above -- there is very good consistency among the seven parametric techniques, very good consistency between the two nonparametric techniques, but poor consistency between the parametric and nonparametric methods. Within the parametric techniques, the correspondence of the best practice 25% of banks ranges from 49.1% to 93.0%, is always statistically significantly higher than .250 at the 1% level, and averages a 69.3% correspondence. Similarly, among the SFA, TFA, and DFA methods, the joint identification of the worst practice 25% of banks ranges from 49.1% to 89.5%, is always statistically significantly different from random chance, and gives an average correspondence of 69.2%. The two DEA methods identify the best- and worst-practice quarter of the banks identically 76.0% and 74.9% of the time, respectively, and both are statistically significantly greater than .250.

In contrast, the correspondences of the best-practice and worst-practice banks between the two DEA methods and the seven parametric methods goes only as high as 37.4%, and is below the random expectation of 25% in several cases. For best-practice, the average correspondence is 31.1%, and for worst-practice, it is 32.8%, and in no cases are the correspondences statistically significantly different from .250. Thus, although the parametric methods tend to identify the same firms as efficient and inefficient and the nonparametric methods are also internally consistent in this regard, the two types of approaches are not consistent in their identification of the best-practice and worst-practice firms. As a result, regulatory policies targeted at either efficient or inefficient firms would hit different targets, depending upon which set of frontier efficiency approaches were used to frame the policy.

Consistency Condition (iv) -- The Stability of Measured Efficiency Over Time. As discussed above, to be useful for regulatory policy purposes, it is important that the efficiency measures demonstrate reasonable stability over time, and do not vary markedly from one year to the next. Although some banks may marginally improve or worsen their performance over short periods of time, it is unlikely that a very efficient bank in one year would become very inefficient the next, only to return to high efficiency in the following year. Consequently, measured efficiency by acceptable approaches should yield efficiencies which are fairly stable over time, and regulatory policies targeted specifically at either very efficient or very inefficient firms should still hit their marks after normal policy and implementation lags.

We now determine the year-to-year stability of the DEA, SFA, and TFA efficiency estimates over time. The three DFA efficiency measures are excluded from this part of the analysis because they measure only "core" efficiency that persists over the entire time period, and so are perfectly stable by construction. We calculated the Spearman rank-order correlations for each of the six time-varying efficiency measures between each pair of years. That is, we computed the rank-order correlation between DEA-S efficiency in each year i, i=1977,...,1987, and DEA-S efficiency in each year j, j=1978,...,1988, with j>i to avoid redundancy, and then repeated this process for the five other techniques. These 396 correlations were positive and statistically significant in all cases. To summarize this large amount of information in the most useful way, Table 5 presents the average correlations by the number of years apart. Each figure in the One-Year-Apart first column reports for a single efficiency method, the average of the correlations of efficiencies in 1977 with 1978, 1978 with 1979, ..., 1987 with 1988, an average of 11 correlations in all. Each figure in the next column reports the average of 10 two-year-apart correlations, 1977 with 1979, 1978 with 1980, ..., 1986 with 1988. In general, the n-year-apart figures are averages of the 12 - n correlations between efficiencies that are n years away from each other. It is these averages that would seem to be most useful for regulatory analysis that must forecast the effects of their policies on firms in the future. For example, if it is thought that the policy and implementation lags are likely to take 3 years to work, then the three-year-apart average correlations may give the best indicator as to whether the policy will hit the intended target banks.

The correlation coefficients decline over time, but remain surprisingly high and statistically significant over all the available lags for all of the methods examined. After three years, the correlations are between 54.7% and 75.9%, suggesting that all the methods are stable. After eleven years, all the efficiencies

still have statistically significant correlations between 16.2% and 31.5%. This suggests that many of the "worst practice" and "best practice" banks tend to remain inefficient or efficient, respectively, over time.[25] All of the DEA, SFA, and TFA methods shown seem to indicate this stability. This also lends some support to the basic assumption of stability that underlies the DFA approach. Importantly, there is little difference in the stability of efficiency between the parametric and nonparametric methods. The only notable difference among the techniques is that the DEA methods generally show slightly more stability than the SFA and TFA methods.

Consistency Condition (v) -- Consistency of Efficiencies with Market Competitive Conditions. As shown above in Table 2 and Figures 1 and 2, the parametric methods generally yield mean efficiencies between about 80% and 90%, with the vast majority of firms having relatively high efficiency, whereas the nonparametric methods yield mean efficiencies between about 20% to 40%, with the vast majority of firms having relatively low efficiency. It seems fairly clear that the parametric approaches are generally more consistent with what are generally believed to be the competitive conditions in the banking industry. The relatively high efficiencies for the vast majority of banks seems consistent with a reasonably competitive industry in local markets that allowed entry by branch banking (recall that all the banks in our sample come from branch-banking states). Moreover, all of these firms survived branching competition over at least a 12-year period of economic turbulence in the industry, which would be difficult to achieve for firms that consumed many more inputs than the best practice banks.

In contrast, the DEA result that the vast majority of firms have measured efficiency of less than 30% does not seem to be consistent with competitive conditions in this industry. One potential explanation of this finding is that DEA does not take account of random error as the parametric approaches do. As discussed above, the dispersion from random error would likely result in lower average efficiency. If there are a few firms with very "lucky" outcomes, the firms that are compared to them may have very low measured efficiency by DEA, and this may have occurred here, since the DEA efficiencies are so much lower than those

---

[25] The stability shown in Table 5 is much longer than was reported by Eisenbeis, Ferrier, and Kwan (1997) for a sample of large multibank holding companies over the period 1986-91, where stability was statistically significant for about three and a half years. DeYoung (1997) found effectively an "optimal" stability of about 6 years for use in DFA analysis that struck a balance between the benefits and costs of the extra information from adding a marginal year of data.

generated by the parametric models and so much lower than are likely to be allowed by market forces. Note that this problem likely is not as serious as a general concern as it appears here, since most prior nonparametric studies of U.S. banks find much higher efficiencies, on average 12 percentage points lower than the parametric studies, as opposed to the average differential of about 53 percentage points here. In part, the DEA efficiency scores here may be lower than most of those found in the bank efficiency literature, because our cost-based DEA methods are based on economic efficiency, rather than technological efficiency as most of the prior DEA studies are. As noted above, technological efficiency scores will tend to be higher than economic efficiency scores because economic efficiency sets a higher standard that includes allocative efficiency. In addition, our DEA efficiency scores may be lower than those in most other DEA financial institution studies because of the steps we take to reduce the self-identifier problem (using input prices, specifying no extra constraints, and having a large number of observations relative to constraints).

Consistency Condition (vi) -- Consistency with Standard Nonfrontier Performance Measures. As indicated above, efficiency measures should be positively correlated with nonfrontier measures of performance generally used by regulators, financial institution managers, and industry consultants. Positive rank-order correlations with these measures would give assurance that the frontier measures are not simply artificial products of the assumptions made regarding the underlying optimization concept (technological efficiency versus economic efficiency), the shape of the efficient frontier, the existence of random error, and any distributional assumptions imposed on the inefficiencies and random error. As also indicated above, the correlations between accurate efficiency measures and the accounting ratios of performance are not expected be close to 1.00, since the accounting ratios embody not only the efficiencies, but also the effects of differences in input prices and other exogenous variables over managers have no little or no control.

Table 6 shows the correlations between the efficiencies generated by the nine techniques and four nonfrontier measures of performance. Both the efficiencies and the more standard ratios are averaged over time to reduce the effects of noise. The standard performance measures are the return on assets (ROA), the negative of the total operating and interest cost per dollar of assets (-TC/TA), the negative of total cost per dollar of revenue (-TC/TR), and the negative of labor employed per banking office (-Labor/Branch). The negative signs are placed on the last three ratios to simply the discussion -- all four measures are positive indicators of performance which should be positively correlated with frontier efficiency. These four

performance ratios and other similar measures are what bank managers and consultants use to generally assess their performance and rank themselves against their peers within the industry. The first three standard performance measures are indicators of economic optimization in terms of bank costs and revenues, whereas -Labor/Branch is a measure of technological optimization.[26]

The results in Table 6 suggest that the parametric-based efficiencies are generally consistent with the standard performance measures, but the DEA-based efficiencies are much less so. Looking first at the DEA columns, only four of the eight correlations between the DEA measures and the standard measures are positive and statistically significant at either the 5% or 1% significance level, while two others are negative and statistically significant at the 5% level. The positive correlations are mostly fairly low, about 10% or less, with only one correlation approaching 20%. The two negative, statistically significant correlations suggest that firms measured as efficient by DEA generally use more labor per branch, which is generally considered to be a signal of poor productivity at the branch level by bankers and consultants. Overall, the simple average of these eight rank-order correlations of .053 suggests that DEA efficiency is at best weakly related to our banking industry indicators of firm performance.

For the seven parametric techniques, all 28 correlations are positive (although some are barely so), 21 are statistically significantly different from zero at the 1% level and another is significant at the 5% level. The TFA frontier efficiency estimates are the most consistent with the standard performance ratios, with all eight correlations statistically significant, and averaging .218. The results for SFA are similar, with an average rank-order correlation with the performance measures of .205, but the correlations for -Labor/Branch ratio are only statistically significant at the 10% significance level (10% level not shown in table). For the three DFA techniques, the average correlation is similar, .199, but there is less consistency than the other parametric measures, with 4 of the 12 correlations below 10% and not statistically significant.

Overall, the parametric approaches, SFA, TFA, and DFA, are fairly consistent with the standard

---

[26]There may be difficulties with using the quantity of labor in this last ratio. It has been shown that the ratio of labor to costs has changed considerably over time (Berger and Humphrey 1992). It has also been suggested that bank holding companies have moved many of their operations into affiliates outside the bank itself, so that the labor input is consumed, but it is not counted because it is employed elsewhere in the holding company (Berger, Kashyap, and Scalise 1995). This is likely more of a problem with measures of productivity growth than efficiency, because it is a change over time, but it does introduce noise into our measure. We include it to have at least one measure of technological efficiency.

nonfrontier measures of performance and are statistically significant and positively correlated with these measures in the vast majority of cases. This gives confidence that the general method of teasing efficiencies from random error in the residuals from cost functions does not do excessive harm to the data. In contrast, there is a much weaker relationship between the DEA nonparametric methods and these same firm performance measures. Even when the correlations are statistically significant, the magnitudes are generally much smaller. This might occur in part because these methods may inadvertently count a substantial amount of the random error as differences in efficiency.

VI. Conclusions

This paper specifies a set of six consistency conditions that frontier efficiency approaches to measuring the performance of financial institutions should meet to be most useful for regulatory purposes. The first three conditions -- that the efficiencies generated by these approaches be consistent with each other in terms of their efficiency levels, rankings, and identification of best and worst firms -- help determine the degree to which the different approaches are consistent with each other. The latter three conditions -- that the efficiencies are consistent over time, consistent with competitive conditions in the market, and consistent with standard nonfrontier measures of performance -- help determine the degree to which the efficiencies generated by the different approaches are consistent with reality and are believable, which is necessary for the efficiency estimates to be useful. These consistency conditions would likely be helpful for evaluating efficiency approaches for other purposes and for firms outside of the financial institutions industry as well, and we encourage others to try applying these conditions elsewhere.

We evaluate the extent to which all four of the main approaches to estimating frontier efficiency or X-efficiency meet these consistency conditions. We employ multiple techniques within each of the four approaches, using single-period and panel methods, for a total of nine efficiency techniques evaluated. To be sure that the applications are comparable, all nine techniques use the same efficiency concept (cost efficiency), the same sample of banks, same time interval, same specifications of inputs and outputs, and (for the parametric methods) the same functional form. Our data set consists of a panel of 683 large U.S. banks (assets over $100 million) that were in operation over the entire 12-year period from 1977-88, and operated in states that allowed branch banking. This was a period of many regulatory changes and many changes in market conditions, making it an almost ideal period to determine how the different frontier approaches

identify and measure bank efficiency over a variety of extreme conditions. It is also under such extreme conditions that it is most important for regulators to be able to determine the effects of their actions on efficiency.

Our findings yield some mixed evidence regarding the consistency of the four main approaches -- nonparametric data envelopment analysis (DEA), and the parametric stochastic frontier approach (SFA), thick frontier approach (TFA), and distribution-free approach (DFA). With regard to the first three consistency conditions, these data suggest that the parametric methods are generally consistent with one another, and the nonparametric methods are generally consistent with one another, but the parametric and nonparametric methods are not generally mutually consistent. The SFA, TFA, and DFA parametric approaches tend to yield about the same distributions of efficiency (condition i), rank banks in roughly the same order (condition ii), and identify mostly the same banks as "best practice" and "worst practice" (condition iii). While there is also consistency within the nonparametric DEA methods, the parametric and nonparametric methods are not consistent with each other in these dimensions. The DEA methods yield much lower average efficiencies, rank the banks differently, and identify the best and worst banks differently from parametric methods. These results suggest that there may be "fragility" in drawing regulatory policy conclusions that may differ according to whether DEA versus the parametric approaches are specified.

Possible "tie-breakers" -- or conditions which may help choose whether the nonparametric versus parametric methods might be "better" -- are whether the efficiencies drawn from the different approaches are consistent with reality and are believable. All of the methods are found to be consistent over time (condition iv), but the parametric methods appear to be more consistent with what are generally believed to be the competitive conditions in banking markets (condition v), and also more consistent with nonfrontier measures of bank performance such as return on assets or various cost ratios that are often used by regulators, managers, and consultants (condition vi). SFA, TFA, and DFA yield relatively high efficiencies for the vast majority of firms, consistent with the state of competition in banking markets, whereas DEA yields relatively low efficiencies for most firms, perhaps reflecting the confounding of random error and inefficiency in this approach that usually does not account for random error. In addition, the parametric measures are generally highly positively correlated with the standard nonfrontier performance measures, whereas DEA measures are much less strongly related to these other indicators of firm performance.

The data also show a high degree of consistency within the parametric methods and within the nonparametric methods. This tends to suggest that regulatory policy conclusions may not be greatly affected by the choice of SFA versus TFA versus DFA, or by the choice between panel and single-year techniques (with the possible exception of single-year TFA). Rather, the only choice that appears to matter greatly for regulatory policy considerations is the choice between the parametric and nonparametric methods, at least for this data set and these techniques.

We hasten to add that these are the results of a single study, and no policy conclusions should be drawn from a single evaluation of these consistency conditions. Our results are generally supported by past research, but there is still relatively little evidence on the consistency conditions, and it is unknown how robust these results are. For example, our findings of very low efficiency for the DEA model is not very typical and may reflect our inclusion of allocative inefficiency or something else about our specification or sample, so more robustness checks are needed using alternative specifications and data sources.

As a final policy conclusion, our results suggest that when performing regulatory analysis -- or really any other analysis that depends on frontier efficiency measurement -- the use of multiple techniques and specifications is likely to be helpful. If the six consistency conditions are met for two or more approaches, then one can be more confident in the conclusions drawn.

REFERENCES

Aigner, D., Lovell C.A.K., and Schmidt P. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6: 21-37.

Ali, A.I. and Seiford L.M. (1993.) The mathematical programming approach to efficiency analysis. in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, U.K.: 120-59.

Banker, R.D., Charnes, A., Cooper W.W., Swarts J., and Thomas D.A. (1989.) An introduction to data envelopment analysis with some of its models and their uses. In *Research in Governmental and Nonprofit Accounting*, 5, (J.L. Chan and J.M. Patton eds.) JAI Press, Greenwich, CN: 125-63.

Barr, R. S., Seiford L.M., and Siems T.F. 1993. An envelopment-analysis approach to measuring the managerial efficiency of banks, *Annals of Operations Research* 45: 1-19.

Battese, G. E., and Coelli T.J. 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in Indian agriculture. *Journal of Productivity Analysis* 3: 153-169.

Bauer, P.W. Oct./Nov. 1990. Recent developments in the econometric estimation of frontiers, *Journal of Econometrics* 46 : 39-56.

Bauer, P.W., Berger A.N., and Humphrey D.B. 1993. Efficiency and productivity growth in U.S. banking, pp. 386-413, in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The measurement of productive efficiency: Techniques and applications*. New York: Oxford University Press.

Bauer, P.W., and Hancock D., April 1993. The efficiency of the Federal Reserve in providing check processing services, *Journal of Banking and Finance* 17: 287-311.

Berger, Allen N., 1993. 'Distribution-free' estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions, *Journal of Productivity Analysis* 4, 61-92.

Berger, A.N., Cummins, D. and Weiss, M. Oct. 1997. The coexistence of multiple distribution systems for financial services: The case of property-liability insurance, *Journal of Business*, 70.

Berger, A.N. and DeYoung, R. June 1997. Problem loans and cost efficiency in commercial banks, *Journal of Banking and Finance* 21: 849-70.

Berger, A.N., and T.H. Hannan, forthcoming, The efficiency cost of market power in the banking industry: A test of the `quiet life' and related hypotheses, *Review of Economics and Statistics*.

Berger, A.N. and Humphrey D.B. 1991. The Dominance of inefficiencies over scale and product mix economies in banking, *Journal of Monetary Economics* 28, 117-148.

Berger, A.N. and Humphrey D.B. 1992. Measurement and efficiency issues in commercial banking, pp. 245-79, in Zvi Griliches, ed., *Output measurement in the service sectors*, National Bureau of Economic Research Studies in Income and Wealth, vol. 56. Chicago: University of Chicago Press.

Berger, A.N. and Humphrey, D.B. April 1997. Efficiency of financial institutions: International survey and directions for future research, *European Journal of Operational Research*, 98, 175-212.

Berger, A.N., Kashyap, A.K and Scalise, J.M. 1995. The transformation of the U.S. banking industry: What a long, strange trip it's been, *Brookings Papers on Economic Activity*, 55-201.

Berger, A.N., Leusner , J. and Mingo, J. Sept. 1997. The Efficiency of Bank Branches, *Journal of Monetary Economics* 40: 141-162.

Berger, A.N., and Mester, L.J. July 1997. Beyond the black box: What explains differences in the efficiencies of financial institutions?, *Journal of Banking and Finance* 21: 895-947.

Cebenoyan, A.S., Cooperman, L.J. and Register, C.A. 1993. Firm inefficiency and the regulatory closure of S&Ls: An empirical investigation, *Review of Economics and Statistics*, 75:540-5.

Charnes, A., Cooper, C.A., Lewin, A., and Seiford, L. 1994. *Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Publishers, Boston, U.S.A.

Charnes, A., Cooper, W.W. and Rhodes, E., 1978. Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2: 429-44.

Charnes, A., Cooper, W.W., Sueyoshi, T. 1988. A goal programming/constrained regression review of the Bell System breakup, *Management Science* 34, 1-26.

Cornwell, C., Schmidt, P. and Sickles, R.C. 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels, *Journal of Econometrics* 46, 185-200.

Cummins, D., and Zi, H. 1998. Comparison of frontier efficiency methods: An application to the U.S. Life Insurance Industry, *Journal of Productivity Analysis*, 10.

D'Agostino, R.B., Balanger, A., and D'Agostino, R.B. Jr. 1990. A suggestion for using powerful and informative tests of normality, *The American Statistician*, 44, no. 4, 316-321.

DeBorger, B., Ferrier, G., and Kerstens, K., forthcoming, The choice of a technical efficiency measure on the free disposal hull reference technology: A Comparison using U.S. Banking Data, *European Journal of Operational Research*.

DeYoung, R., 1997. A Diagnostic test for the distribution-free efficiency estimator: An example using U.S. Commercial Bank Data, *European Journal of Operational Research* 98, 243-249.

DeYoung, R., Feb.1998. Management Quality and X-Efficiency in National Banks, *Journal of Financial Services Research*, 13

DeYoung, R., Hasan, I., and Kirchhoff, B. 1998. The impact of out-of-state entry on the efficiency of local banks, *Journal of Economics and Business*.

Eisenbeis, R.A., Ferrier, G.D. and Kwan S. 1997. The Informativeness of Linear Programming and Econometric Efficiency Scores: An Analysis Using US Banking Data, Bureau of Business and Economic Research. University of Arkansas working paper.

Färe, R., Grosskopf, S. and Lovell, C.A.K. 1994, *Production Frontiers*. Cambridge: Cambridge University Press.

Färe, R., and Whittaker, G. 1996. Dynamic Measurement of Efficiency: An Application to Western Public Grazing, pp. 168-185, in *Intertemporal Production Frontiers: With Dynamic DEA*, by Rolf Färe and Shawna Grosskopf . Boston: Kluwer Academic Publishers.

Fecher, F., Kessler, D., Perelman, S. and Pestieau, P. Productive performance of the French insurance industry, *Journal of Productivity Analysis*, 4 1993: 77-93.

Ferrier, G., Grosskopf, S., Hayes, K., and Yaisawarng, S. 1993. Economies of diversification in the banking industry: A frontier approach, *Journal of Monetary Economics*, 31: 229-49.

Ferrier, G.D., and Hirschberg, J.D. March 1997. Bootstrapping Confidence Intervals for Linear Programming Efficiency Scores: With an Illustration Using Italian Banking Data, *Journal of Productivity Analysis*, 8, 19-33.

Ferrier, G., Kerstens, K., Vanden Eeckaut, P. 1994. Radial and Nonradial Technical Efficiency Measures on a DEA Reference Technology: A Comparison Using Banking Data, *Recherches Économiques de Louvain*, 60:449-79.

Ferrier, G. D. and Lovell, C.A.K. 1990. Measuring cost efficiency in banking: Econometric and linear programming evidence, *Journal of Econometrics* 46, 229-245.

Ferrier, G. D. and Philpot, J. 1994. The Relative Efficiency of Mutual Funds: A Frontier Analysis of Their Performance, University of Arkansas working paper.

Giokas, D., 1991. Bank Branch Operating Efficiency: A Comparative Application of DEA and the Loglinear Model, *OMEGA International Journal of Management Science*, 19:549-57.

Greene, W.H., 1993. The Econometric Approach to Efficiency Analysis, in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, U.K. 68-119.

Grosskopf, S., 1993. Efficiency and Productivity, in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, U.K. 160-94.

Grosskopf, S., July 1996. Statistical Inference and Nonparametric Efficiency: A Selective Survey, *Journal of Productivity Analysis*, 7: 61-76.

Hasan, I., and Hunter, W.C. 1996. Efficiency of Japanese Multinational Banks in the U.S., in A.H. Chen, ed., *Research in Finance, Volume 14*, JAI Press, Greenwich, CT 157-73.

Hermalin, B.E., and Wallace N.E. Autumn 1994. The determinants of efficiency and solvency in savings and loans, *Rand Journal of Economics*, 25: 361-81.

Kumbhakar, S. C., Oct./Nov. 1990. Production frontiers, panel data, and time-varying technical inefficiency, *Journal of Econometrics*, 46: 201-212.

Land, K., Lovell, C.A.K., and Thore, S. Nov/Dec 1993. Chance-Constrained Data Envelopment Analysis, *Managerial and Decision Economics*, 14: 541-54.

Leamer, E. E., 1994. *Sturdy econometrics*. Aldershot, U.K.: Edward Elgar.

Leamer, E.E. and Leonard, H.B. 1983. Reporting the Fragility of Regression Estimates, *Review of Economics and Statistics* 65, 306-317.

Lovell, C.A.K., 1993. Production Frontiers and Productive Efficiency, in H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, U.K. 3-67

Lovell, C.A.K., and Pastor, J. 1997. Target setting: An application to a bank branch network, *European Journal of Operational Research*, 98, 290-299.

Maudos, J., June 1996. A Comparison of Different Stochastic Frontier Techniques with Panel Data: An Application for Efficiency of Spanish Banks, working paper, University of Valencia, Valencia, Spain.

Mester, L. J., 1996. A study of bank efficiency taking into account risk-preferences, *Journal of Banking and Finance* 20, 1025-1045.

Oral, M., and Yolalan, R., 1990. An Empirical Study on Measuring Operating Efficiency and Profitability of Bank Branches, *European Journal of Operational Research* 46: 282-94.

Parkan, C. 1987. Measuring the Efficiency of Service Operations: An Application to Bank Branches, *Engineering Costs and Production Economics* 12: 237-42.

Pastor, J.T., July 1993. Efficiency of Bank Branches Through DEA: The Attracting of Liabilities, Working Paper, Universidad de Alicante, Alicante, Spain.

Resti, A. Feb. 1997. Evaluating the cost-efficiency of the Italian banking system: What can be learned from the joint application of parametric and nonparametric techniques, *Journal of Banking and Finance* 20: 221-50.

Schmidt, P. and Sickles, R.C. 1984., Production Frontiers and Panel Data, *Journal of Business and Economic Statistics* 2, 367-374.

Seiford, L.M., and Thrall, R.M. Oct./Nov. 1990. Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis, *Journal of Econometrics* 46: 7-38.

Sherman, H.D., and F. Gold, Bank Branch Operating Efficiency: Evaluation with Data Envelopment Analysis, *Journal of Banking and Finance*, 9 1985: 279-315.

Simar, L., and Wilson, P.W. 1995. Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models, working paper, Institute of Statistics, Universite Catholique de Louvain, Belgium.

Spong, K., Sullivan R., and DeYoung, R. December 1995. What makes a bank efficient? A look at financial characteristics and bank management and ownership structure, Financial Industry Perspectives, Federal Reserve Bank of Kansas City, pp. 1-19.

Vassiglou, M., and Giokas, D. 1990. A study of the relative efficiency of bank branches: An Application of Data Envelopment Analysis, *Journal of the Operational Research Society* 41:91-97.

Yuengert, A., 1993. The Measurement of Efficiency in Life Insurance: Estimates of a Mixed Normal-Gamma Error Model, *Journal of Banking and Finance*, 17: 483-96.

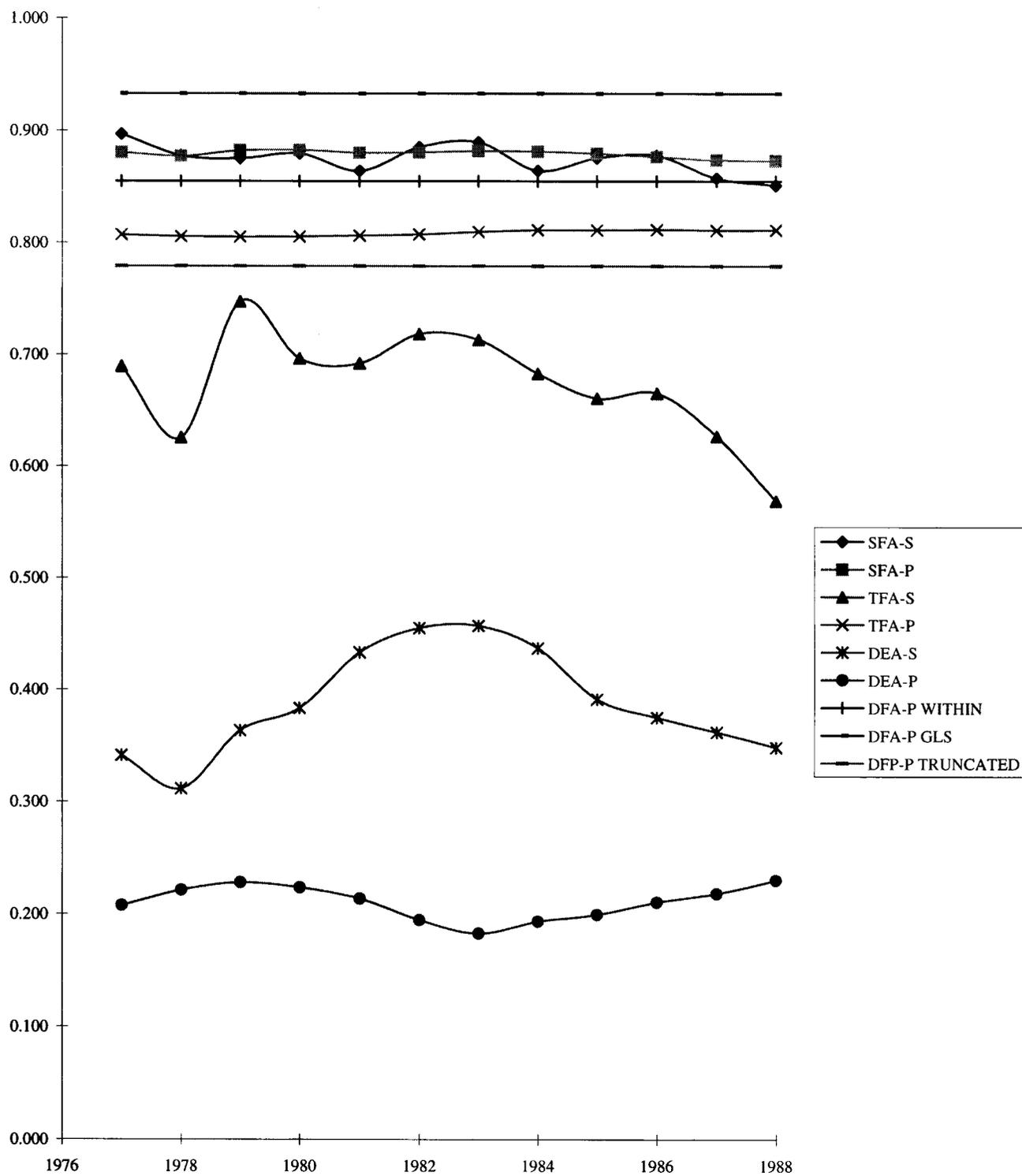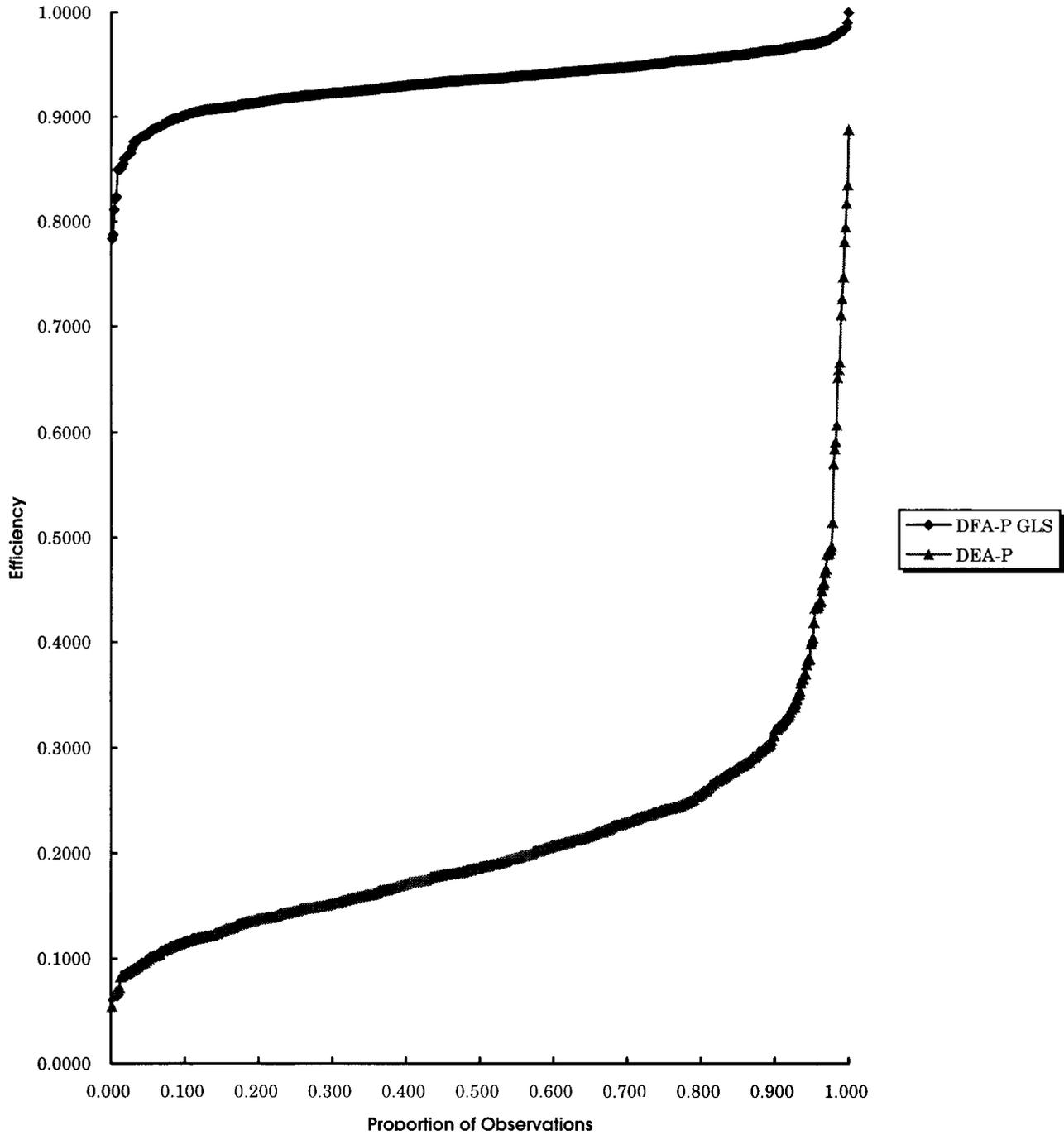**Figure 1:** Average Efficiency over Time for the Nine Techniques

Figure 2: The Cumulative Distributions of the DFA-P GLS and DEA-P Efficiency Scores

Table 1: Descriptive Statistics of the Aggregate Data (683 banks over 12 years: 1977-88)[1]

	Mean	Standard Deviation	Minimum Value	Maximum Value
Total Cost	214155.57	944194.56	8277.12	20511576.66
Output Quantities:				
Demand Deposits	6785700.65	3400066.25	1090526.00	12531125.00
Real Estate Loans	323196.21	1215954.37	0.00	30159182.64
Installment Loans	214890.11	637394.20	0.00	14155304.67
Commercial Loans	755611.30	3780350.97	1306.88	66260201.77
Input Quantities:				
Labor[2]	1216.45	3971.48	12.00	79399.71
Physical Capital	32088.73	112207.91	495.32	66260201.73
Time Deposits	611615.98	1602807.47	605.65	35647871.59
Purchased Funds	1175298.34	6689270.60	2727.43	115298070.00
Input Prices:				
Labor	24.07	5.30	8.47	121.75
Physical Capital	76.71	9.96	43.36	121.85
Time Deposits[3]	0.05	0.02	0.00	0.56
Purchased Funds[3]	0.08	0.04	0.00	1.73

[1] All financial data are annual real values in 1000's of 1988 dollars, unless otherwise indicated.

[2] Number of full-time equivalent employees.

[3] Prices of financial inputs are interest rates.

Table 2: Descriptive Statistics of the Efficiency Scores by Technique

	DEA-S	DEA-P	SFA-S	SFA-P	TFA-S	TFA-P	DFA-P WITHIN	DFA-P GLS	DFA-P TRUNCATED
Mean	0.385	0.210	0.875	0.879	0.674	0.808	0.855	0.933	0.779
Median	0.348	0.187	0.883	0.889	0.673	0.890	0.861	0.936	0.780
Mode	0.276	0.141	0.900	0.907	0.671	0.813	0.872	0.942	0.780
Minimum	0.103	0.054	0.471	0.479	0.309	0.622	0.546	0.783	0.548
Maximum	1.000	0.888	1.000	1.000	0.955	1.000	1.000	1.000	1.000
Standard Deviation	0.170	0.109	0.074	0.073	0.082	0.022	0.064	0.027	0.0835
Skewness	1.283	2.649	-1.043	-1.168	-0.274	-0.601	-0.779	-1.195	-0.116
Kurtosis	1.942	10.238	2.685	2.972	1.389	17.363	1.485	3.462	0.271
Num. Banks	683	683	683	683	683	683	683	683	683

Notes: The efficiencies are calculated using 12 years of data for 683 banks (8,196 observations), and the numbers in the table are based on the average efficiency for each bank over the entire sample period.

DEA-S ≡ Data Envelopment Analysis (DEA) using a Single (S) year of data as a reference set.

DEA-P ≡ Data Envelopment Analysis (DEA) using the entire 12-year Panel (P) of data as a reference set.

SFA-S ≡ Stochastic Frontier Approach (SFA) allowing the cost function parameters to vary for each Single (S) year of data.

SFA-P ≡ Stochastic Frontier Approach (SFA) forcing the cost function parameters (other than the intercepts) to be constant over the 12-year Panel (P) of data.

TFA-S ≡ Thick Frontier Approach (SFA) allowing the cost function parameters to vary for each Single (S) year of data.

TFA-P ≡ Thick Frontier Approach (SFA) forcing the cost function parameters (other than the intercepts) to be constant over the 12-year Panel (P) of data.

DFA-P WITHIN ≡ Distribution-Free Approach (DFA) forcing the cost function parameters (other than the intercepts) to be constant over the 12-year Panel (P) of data, and estimating fixed-effect dummies for individual banks (WITHIN) to measure inefficiencies.

DFA-P GLS ≡ Distribution-Free Approach (DFA) forcing the cost function parameters (other than the intercepts) to be constant over the 12-year Panel (P) of data, controlling for serial correlation and estimating average residuals for individual banks (GLS) to measure inefficiencies.

DFA-P TRUNCATED ≡ Distribution-Free Approach (DFA) allowing the cost function parameters to vary for each year of data, but forcing the inefficiencies to be constant for an individual bank over the 12-year Panel (P) of data, estimated using the average residuals, and truncating the extreme values by assigning slightly less extreme values (TRUNCATED).

Table 3: Spearman Rank-Order Correlations Among the Efficiency Scores Created by Various Techniques

	DEA-S	DEA-P	SFA-S	SFA-P	TFA-S	TFA-P	DFA-P WITHIN	DFA-P GLS	DFA-P TRUNCATED
DEA-S	1.000	0.895**	0.150**	0.104**	0.104**	0.083*	-0.135**	-0.092*	0.154**
DEA-P		1.000	0.167**	0.140**	0.214**	0.174**	0.054	0.061	0.188**
SFA-S			1.000	0.976**	0.837**	0.876**	0.463**	0.550**	0.981**
SFA-P				1.000	0.849**	0.897**	0.495**	0.566**	0.964**
TFA-S					1.000	0.945**	0.651**	0.684**	0.889**
TFA-P						1.000	0.645**	0.701**	0.918**
DFA-P WITHIN							1.000	0.936**	0.484**
DFA-P GLS								1.000	0.567**
DFA-P TRUNCATED									1.000

Notes: To reduce the effects of noise, the rank-order correlations of the **average** efficiencies of the individual banks over time are reported, although the correlations using the efficiency estimates from the separate years are quite similar.

* Correlation is statistically significantly different from zero at the 5% level, two-sided.
** Correlation is statistically significantly different from zero at the 1% level, two-sided.

Table 4: Correspondence of "Best Practice" and "Worst Practice" Banks across Techniques
(Upper triangle shows "best practice" and lower triangle shows "worst practice")

	DEA-S	DEA-P	SFA-S	SFA-P	TFA-S	TFA-P	DFA-P WITHIN	DFA-P GLS	DFA-P TRUNCATED
DEA-S		0.760**	0.357	0.339	0.310	0.304	0.187	0.211	0.345
DEA-P	0.749**		0.351	0.345	0.363	0.333	0.263	0.281	0.368 ---
SFA-S	0.345	0.368		0.912**	0.743**	0.790**	0.491**	0.526**	0.930**
SFA-P	0.298	0.316	0.877**		0.743**	0.795**	0.509**	0.521**	0.906**
TFA-S	0.304	0.374	0.725**	0.719**		0.854**	0.567**	0.602**	0.778**
TFA-A	0.322	0.368	0.778**	0.790**	0.854**		0.591**	0.632**	0.813**
DFA-P WITHIN	0.240	0.374	0.491**	0.521**	0.661**	0.597**		0.825**	0.503**
DFA-P GLS	0.257	0.345	0.532**	0.544**	0.673**	0.620**	0.778**		0.526**
DFA-P TRUNCATED	0.322	0.357	0.895**	0.854**	0.760**	0.819**	0.509**	0.544**	

Notes: Each number in the upper triangle are the proportion of banks that are identified by one technique as having efficiency scores in the **most** efficient 25% of banks that are also identified in the **most** efficient 25% by the other technique.

Each number in the lower triangle is the proportion of banks that are identified by one technique as having efficiency scores in the **least** efficient 25% of banks that are also identified in the **least** efficient 25% by the other technique.

* Correspondence is statistically significantly different from 0.250 at the 5% level by χ^2 test.
** Correspondence is statistically significantly different from 0.250 at the 1% level by χ^2 test.

Table 5: The Persistence of Efficiency -- The Correlations of n-Year-Apart Efficiencies

	One Year Apart	Two Years Apart	Three Years Apart	Four Years Apart	Five Years Apart	Six Years Apart	Seven Years Apart	Eight Years Apart	Nine Years Apart	Ten Years Apart	Eleven Years Apart
DEA-S	0.920	0.815	0.723	0.640	0.575	0.520	0.484	0.436	0.377	0.354	0.313
DEA-P	0.944	0.850	0.759	0.679	0.614	0.561	0.514	0.468	0.419	0.367	0.315
SFA-S	0.855	0.703	0.581	0.472	0.400	0.347	0.299	0.276	0.260	0.252	0.231
SFA-P	0.864	0.710	0.583	0.466	0.378	0.312	0.258	0.223	0.194	0.175	0.162
TFA-S	0.829	0.664	0.547	0.439	0.372	0.323	0.268	0.202	0.242	0.228	0.200
TFA-P	0.875	0.740	0.628	0.520	0.450	0.386	0.328	0.293	0.258	0.251	0.252

Notes: Each entry in the table is the average of the correlations of the n-year apart efficiencies for a single efficiency technique within our 12-year time span, so for each n, the number reported is the average of 12 - n correlations. For example, there are 9 different correlations for the 3-year apart correlations -- 1977 with 1980, 1978 with 1981, ..., 1985 with 1988. All 396 of the individual correlations that used in the averages are statistically significantly different from zero at the 1% level, two-sided.

The three DFA efficiency measures are excluded from this table because they measure only "core" efficiency that persists over the entire time period, and so are perfectly persistent by construction.

Table 6: Efficiency Correlations with Standard Nonfrontier Performance Measures

	DEA-S	DEA-P	SFA-S	SFA-P	TFA-S	TFA-P	DFA-P WITHIN	DFA-P GLS	DFA-P TRUNCATED
ROA	0.109*	0.026	0.235**	0.229**	0.125**	0.169**	0.007	0.019	0.232**
-TC/TA	0.102*	0.196**	0.220**	0.224**	0.339**	0.342**	0.290**	0.285**	0.264** ---
-TC/TR	0.115**	0.067	0.285**	0.282**	0.215**	0.258**	0.095*	0.072	0.291**
-Labor/ Branch	-0.100*	-0.092*	0.078	0.083	0.160**	0.137**	0.357**	0.411**	0.068

Notes: To reduce the effects of noise, the rank-order correlations of the **average** efficiencies and **average** standard nonfrontier performance measures of the individual banks over time are reported, although the correlations using the efficiency estimates from the separate years are quite similar.

* Correlation is statistically significantly different from zero at the 5% level, two-sided.
** Correlation is statistically significantly different from zero at the 1% level, two-sided.

ROA   ≡ return on assets
-TC/TA   ≡ negative of total costs/total assets
-TC/TR   ≡ negative of total costs/total revenues
-Labor/Branch   ≡ negative of number of employees per banking office