Robust Nonparametric Estimation of Efficiency and Technical Change in U.S. Commercial Banking

David C. Wheelock

and

Paul W. Wilson*

December 2003

JEL classification nos.: C14, C43, G21
Keywords: order-m, technical efficiency, productivity, frontier, Malmquist

Abstract

This paper examines the performance of the U.S. commercial banking industry over 1984–2002. Rather than measuring performance relative to the unknown (and difficult-to-estimate) boundary of the production set, performance for a given bank is measured relative to *expected* maximum output among m banks using no more of each input than the given bank. This approach permits fully non-parametric estimation with $\sqrt{n}$-consistency, avoiding the usual curse of dimensionality that plagues traditional non-parametric efficiency estimators. The resulting estimates are robust with respect to outliers and noise in the data.

*Wheelock: Research Department, Federal Reserve Bank, P.O. Box 442, St. Louis, MO 63166 (email: wheelock@stls.frb.org). Wilson: Department of Economics, University of Texas, Austin, Texas 78712 USA (email: wilson@eco.utexas.edu). This research was conducted while Wilson was a visiting scholar at the Federal Reserve Bank of St. Louis. We have benefited from discussions with Léopold Simar and comments on a previous version of the paper from Alton Gilbert and Frank Schmid. The usual caveats apply, *i.e.,* any errors, omissions, *etc.* are solely our own.

1. Introduction

The U.S. banking industry has witnessed dramatic changes in regulation and market structure during the past two decades, with the number of commercial banks declining from a peak of 14,496 in 1984 to 7,887 at the end of 2002. Although the number of banks has declined, many banking markets have become more competitive with the elimination of branch banking regulations and other entry barriers, as well as a blurring of lines separating commercial banks from other financial service firms. Banks have also invested heavily in new data processing and telecommunications technologies with the expectation that such investment would lead to improved productivity and higher profits.

Thus far, evidence for productivity and efficiency improvement in banking attendant with increased competition and capital investment has been mixed. Many studies, using data from the 1980s and early 1990s, found that banks tend to suffer from considerable managerial, or "x-", inefficiency (see Berger and Humphrey 1997, for a survey). Some studies find that technological progress raised average bank productivity in these years, but relative to production possibilities, banks seem not to have become more efficient (see Wheelock and Wilson, 1999; Alam, 2001). Using more recent data (1991-97), Berger and Mester (2003) find that technological improvements mainly increased banks' profit productivity, as early adopters of technology earned higher profits, at least temporarily. Over the same period, however, average cost productivity declined.

This paper examines further the evolution of productivity, efficiency and technical progress in commercial banking by applying alternative, new concepts of efficiency, as well as new nonparametric estimators. The banking industry has continued to consolidate since the mid-1990s while deregulation, such as the removal of barriers to interstate branching in 1997, and heavy investment in new computer and telecommunications capital has continued. Using data for 1984-2002, we follow Alam (2001) and Wheelock and Wilson (1999) in decomposing a Malmquist index of total factor productivity into changes in efficiency and technology, though we use the more general model of bank production of

Berger and Mester (2003). Unlike previous studies, however, our analysis is based on the order-m frontier described by Cazals *et al.* (2002), which offers several advantages over previously used methods of efficiency estimation.

Prior studies often relied on methods that imposed strong assumptions on production and cost relationships, as well as on the distribution of efficiency scores. Many studies estimated translog cost or profit functions that include a two-sided random noise term and a one-sided random inefficiency term. The translog function, however, has been shown to mis-specify bank cost relationships (see, *e.g.,* McAllister and McManus, 1993; Wheelock and Wilson, 2001), and the commonly used "correction" of augmenting the translog function with trigonometric terms has several other drawbacks as implemented in the banking literature.[1] Other studies, *e.g.,* Wheelock and Wilson (1999), use nonparametric envelopment estimators of the efficient frontier, such as the data envelopment analysis (DEA) or free disposal hull (FDH) estimators. Unlike parametric estimators, DEA and FDH do not impose a particular functional form on the relationship between production inputs and outputs. DEA and FDH do have drawbacks, however. For example, both are highly sensitive to extreme values and noise in the data. The only difference between DEA and FDH is that the DEA estimator assumes that the efficient frontier is convex. In the present study of U.S. banks, we find that efficiency estimates are extremely sensitive to this assumption. Using FDH, we find that *all* banks lie on the estimated frontier, implying that all banks are efficient. By contrast, when we use DEA, less than 3 percent of banks lie on the estimated frontier.

[1]Applications in banking have typically added an arbitrary number of sine and cosine terms to the traditional translog cost function to achieve greater flexibility in fit. Addition of trigonometric terms to translog cost or profit functions represent attempts at semi-non-parametric series estimation (see Efromovich, 1999, for discussion). None of the banking studies implementing series estimation have, to our knowledge, attempted to optimize the number of included terms by cross-validation or other data-based methods, and so it remains unknown whether these models under- or over-fit the data. The large number of terms (Gallant, 1981, 1982, proposed $n^{2/3}$ as a rule-of-thumb for the number of terms to include) typically required for series estimators in the regression context make it impractical to use maximum likelihood to estimate composite-error models, where a one-sided inefficiency process is convolved with a two-sided noise process. Consequently, a number of recent studies using this approach have included only a small number of trigonometric terms, but this likely results in under-fitting the data.

The order-m estimator proposed by Cazals *et al.* (2002) requires no convexity assumptions, and has several desirable properties that make it useful for drawing inferences about efficiency. As with DEA and FDH estimators, order-m estimators are fully non-parametric, but unlike DEA and FDH estimators, order-m estimators are root-n consistent and do not suffer from the well-known curse of dimensionality. In addition, order-m estimators are robust with respect to extreme values and noise, in stark contrast to DEA and FDH estimators which are especially sensitive. We use the order-m estimator to estimate a modified measure of output technical efficiency, as well as modified measures of changes in productivity, efficiency, and technology over time. We further decompose efficiency changes into pure technical and scale inefficiency, and decompose technical change into pure technical change and changes in the scale of the technology.

Section 2 describes our statistical model and the FDH, DEA, and order-m estimators, which are used to define modified measures of productivity and other changes in Section 3. We specify the inputs and outputs of bank production and describe our data in Section 4. Empirical results are given in Section 5, with conclusions in Section 6.

2. Technology, Distance Functions, and Estimators

2.1. Statistical Model:

We begin by defining notation and summarizing the traditional nonparametric estimators of efficiency and their properties. Denote the production possibilities set at time t by

$$\mathcal{P}^t = \{(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{x} \text{ can produce } \boldsymbol{y} \text{ at time } t\}, \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{R}_+^p$ and $\boldsymbol{y} \in \mathbb{R}_+^q$ denote vectors of inputs and outputs, respectively. The production possibilities set can be described in terms of its sections

$$\mathcal{X}^t(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathbb{R}_+^p \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t\}, \tag{2.2}$$

and

$$\mathcal{Y}^t(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}_+^q \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t\}, \tag{2.3}$$

or input requirement sets and output correspondence sets, respectively. Typical economic assumptions (*e.g.,* Shephard, 1970; Färe, 1988) include: (i) $\mathcal{P}^t$ is convex, $\mathcal{X}^t(\boldsymbol{y})$ is convex and closed for all $y \in \mathbb{R}_+^q$, and $\mathcal{Y}^t(\boldsymbol{x})$ is convex, bounded, and closed for all $\boldsymbol{x} \in \mathbb{R}_+^p$; (ii) all production requires the use of some inputs, *i.e.,* $(\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}^t$ if $\boldsymbol{y} \geq 0,\ \boldsymbol{x} = 0$; and (iii) both inputs and outputs are strongly disposable, *i.e.,* if $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t$ then $\widetilde{\boldsymbol{x}} \geq \boldsymbol{x} \Rightarrow (\widetilde{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{P}^t$ and $\widetilde{\boldsymbol{y}} \leq \boldsymbol{y} \Rightarrow (\boldsymbol{x}, \widetilde{\boldsymbol{y}}) \in \mathcal{P}^t$.

The upper boundary of $\mathcal{P}^t$, denoted $\mathcal{P}^{\partial t}$, is sometimes referred to as the *technology* or the *production frontier,* and is given by the intersection of $\mathcal{P}^t$ and the closure of its compliment. Assumption (iii) above is equivalent to an assumption of monotonicity for $\mathcal{P}^{\partial t}$. Similarly, the closure of the compliment of $\mathcal{X}^t(\boldsymbol{y})$—denoted $\mathcal{X}^{\partial t}(\boldsymbol{y})$—represents an isoquant. The closure of the compliment of $\mathcal{Y}^t(\boldsymbol{x})$, denoted $\mathcal{Y}^{\partial t}(\boldsymbol{x})$, gives an iso-output or product transformation curve.

The Shephard (1970) output distance function measures distance from an arbitrary point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$ to $\mathcal{P}^{\partial t}$ in a direction orthogonal to $\boldsymbol{x}$, and is defined by

$$\begin{aligned} D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t) &\equiv \inf \left\{\theta > 0 \mid (\boldsymbol{x}, \boldsymbol{y}/\theta) \in \mathcal{P}^t\right\} \\ &= \inf \left\{\theta > 0 \mid \boldsymbol{y}/\theta \in \mathcal{Y}^t(\boldsymbol{x})\right\}. \end{aligned} \tag{2.4}$$

For $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t$, we have $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t) \leq 1$ by definition.

Although the distance function in (2.4) is defined in terms of the production set $\mathcal{P}^t$, different distance functions can be defined by replacing $\mathcal{P}^t$ (2.4) with some other set to measure distance from $(\boldsymbol{x}, \boldsymbol{y})$ to the boundary of the other set. Let $\mathcal{V}(\mathcal{A})$ denote the convex cone (with vertex at the origin) spanned by the set $\mathcal{A} \subset \mathbb{R}_+^{p+q}$. Clearly, $\mathcal{P}^t \subseteq \mathcal{V}(\mathcal{P}^t)$. If $\mathcal{P}^{\partial t}$ exhibits constant returns to scale (CRS) everywhere, then the technology $\mathcal{P}^{\partial t}$ implies a mapping $\boldsymbol{x} \rightarrow \boldsymbol{y}$ that is homogeneous of degree 1; *i.e.,* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^{\partial t}$ implies $(\lambda \boldsymbol{x}, \lambda \boldsymbol{y}) \in \mathcal{P}^{\partial t}$ for all $\lambda > 0$. In this case, $\mathcal{P}^t = \mathcal{V}(\mathcal{P}^t)$ and $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t) = D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}^t))$; otherwise, $\mathcal{P}^t \subset \mathcal{V}(\mathcal{P}^t)$ and $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t) \geq D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}^t))$.

Of course, the production set $\mathcal{P}^t$ and hence the distance function defined by (2.4) are unobserved and must be estimated from data. Before anything can be estimated, a statistical model must be defined. To ensure consistent estimation using the DEA estimator described below, assumptions (i)–(iii) listed above and the following assumptions of Kneip *et al.* (1998) are required: (iv) the sample observations at time t, $\mathcal{S}^t_{n_t} = \{(\boldsymbol{x}^t_i, \boldsymbol{y}^t_i\}^{n_t}_{i=1}\}$, are realizations of identically, independently distributed (iid) random variables with probability density function $f^t(\boldsymbol{x}, \boldsymbol{y})$ with support over $\mathcal{P}^t$; (v) the density $f^t(\boldsymbol{x}, \boldsymbol{y})$ is continuous except along the frontier, with $f^t(\boldsymbol{x}, \boldsymbol{y}) = 0 \; \forall \; (\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}^t$ and $f^t(\boldsymbol{x}, \boldsymbol{y}) > 0 \; \forall \; (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^{\partial t}$; and (vi) for all $(\boldsymbol{x}, \boldsymbol{y})$ in the interior of $\mathcal{P}^t$, $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t)$ is differentiable in both its arguments. Together, assumptions (i)–(vi) define a statistical model.[2]

2.2. Traditional Estimators:

Several estimators of $\mathcal{P}^t$ and $\mathcal{V}(\mathcal{P}^t)$, and hence the distance functions $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t)$ and $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}^t))$, are possible. Two common estimators of $\mathcal{P}^t$ are the free-disposal hull of $\mathcal{S}^t_{n_t}$, suggested by Deprins *et al.* (1984) and defined by

$$\widehat{\mathcal{P}}^t_{\mathrm{FDH}} = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{p+q}_+ \mid \boldsymbol{y} \leq \boldsymbol{y}^t_i, \; \boldsymbol{x} \geq \boldsymbol{x}^t_i \; \forall \; (\boldsymbol{x}^t_i, \boldsymbol{y}^t_i) \in \mathcal{S}^t_{n_t}\}, \tag{2.5}$$

and the convex hull of $\widehat{\mathcal{P}}^t_{\mathrm{FDH}}$ given by

$$\begin{aligned} \widehat{\mathcal{P}}^t_{\mathrm{DEA}} = \Big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{p+q}_+ \mid \boldsymbol{y} \leq \sum_{i=1}^{n} \gamma_i \boldsymbol{y}^t_i, \; \boldsymbol{x} \geq \sum_{i=1}^{n} \gamma_i \boldsymbol{x}^t_i, \\ \sum_{i=1}^{n} \gamma_i = 1, \; \gamma_i \geq 0 \; \forall \; i = 1, \; \ldots, \; n \Big\}. \end{aligned} \tag{2.6}$$

The convex cone $\mathcal{V}(\widehat{\mathcal{P}}^t_{\mathrm{DEA}})$ spanned by $\widehat{\mathcal{P}}^t_{\mathrm{DEA}}$ (or, equivalently, by $\widehat{\mathcal{P}}^t_{\mathrm{FDH}}$ or $\mathcal{S}^t_{n_t}$), is obtained by dropping the constraint $\sum_{i=1}^{n} \gamma_i = 1$ in (2.6) and provides an estimator of $\mathcal{V}(\mathcal{P}^t)$. The asymptotic properties of $\widehat{\mathcal{P}}^t_{\mathrm{FDH}}$ and $\widehat{\mathcal{P}}^t_{\mathrm{DEA}}$ have been examined by Korostelev *et al.* (1995a, 1995b); see Simar and Wilson (2000) for a summary.

Estimators of $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t)$ are obtained by replacing $\mathcal{P}^t$ with either $\widehat{\mathcal{P}}^t_{\mathrm{FDH}}$ or $\widehat{\mathcal{P}}^t_{\mathrm{DEA}}$ in (2.4). Similarly, $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\widehat{\mathcal{P}}^t_{\mathrm{DEA}}))$, which is equivalent to $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\widehat{\mathcal{P}}^t_{\mathrm{FDH}}))$, yields

[2]Our assumption (vi) is stronger, but simpler, than the one used by Kneip *et al.* (1998); both are assumptions about the smoothness of the frontier $\mathcal{P}^{\partial t}$.

an estimator of $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}^t))$. The resulting estimators $D(\boldsymbol{x}, \boldsymbol{y} \mid \widehat{\mathcal{P}}^t_{\text{DEA}})$ and $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\widehat{\mathcal{P}}^t_{\text{DEA}})$ are easily computed by linear programming methods, while $D(\boldsymbol{x}, \boldsymbol{y} \mid \widehat{\mathcal{P}}^t_{\text{FDH}})$ can be computed by simple numerical algorithms. The estimators based on $\widehat{\mathcal{P}}^t_{\text{DEA}}$ and $\mathcal{V}(\widehat{\mathcal{P}}^t_{\text{DEA}})$ are commonly referred to as DEA estimators. The estimators based on the convex hull permit varying returns to scale, while those based on the convex cone incorporate a restriction of constant returns to scale.

The asymptotic properties of the DEA and FDH distance function estimators are discussed in Gijbels (1999), Park *et al.* (2000), Simar and Wilson (2000), and Kneip *et al.* (2003). In particular, $D(\boldsymbol{x}, \boldsymbol{y} \mid \widehat{\mathcal{P}}^t_{\text{DEA}}) = D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t) + O_p\left(n^{-2/(p+q+1)}\right)$ and $D(\boldsymbol{x}, \boldsymbol{y} \mid \widehat{\mathcal{P}}^t_{\text{FDH}}) = D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t) + O_p\left(n^{-1/(p+q)}\right)$. The convergence rates are slow, reflecting the curse of dimensionality which is common with nonparametric estimators. The rate of convergence for the FDH estimator is slower than for the DEA estimator, but if $\mathcal{P}^t$ is non-convex, the DEA estimator is inconsistent. In addition to slow convergence rates and the curse of dimensionality, the DEA and FDH estimators also suffer from extreme sensitivity to outliers. For many applications, these problems are potentially acute.[3]

2.3. Order-m Estimators:

As an alternative to estimators tied to the frontier $\mathcal{P}^{\partial t}$, we consider estimators based on the expected maximum output frontiers of order m proposed by Cazals *et al.* (2002). These allow the convexity assumption to be relaxed, and in addition permit noise (with zero expected value) in the output measures. Recall that the density $f^t(\boldsymbol{x}, \boldsymbol{y})$ has bounded support over the production set $\mathcal{P}^t$. Then $f^t(\boldsymbol{x}, \boldsymbol{y})$ implies the conditional distribution function $F^t_{y|x}(\boldsymbol{y}_0 \mid \boldsymbol{x}_0) = \Pr(\boldsymbol{y} \leq \boldsymbol{y}_0 \mid \boldsymbol{x} \leq \boldsymbol{x}_0)$. For a given level of inputs $\boldsymbol{x}_0$ in the interior of the support of $\boldsymbol{x}$, consider the m iid random variables $\{\boldsymbol{v}_j\}_{j=1}^m$, $\boldsymbol{v}_j \in \mathbb{R}^q_+$, drawn from

[3]Several algorithms for detecting outliers in high dimensional spaces have been proposed (*e.g.,* Wilson, 1993, 1995), but these involve substantial computational burden with large sample sizes.

the conditional distribution $F^t_{y|x}(\cdot \mid \boldsymbol{x}_0)$. Define the set

$$\mathcal{A}^t_m(\boldsymbol{x}_0) = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{p+q}_+ \mid \boldsymbol{x} \le \boldsymbol{x}_0,\ \bigcup_{j=1}^{m} \boldsymbol{y} \le \boldsymbol{v}_j \right\}. \tag{2.7}$$

Note that $\mathcal{A}^t_m(\boldsymbol{x}_0)$ is random, depending on the particular draw of m vectors from $F^t_{y|x}(\cdot \mid \boldsymbol{x}_0)$.

Analogous to (2.4), we may define the random distance function

$$D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{A}^t_m(\boldsymbol{x})) \equiv \inf \left\{ \theta > 0 \mid (\boldsymbol{x}, \boldsymbol{y}/\theta) \in \mathcal{A}^t_m(\boldsymbol{x}) \right\}. \tag{2.8}$$

For any $\boldsymbol{y} \in \mathbb{R}^q_+$, define the *expected maximum output level of order* m for all $\boldsymbol{x}$ such that $f^t_x(\boldsymbol{x}) = f^t(\boldsymbol{x}, \boldsymbol{y})/f^t(\boldsymbol{y} \mid \boldsymbol{x}) > 0$ as

$$\boldsymbol{y}^{\partial t}_m(\boldsymbol{x}) \equiv \boldsymbol{y}/E\left[D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{A}^t_m(\boldsymbol{x}_0)) \right]. \tag{2.9}$$

This is the output-oriented analog of the input measure defined by Cazals *et al.* (2002). The order-m analog of $\mathcal{P}^t$ may be formed by defining

$$\mathcal{P}^t_m \equiv \left\{ (\boldsymbol{x}, \boldsymbol{y}) \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t,\ y \le \boldsymbol{y}^{\partial t}_m(\boldsymbol{x}) \right\}, \tag{2.10}$$

which we call the *expected production set of order* m. Finally, we denote the closure of the compliment of $\mathcal{P}^t_m$ as $\mathcal{P}^{\partial t}_m$, and call this the *order-m frontier.*

To understand the order-m idea, consider $(\boldsymbol{x}, \boldsymbol{y})$ lying in the interior of $\mathcal{P}^t$. Then $(\boldsymbol{x}, \boldsymbol{y}/D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t))$ gives the projection of $(\boldsymbol{x}, \boldsymbol{y})$ onto the frontier $\mathcal{P}^{\partial t}$; given input quantities $\boldsymbol{x}$, $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t)^{-1}$ is the maximum feasible proportionate increase in the output quantities $\boldsymbol{y}$. On the other hand, $\boldsymbol{y}^{\partial t}_m(\boldsymbol{x})$ is the *expected* maximum output vector (with the same output proportions as $\boldsymbol{y}$) among m firms chosen randomly, conditional on their inputs being less than or equal to $\boldsymbol{x}$. Clearly, $\boldsymbol{y}^{\partial t}_m(\boldsymbol{x}) \le \boldsymbol{y}/D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t)$, and it can be shown that (i) $\lim_{m \to \infty} \boldsymbol{y}^{\partial t}_m(\boldsymbol{x}) = \boldsymbol{y}/D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}^t)$, and hence (ii) $\mathcal{P}^t_m \to \mathcal{P}^t$ as $m \to \infty$.[4] The

[4]Cazals *et al.* (2002, theorem 5.2) give a proof for the input-oriented analog of (i). Straightforward changes in notation lead to a proof for the output orientation used here, and (ii) follows directly from (i).

order-m concept relies on a different benchmark than traditional efficiency studies; rather than comparing a given firm's output to (an estimate of) the maximum *feasible* output, the firm's observed output quantities are compared to what could be expected from any m randomly chosen firms that use no more input quantities than the given firm.

Cazals *et al.* (2002) suggest a simple Monte Carlo technique that can be used to obtain nonparametric estimates of $E\left[D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{A}_m^t(\boldsymbol{x}))\right]$ and hence $\boldsymbol{y}_m^{\partial t}$. Note that a realization of the random distance function defined in (2.8) can be computed for a particular draw $\{\boldsymbol{v}_j\}_{j=1}^m$ by

$$D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{A}_m^t(\boldsymbol{x})) = \min_{j=1,\,\dots,\,m}\left[\max_{\ell=1,\,\dots,\,q}\left(\frac{y_\ell}{v_{j\ell}}\right)\right], \tag{2.11}$$

where y_ℓ and $v_{j\ell}$ are the ℓ-th elements of $\boldsymbol{y}$ and $\boldsymbol{v}_j$. To implement the Monte Carlo method, we draw variates $\boldsymbol{v}_j$ from the empirical analog of the conditional distribution $F_{y|x}^t(\cdot \mid \boldsymbol{x}_0)$, given by

$$\widehat{F}_{y|x,\,n_t}^t(\boldsymbol{y}_0 \mid \boldsymbol{x}_0) = \frac{\sum_{i=1}^{n_t} I(\boldsymbol{x}_i \leq \boldsymbol{x}_0,\ \boldsymbol{y}_i \leq \boldsymbol{y}_0)}{\sum_{i=1}^{n_t} I(\boldsymbol{x}_i \leq \boldsymbol{x}_0)}, \tag{2.12}$$

where $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{S}_{n_t}^t\ \forall\ i = 1,\ \dots,\ n_t$. If the point of interest is $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, the procedure works as follows:

[1] Draw m times, independently, with replacement, from the observations in $\mathcal{S}_{n_t}^t$ such that $\boldsymbol{x}_i \leq \boldsymbol{x}_0$; discard the input vectors and denote the sample of remaining output vectors by $\{\boldsymbol{v}_{kj}\}_{j=1}^m$.

[2] Compute

$$\widetilde{D}_k(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{S}_{n_t}^t, m) = \min_{j=1,\,\dots,\,m}\left[\max_{\ell=1,\,\dots,\,q}\left(\frac{y_{0\ell}}{v_{kj\ell}}\right)\right]$$

where $v_{kj\ell}$ and $y_{0\ell}$ are the ℓ-th elements of $\boldsymbol{v}_{kj}$ and $\boldsymbol{y}_0$.

[3] Repeat steps [1]–[2] K times to obtain $\left\{\widetilde{D}_k(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{S}_{n_t}^t, m)\right\}_{k=1}^K$.

[4] Compute

$$\widehat{D}_{m,n_t}(\boldsymbol{x}_0, \boldsymbol{y}_0) = \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{S}_{n_t}^t, m) = K^{-1}\sum_{k=1}^K \widetilde{D}_k(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{S}_{n_t}^t, m), \tag{2.13}$$

an estimator of $E\left[D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{A}_m^t(\boldsymbol{x}))\right]$.

An estimator $\widehat{\boldsymbol{y}}^{\partial t}_{m,n_t}$ of $\boldsymbol{y}^{\partial t}_m$ can be computed by replacing $E\left[D(\boldsymbol{x},\boldsymbol{y} \mid \mathcal{A}^t_m(\boldsymbol{x}))\right]$ with $\widehat{D}_{m,n_t}(\boldsymbol{x}_0,\boldsymbol{y}_0)$ in (2.9).

Additional insight can be gained by considering the following simple example. Suppose $p = q = 1$, and consider a DGP where

$$f(x,y) = \begin{cases} 2 & \forall\, x \in [0,1],\ y \leq x; \\ 0 & \text{otherwise.} \end{cases} \tag{2.14}$$

Then the production set corresponds to the right triangle with corners at (0,0), (0,1), and (1,1). It is easy to show that $f(y \mid x \leq x_0) = 2x_0^{-1}\left[1 - x_0^{-1}y\right]$ and hence $F_{y|x}(y_0 \mid x \leq x_0) = 2x_0^{-1}y_0 - x_0^{-2}y_0^2$. The order-$m$ frontier at input level $x_0 \in [0,1]$ is given by

$$\begin{aligned} \mathcal{P}^{\partial}_m(x_0) &= E\left[\max(y_1,\ \ldots\ ,\ y_m) \mid x \leq x_0\right] \\ &= \int_0^{x_0} 1 - \left[F_{y|x}(y \mid x_0)\right]^m\, dy \\ &= \int_0^{x_0} 1 - \left[2x_0^{-1}y - x_0^{-2}y^2\right]^m\, dy. \end{aligned} \tag{2.15}$$

The integral can be computed easily since the integrand involves a polynomial in y.

Draws of size $n = 100$ and $n = 1000$ were taken from this simple DGP to produce the plots shown in Figure 1. Both panels show two lines; the one with slope equal to 1 is the *true* boundary of the production set, while the line with lesser slope is the *true* order-m frontier with $m = 50$. In each panel, the stair-step pattern just below the production set boundary is the FDH frontier *estimate*, while the order-m frontier estimate lies below the FDH *estimate*, closer to the true order-m frontier.[5] For the case $p = q = 1$, both the FDH and order-m estimators have convergence rates of $n^{-1/2}$, and the plots in Figure 1 indicate that the estimates move closer to the true frontier as n is increased from 100 to 1000. It is also apparent from the plots that the order-m frontier deviates farther from the production set boundary as we move left to right along the horizontal axes. This is a consequence of the definition in (2.9) and (2.15). In general, the distance between the two

[5]For small values of x, the FDH and order-m estimates coincide. With $n = 100$, the range of values of x over which the two estimates coincide is larger than for the case with $n = 1000$.

frontiers at any given input level will depend on the particular choice of m, the variance of y, conditional on x, and the slope and curvature of the true production set boundary.

Cazals *et al.* (2002) prove that the order-m estimators have some interesting and useful properties. In particular, for finite, fixed m, $\widehat{\boldsymbol{y}}^{\partial t}_{m,n_t}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{y}^{\partial t}_m$. Root-$n$ consistency is unusual among non-parametric estimators; this result means that the order-m estimator avoids the curse of dimensionality that plagues DEA and FDH estimators. In addition, for fixed n_t, $\widehat{D}_{m,n_t}(\boldsymbol{x}_0, \boldsymbol{y}_0) \rightarrow D(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \widehat{\mathcal{P}}^t_{\text{FDH}})$ as $m \rightarrow \infty$; *i.e.*, for a given sample size, the order-m estimator converges to the FDH estimator as $m \rightarrow \infty$. Moreover, for finite m, the order-m estimator is far more robust to extreme values, noise, or outliers than either the DEA or FDH estimators, provided m is not too large relative to n.

The root-n consistency property is lost if the order-m estimator is used to estimate $\mathcal{P}^{\partial t}$. Consequently, we use the order-m estimator to estimate the order-m frontier, $\mathcal{P}^{\partial t}_m$, rather than $\mathcal{P}^{\partial t}$. Readers familiar with DEA and FDH estimators may find this puzzling at first, but should realize that the order-m frontier is merely an alternative benchmark by which to gauge the efficiency of production units. Rather than measuring a firm's performance relative to a potentially unreliable estimate of the *maximum feasible output* for the firm's observed inputs, we measure the firm's performance relative to the *expected maximum output* among m firms using input quantities *no greater than* those of the firm of interest.

The only remaining issue regarding the order-m estimators concerns the particular choice of m. Cazals *et al.* (2002) remark that the value of m can be viewed as a trimming parameter; its role is similar to that of the trimming parameter in trimmed mean estimators. They write (p. 7) that in practice, "a few values of m could be used to guide the manager of the production unit to evaluate its own performance." For the simple example described above, Figure 2 shows, descending from top to bottom, the true production set boundary and true order-m frontiers corresponding to $m = 1500$, 150, and 50, illustrating that increasing m moves the order-m frontier closer to $\mathcal{P}^{\partial t}$.

3. Dynamic Effects

The order-m concept discussed in the previous section is perhaps most useful in our application when it is used to examine dynamic changes in the banking industry. In the case of only one input and one output ($p = q = 1$), productivity can be measured by the ratio of output to input, and changes in productivity can be examined by comparing output-input ratios of firms at different points in time. With multiple inputs and multiple outputs, however, this simple approach does not work. Malmquist indices are typically used to examine productivity changes in multivariate settings; see Färe and Grosskopf (1996, 1998) for discussion.

Consider a bank with input and output quantities $(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1})$ at time t_1 and $(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2})$ at time $t_2 > t_1$. To measure changes in this bank's productivity from t_1 to t_2, we could use a Malmquist index similar to the one proposed by Färe, Grosskopf, Lindgren, and Roos (1992, 1994) and given by

$$\mathcal{M}(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1}, \boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}^{t_1}, \mathcal{P}^{t_2}) \equiv \left[\frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{V}(\mathcal{P}^{t_1}))}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{V}(\mathcal{P}^{t_1}))} \times \frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{V}(\mathcal{P}^{t_2}))}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{V}(\mathcal{P}^{t_2}))} \right]^{1/2}. \quad (3.1)$$

This index is the geometric mean of two ratios. The first ratio inside the brackets measures the change in productivity relative to the technology at time t_1, while the second ratio measures change in productivity relative to the technology prevailing at time t_2. Productivity change is measured relative to the conical hull of the production set defined by (2.1) for either time t_1 or t_2. A particular bank either moves closer to the boundary of this conical hull (becoming more productive), or farther from the boundary (becoming less productive). Values of the Malmquist index greater than 1 indicate an improvement in productivity; values less than 1 indicate a decrease in productivity, while a value of 1 indicates no change.

The true, unknown Malmquist index in (3.1) is typically estimated by replacing $\mathcal{P}^{t_1}$ and $\mathcal{P}^{t_2}$ with $\widehat{\mathcal{P}}^{t_1}_{\text{DEA}}$ and $\widehat{\mathcal{P}}^{t_2}_{\text{DEA}}$; consequently, this estimator inherits all the problems of the DEA estimator. Since we doubt the viability of DEA estimators for our application, we define an *order-m Malmquist index* to measure productivity change relative to (the conical hull of) the frontier of the expected production set of order-m ($\mathcal{P}^t_m$) defined by (2.10), instead of the difficult-to-estimate $\mathcal{P}^{\partial t}$. Our order-m Malmquist index is defined by replacing $\mathcal{P}^{t_1}$ and $\mathcal{P}^{t_2}$ in (3.1) with $\mathcal{P}^{t_1}_m$ and $\mathcal{P}^{t_2}_m$ to obtain

$$\mathcal{M}(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1}, \boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}^{t_1}_m, \mathcal{P}^{t_2}_m) \equiv \left[\frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{V}(\mathcal{P}^{t_1}_m))}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{V}(\mathcal{P}^{t_1}_m))} \times \frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{V}(\mathcal{P}^{t_2}_m))}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{V}(\mathcal{P}^{t_2}_m))}\right]^{1/2}. \quad (3.2)$$

Estimates $\widehat{\mathcal{M}}_{m,n_{t_1},n_{t_2}}(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1}, \boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2})$ of the modified index in (3.2) can be obtained by replacing the unknown, true distance functions on the right-hand side with consistent estimates. For an arbitrary point $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, $D(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{V}(\mathcal{P}^t_m))$ can be estimated by first computing estimates $\widehat{D}_{m,n_t}(\boldsymbol{x}^t_i, \boldsymbol{y}^t_i)$, where $i = 1, \ldots, n_t$ indexes the sample observations at time t. Then project the sample observations onto the frontier of the expected production set of order m by computing $\{(\boldsymbol{x}^t_i, \widetilde{\boldsymbol{y}}^t_i)\}_{i=1}^{n_t}$, where $\widetilde{\boldsymbol{y}}_i = \boldsymbol{y}_i / \widehat{D}_{m,n_t}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ is the empirical analog of (2.9). Next, analogous to (2.6), we can write

$$\mathcal{V}\left(\widehat{\mathcal{P}}^t_m\right) = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{p+q}_+ \mid \boldsymbol{y} \leq \sum_{i=1}^n \gamma_i \widetilde{\boldsymbol{y}}^t_i,\ \boldsymbol{x} \geq \sum_{i=1}^n \gamma_i \boldsymbol{x}^t_i,\ \gamma_i \geq 0\ \forall\, i = 1, \ldots, n\right\}, \quad (3.3)$$

which describes the convex cone of $\widehat{\mathcal{P}}^t_m$. Then an estimate $D(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{V}(\widehat{\mathcal{P}}^t_m))$ of $D(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{V}(\mathcal{P}^t_m))$ can be computed by linear programming techniques. Substituting estimates for the true distance functions in (3.2) yields an estimate of the order-m Malmquist index.

Just as the traditional Malmquist index in (3.1) can be decomposed in various ways to identify the sources of changes in productivity, our order-m Malmquist index can be decomposed in analogous ways. Various decompositions have been proposed in the literature. While there are perhaps infinitely many possibilities, we apply the order-m analogy of the decomposition proposed independently by Wheelock and Wilson (1999) and Zofío and Lovell (1998) by writing

$$\begin{aligned}
\mathcal{M}(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1}, \boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}_m^{t_1}, \mathcal{P}_m^{t_2}) = &\underbrace{\left(\frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}_m^{t_2})}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{P}_m^{t_1})}\right)}_{=\Delta M_Eff} \times \\
&\underbrace{\left(\frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{V}(\mathcal{P}_m^{t_2}))/D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}_m^{t_2})}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{V}(\mathcal{P}_m^{t_1}))/D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{P}_m^{t_1})}\right)}_{=\Delta M_SEff} \times \\
&\underbrace{\left(\frac{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{P}_m^{t_1})}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{P}_m^{t_2})} \times \frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}_m^{t_1})}{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}_m^{t_2})}\right)^{1/2}}_{=\Delta M_Fron} \times \\
&\underbrace{\left\{\left[\frac{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{V}(\mathcal{P}_m^{t_1}))/D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{P}_m^{t_1})}{D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{V}(\mathcal{P}_m^{t_2}))/D(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1} \mid \mathcal{P}_m^{t_2})}\right] \times \left[\frac{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{V}(\mathcal{P}_m^{t_1}))/D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}_m^{t_1})}{D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{V}(\mathcal{P}_m^{t_2}))/D(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2} \mid \mathcal{P}_m^{t_2})}\right]\right\}^{1/2}}_{=\Delta M_SFron}.
\end{aligned} \tag{3.4}$$

As with the order-m Malmquist index defined in (3.2), the components ΔM_Eff, ΔM_SEff, ΔM_Fron, and ΔM_SFron can be estimated by replacing the true, unknown sets $\mathcal{P}_m^{t_j}$ and $\mathcal{V}\left(\mathcal{P}_m^{t_j}\right)$ with estimates $\widehat{\mathcal{P}}_m^{t_j}$ and $\mathcal{V}\left(\widehat{\mathcal{P}}_m^{t_j}\right)$, $j = 1, 2$.

The first term in the decomposition shown in (3.4), labeled ΔM_Eff, measures changes in order-m technical efficiency, with values greater than (equal to, less than) 1 indicating improving (unchanged, decreasing) efficiency. Order-m technical efficiency may change over time because a bank moves relative to the order-m frontier, because the order-m frontier changes over time, or because of a combination of both factors.

The second term, ΔM_SEff, measures changes in the order-m *scale efficiency* faced by a particular bank. To understand this term, consider the ratio $D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{P}_m^t))/D(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P}_m^t)$, which compares distances from a particular point $(\boldsymbol{x}, \boldsymbol{y})$ to (i) the order-m frontier and (ii) the conical hull of the order-m frontier in the output direction (*i.e.,* orthogonal to the input axes). If these distances are the same, then the projection of $(\boldsymbol{x}, \boldsymbol{y})$ onto the order-m frontier in the output direction is in a region where the order-m frontier is locally homogeneous of degree 1, *i.e.,* where constant returns to scale prevail. In this case, $(\boldsymbol{x}, \boldsymbol{y})$ is said to be order-m scale-efficient. If the distances are not the same, then

$D(\boldsymbol{x},\boldsymbol{y} \mid \mathcal{V}(\mathcal{P}^t_m))/D(\boldsymbol{x},\boldsymbol{y} \mid \mathcal{P}^t_m) < 1$, and the projection of $(\boldsymbol{x},\boldsymbol{y})$ onto the order-m frontier in the output direction is in a region where the order-m frontier displays either increasing or decreasing returns to scale. The denominator of Δ*M_SEff* measures scale efficiency faced by a particular bank at time t_1, while the numerator measures scale efficiency at time t_2, so that the ratio gives a measure of change in scale efficiency. Hence, Δ*M_SEff*$(>,=,<)1$ as scale efficiency improves, remains constant, or decreases for a particular bank.

The third term on the right-hand side of (3.4), Δ*M_Fron*, measures changes in the order-m frontier over time. The first ratio inside the parentheses will be greater than (equal to, less than) 1 when the order-m frontier shifts upward (remains unchanged, shifts downward) at the location where $(\boldsymbol{x}^{t_1},\boldsymbol{y}^{t_1})$ is projected in the output direction onto the order-m frontier. The second ratio behaves similarly for shifts in the order-m frontier at the location where $(\boldsymbol{x}^{t_2},\boldsymbol{y}^{t_2})$ is projected in the output direction onto the order-m frontier; Δ*M_Fron* is simply the geometric mean of these two ratios.

The fourth term on the right-hand side of (3.4), Δ*M_SFron*, measures changes in order-m scale efficiency due to changes in the order-m frontier, as opposed to changes in banks' locations. We label this effect *changes in scale of the order-m frontier,* analogous to the terminology in Wheelock and Wilson (1999). To understand this term, consider the first term in square brackets in Δ*M_SFron*. The numerator is the same as the denominator in Δ*M_SEff*; hence this numerator measures order-m scale efficiency at the point where $(\boldsymbol{x}^{t_1},\boldsymbol{y}^{t_1})$ is projected (in the output direction) onto the order-m frontier at time t_1. The corresponding denominator is similar, but $\mathcal{P}^{t_2}_m$ replaces $\mathcal{P}^{t_1}_m$; consequently, the denominator measures order-m scale efficiency of the order-m frontier in the second period, at the location where $(\boldsymbol{x}^{t_1},\boldsymbol{y}^{t_1})$ is projected (in the output direction) onto $\mathcal{P}^{t_2}_m$. Therefore, the ratio inside the square brackets will be less than (equal to, greater than) 1 if the distance between $\mathcal{P}^{t_2}_m$ and $\mathcal{V}(\mathcal{P}^{t_2}_m)$ is smaller than (equal to, greater than) the distance between $\mathcal{P}^{t_1}_m$ and $\mathcal{V}(\mathcal{P}^{t_1}_m)$ along the path where $(\boldsymbol{x}^{t_1},\boldsymbol{y}^{t_1})$ is projected toward the frontiers in a direction orthogonal to the input axes and parallel to the output axes. In other words, the first term in square brackets in ΔM_SFron compares scale efficiency of $(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1})$ relative to $\mathcal{P}_m^{\partial t_1}$ and $\mathcal{P}_m^{\partial t_2}$ (in the output direction); values less than (equal to, greater than) one correspond to increasing (constant, decreasing) order-m scale *in*efficiency for a firm located at $(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1})$ in both periods.

The second term in square brackets in ΔM_SFron measures similar phenomena, but relative to $(\boldsymbol{x}^{t_2}, \boldsymbol{y}^{t_2})$ instead of $(\boldsymbol{x}^{t_1}, \boldsymbol{y}^{t_1})$; *i.e.,* relative to a particular bank's location at t_2 as opposed to its location at t_1. The phenomena could be different at different locations, and ΔM_SFron is simply the geometric mean of terms which measure the effect relative to a bank's location at t_1 and at t_2. Thus, ΔM_SFron measures changes in order-m scale efficiency that would result only from shifts or changes in shape of the order-m frontier, with $\Delta M\_SFron(<,=,>)1$ as order-m scale efficiency (increases, remains constant, decreases) along fixed paths in the output direction.

To illustrate factors that influence ΔM_SFron, consider Figure 3, which illustrates two extreme possibilities for the simple case of one input and one output ($p = q = 1$). In Panel A, $\mathcal{V}(\mathcal{P}_m^{t_1}) = \mathcal{V}(\mathcal{P}_m^{t_2})$, but the order-$m$ frontier is less curved at t_2 than at t_1. In this case, $\Delta M\_SFron < 1$ for the firm located at point A at t_1 and point B at t_2. In Panel B, the order-m frontier shifts upward by the same distance everywhere, so $\mathcal{V}(\mathcal{P}_m^{t_1}) \subset \mathcal{V}(\mathcal{P}_m^{t_2})$. For a firm located at point C in both periods, $\Delta M\_SFron > 1$. Although the firm does not move, its order-m scale efficiency decreases from t_1 to t_2 due to the shift in the order-m frontier.

4. Bank Production and Data

Distance function estimation using the estimators described in Section 2 requires the specification of production inputs and outputs. We define five inputs and five outputs which, with one exception (the measure of off-balance sheet output), are those used by Berger and Mester (2003). Our inputs are purchased funds (*x1*), which consists of time deposits over \$100,000, foreign deposits, federal funds purchased, and various other borrowed funds; core deposits (*x2*), which consists of domestic transactions accounts, time deposits under \$100,000 and savings deposits; labor (*x3*); physical capital (*x4*), which consists of premises and other fixed assets; and financial equity capital (*x5*). Our outputs are consumer loans (*y1*), business loans (*y2*), real estate loans (*y3*), securities (*y4*), and off-balance sheet items (*y5*), which consist of total non-interest income minus service charges on deposits.[6] With the exception of labor input (which is measured as full-time equivalent employees) and off-balance sheet items (which are measured in terms of net flow of income), inputs and outputs are stocks measured by dollar amounts reported on banks' balance sheets, rather than number of loans or deposits, or loan income or deposit interest expenses. This approach is consistent with the widely used "intermediation" model of Sealey and Lindley (1977).

Our data come from Reports of Income and Condition (Call Reports) for all U.S. commercial banks at year-end 1984, 1993, and 2002. We omitted banks with missing or negative values for any input or output, and we converted dollar values to 1996 prices using the GDP deflator. After examining the marginal distributions of each variable and omitting observations with impossible values, we used a leave-one-out version of the order-m estimator to search for outliers as described by Simar (2003). This approach did not suggest any obvious outliers, and so we deleted no additional observations (complete details are available from the authors upon request). Hence, we retain 13,845, 10,661, and 7,561 observations for 1984, 1993, and 2002, respectively. For each year, our sample consists of at least 95 percent of all commercial banks in operation. Descriptive statistics for each input and output are reported in Table 1.

Although our sample sizes may seem large, at least by parametric standards, they are in fact small for the non-parametric DEA and FDH estimators given the high dimensionality of our application. With five inputs (p) and five outputs (q), we have $(p + q) = 10$ dimen-

[6] Of the various commonly used measures of off-balance sheet output, this definition is the most consistently measurable across banks and over time (see Clark and Siems, 2002). See Berger and Mester (2003) for additional details about the computation of the other inputs and outputs.

sions. The potential for the curse of dimensionality to affect DEA and FDH estimation can be gauged by a rough comparison of equivalent sample sizes. For the order-m, DEA, and FDH estimators, we have convergence rates of $n^{-1/2}$, $n^{-2/(p+q+1)}$, and $n^{-1/(p+q)}$, respectively. Thus, to achieve the same order of magnitude in estimation error as obtained with the order-m estimator with $n = 100$ observations, the DEA estimator would require $\left(100^{-1/2}\right)^{-11/2} = 316,227$ observations, while the FDH estimator would require $\left(100^{-1/2}\right)^{-10} = 10^{10}$ observations. As we show below, our FDH and DEA estimates are indeed affected by the curse of dimensionality.

5. Empirical Results

To illustrate the problems with measuring efficiency relative to estimates of the production set boundary $\mathcal{P}^{\partial t}$, we first used the FDH and DEA estimators to produce distance function estimates for all banks in each year of our sample. Table 2 reports summary statistics for these estimates, by estimator and by year. The FDH estimates are strictly equal to 1.0 for all banks in each year, indicating that *all* banks lie on the (estimated) efficient frontier, implying that no banks are inefficient. This implausible result reflects the curse of dimensionality—even with several thousand observations, the FDH estimator yields no useful information about inefficiency here.

The DEA estimates are similarly problematic. DEA differs from FDH only in that DEA imposes convexity on the production frontier $\mathcal{P}^t$. Because FDH suggested that all banks are perfectly efficient, any inefficiency detected by DEA would necessarily result solely from the convexity assumption. As shown in Table 2, the DEA estimates of mean *in*efficiency range from about $(1.0 - 0.868) \times 100 = 0.132$ percent in 1984 to $(1.0 - 0.975) \times 100 = 0.025$ percent in 2002 (recall that DEA distance function estimates are weakly bounded above at 1.0, with an estimate of 1.0 indicating that an observation lies on the estimated efficient frontier). Thus, the DEA estimates suggest that on average banks became more efficient over time. One should have little confidence in these results, however, since the DEA frontier is merely the convexified FDH frontier. The convergence rates of both the DEA and FDH estimators in our application are too slow, and the dispersion of the data in any particular year is too great, to allow any reasonable level of confidence in these estimates. Simply put, neither the FDH nor DEA estimator conveys useful information in our particular application.

As discussed previously, by switching to a different benchmark than $\mathcal{P}^{\partial t}$, namely the order-m frontier, we can employ $\sqrt{n}$-consistent estimators and avoid the curse of dimensionality that plagues the DEA and FDH estimators. Moreover, the order-m estimator is robust with respect to outliers and other noise in the data.

We computed order-m efficiency estimates for all banks in each year 1984, 1993, and 2002 using a variety of values of m, ranging from 75 to 3000. Because the estimated order-m frontier approaches the FDH frontier as m increases, and because every bank in our sample has an FDH efficiency estimate equal to 1 in each year (*i.e.,* every bank lies on its contemporaneous FDH frontier), it is necessarily the case that all banks in each year of our sample lie on or above the contemporaneous estimated order-m frontier. Consequently, our contemporaneous order-m efficiency estimates are equal to or greater than 1 in every case.

Figure 4 shows kernel estimates of the densities of the contemporaneous order-m efficiency estimates for 2002, by quintiles of banks' total assets, for $m = 75$, 150, 300, and 1500.[7] As expected, the densities shown in Figure 4 shift to the left and collapse toward 1 as m is increased. For each value of m, the densities become more disperse moving from the first quintile toward the fifth quintile. As noted in Section 2.3, this reflects the properties of the conditional distribution of outputs, given input quantities. Because banks vary widely in terms of their sizes (as measured by total assets), it is not surprising that this conditional density becomes more disperse as banks become larger. We would expect

[7] To avoid the problem of bias in kernel density estimates near boundaries of support, we used the reflection method described by Silverman (1986) and Scott (1992). We used a Gaussian kernel, and chose bandwidths using the Sheather and Jones (1991) two-stage plug-in procedure.

heteroskedasticity in bank outputs for the usual reasons.[8]

The large numbers of banks in our samples prevent us from displaying estimates for individual banks. A bank manager might want to estimate order-m efficiency to assess her bank's performance relative to the expected performance of peer banks; regulators and shareholders, including prospective acquirers, might also find this information useful. Our goal is to gauge industry performance, however, rather than the performance of specific banks. Consequently, we divided banks into deciles according to total assets for each sample year. We constructed hypothetical "median" banks having the median values of each input and output within each decile in each year.

Table 3 shows contemporaneous order-m efficiency estimates with $m = 150$ for the median banks in each year, with bootstrap estimates of 95-percent confidence intervals. We report similar estimates for $m = 75$, 300, and 1500 in a separate appendix (available from the authors on request). As expected, efficiency estimates are uniformly smaller as m is increased, consistent with the density estimates shown in Figure 4, but otherwise the patterns are similar for different values of m.

The estimates shown in Table 3 (and the corresponding tables in the appendix) indicate some tendency for efficiency to increase with bank size. Many of the differences across deciles are not statistically significant, however. For 2002, for example, confidence intervals for the 1st–5th deciles overlap, as do the confidence intervals of the 1st, 6th and 7th deciles. Confidence intervals for the 6th–9th deciles also overlap, but those for the 8th and 9th deciles lie above the confidence intervals for deciles 1–5. The estimated confidence interval for the 10th decile lies above the intervals for each of the first 9 deciles.

The efficiency estimates show little variation over time; comparing estimates for 1984 and 2002, the confidence intervals overlap for each of the corresponding deciles except the first. Comparisons across time are complicated, however, by the fact that order-m efficiency

[8] The patterns for 1984 and 1993 are very similar to those shown in Figure 4. Density estimates by quintile for 1984 and 1993 are shown in a separate appendix available from the authors upon request.

can change due to either (i) movement of the median banks over time; (ii) movement of the order-m frontier over time; or (iii) a combination of (i) and (ii). Estimates of the order-m Malmquist productivity index and its components defined in (3.4) provide insight to help disentangle these phenomena.[9]

Table 4 reports estimates of the order-m ($m = 150$) Malmquist productivity index and its decomposition for 1984–1993, 1993–2002, and the entire sample period 1984–2002. Similar estimates for $m = 75$, 300, and 1500 are reported in corresponding tables in the separate appendix mentioned earlier. With few exceptions, the estimates are qualitatively unaffected by the choice of m. Estimates of the order-m Malmquist productivity index are reported in the second column of Table 4, along with an indication of whether the estimates are significantly different from 1.[10] Recall from the discussion in Section 3 that values of the index greater (less) than 1 indicate improving (decreasing) productivity. For the first half of our study period, we find statistically significant increases in productivity for deciles 2 and 3, and a significant decrease in decile 9. The changes are small, however, with increases of 0.7 and 2.1 percent, while the decline in decile 9 is 1.6 percent over 9 years.

The results for the second half of our study period are more dramatic. The estimates in Table 4 indicate that for 1993–2002, productivity increased significantly in all deciles, with the increases ranging from 3.1 percent in the 3rd decile to 19.7 percent in the 8th decile. The first nine years of our study period were tumultuous years for the U.S. banking industry, with low profits and many failures. Our results indicate that productivity changed little in this period. The industry turned around during the 1990s, however, and saw large

[9] One of the advantages of focusing on median-decile banks becomes clear here if one recalls that the Malmquist index for a particular bank requires its existence in both periods. By defining median banks, we do not have to worry about unbalanced panels as we would if we tried to compute the Malmquist index for each bank, and then take medians within a decile.

[10] Statistical significance is determined by bootstrapping. Since the order-m frontier is not at the boundary of support for $f(\boldsymbol{x}, \boldsymbol{y})$, a simple bootstrap that resamples from the empirical distribution of the data can be used here, avoiding the complexity required when bootstrapping DEA estimators as described in Kneip *et al.* (2003).

gains in productivity accompanying record profits.

Looking across the entire period 1984–2002, our results suggest that productivity gains were modest. Our estimates of productivity change are significantly different from 1 in just the first four deciles, with small increases in deciles 1, 2, and 4, and a small decrease in decile 3. While significant statistically, the estimates for these deciles are small, ranging from −1.8 percent to 3.2 percent over 18 years. Although the productivity gains in the second half of the period were dramatic, we obtain insignificant estimates for deciles 4–10 for the period as a whole. Moreover, the results for decile 3 indicate a significant *decrease* in productivity for the overall period, but significant *increases* in each of the two halves, *i.e.,* for 1984–1993 and 1993–2002. These results illustrate that the Malmquist productivity indices defined in (3.1) and (3.2) do not satisfy the circular test, as noted by Färe, Grosskopf, Norris, and Zhang (1994). In other words, the index is path-dependent; the square root of the geometric average of estimates for the two sub-periods will typically differ from the estimate for the overall period. In the case of decile 3, the results should be taken with some caution.[11]

Columns 3–6 of Table 4 give, for median-decile banks, estimates of (i) changes in order-m efficiency, $\widehat{\Delta M\_Eff}$; (ii) changes in order-m scale efficiency, $\widehat{\Delta M\_SEff}$; (iii) changes in the order-m frontier, $\widehat{\Delta M\_Fron}$; and (iv) changes in scale of the order-m frontier, $\widehat{\Delta M\_SFron}$. Note that each of these estimates ask more of the data than our estimates of productivity change since these estimates attempt to identify the *sources* of productivity change. Just as F-tests frequently reject the null hypothesis that all slope coefficients in a linear regression equal zero when individual t-statistics fail to reject for each coefficient, it is not surprising that fewer estimates in columns 3–6 are significant than was the case for estimates of productivity change in column 2.

Comparing columns 3–4 in Table 4, we see that except in three instances (decile 5,

[11]Overall, the results for different values of m are similar. However, for 1984–2002, estimates of productivity gains for the higher deciles are larger using m =75 than those for $m = 150$ and are statistically significant. Using $m = 1500$, the estimates are similar to those for $m = 150$, but are statistically significant.

1984–1993; decile 10, 1993–2002; and decile 7, 1984–2002), whenever either $\widehat{\Delta M\_Eff}$ or $\widehat{\Delta M\_SEff}$ is significantly different from one, the other is also significantly different from one. In addition, whenever either estimate is significantly greater (less) than one, the other is less (greater) than one. In other words, in every case where one estimate is significant, ΔM_Eff and ΔM_SEff partially off-set each other in determining productivity change. Where order-m scale efficiency improves, the median bank becomes less order-m efficient; where the median bank becomes more order-m efficient, order-m scale efficiency declines. In each of these cases, these off-setting effects result in estimates of (order-m) productivity change that are small, *i.e.,* close to one. Thus, although overall productivity changes are small in these cases, other changes occur.

Only two estimates of changes in the order-m frontier ($\widehat{\Delta M\_Fron}$) for 1984–1993 and 1984–2002 are significantly different from 1, but five are significant for 1993–2002. For the two halves of our study period, the significant estimates are greater than 1 in each case, suggesting the order-m frontier advanced for the 2nd and 7th deciles in 1984–1993, and for the 3rd and 7–10th deciles for 1993–2002. The advance in the order-m frontier over 1993–2002 among the largest deciles drive, in part, the large increases in productivity for these deciles over that period.

Estimates of ΔM_Fron for the overall period, 1984–2002, are significant (from 1) only for the 1st and 6th deciles. The estimate for the 1st decile indicates a large decline of 18.3 percent. Interestingly, this large decline is entirely off-set by $\widehat{\Delta M\_SFron}$, suggesting a shift in the order-m frontier opposite the direction depicted in Figure 3(b). The estimates of ΔM_SFron in the last column of Table 4 are significantly different from 1 in several other cases for both 1984–1993 and 1993–2002, but show no clear pattern.

Taken as a whole, our estimates reveal little or no increase in productivity across 1984–1993, but a large increase during 1993–2002, when the U.S. banking industry enjoyed high profits and few failures, and the U.S. economy as a whole saw an increased rate of productivity growth. The Malmquist decomposition reveals a less clear cut pattern of changes in technology and efficiency, with considerable variation across banks of different sizes and across different periods. For larger banks, the large gains in productivity during 1993–2002 coincide with a significant outward shift in the order-m frontier, consistent with advancing technology. Such was not the case during 1984–1993, however, when the order-m frontier showed little movement.

6. Conclusions

Less-than-fully parametric estimators have become increasingly popular for studying the performance of U.S. commercial banks as many researchers have concluded that the widely used parametric functions fail to adequately represent cost or profit relationships in banking. Data Envelopment Analysis (DEA) has been the most widely applied non-parametric estimator to study commercial banks. Although DEA is flexible in the sense that no functional form assumptions are needed, it is potentially very sensitive to outliers and other noise in the data. Moreover, as estimators of the boundary of the production set, DEA and FDH estimators suffer from slow convergence rates due to the curse of dimensionality, limiting their usefulness for estimating efficiency in cases with several inputs or outputs, or when sample sizes are not extremely large. Although our samples consist of several thousand bank observations, we are unable to produce meaningful estimates of inefficiency using the FDH estimator because of the curse of dimensionality—all banks are estimated to lie on the efficient frontier. DEA differs from FDH in that DEA assumes that the efficient frontier is convex, and because of this assumption our DEA estimates of inefficiency vary across banks.

The order-m idea relies on a different benchmark for gauging efficiency, productivity changes, etc., and permits robust, $\sqrt{n}$-consistent estimation, thus avoiding many of the problems with DEA and FDH estimators. In addition, because we are not estimating the boundary of support of the density $f(\boldsymbol{x}, \boldsymbol{y})$—*i.e.,* the density of inputs and outputs—the bootstrap we use for inference is much simpler than what is required for DEA estimators (see Simar and Wilson, 2000, for discussion).

The order-m frontier provides, for a given bank, the output level that is the *expected* best among m draws of banks using no larger input quantities than the given bank; as such, it gives a measure of what is best *on average* among any m of the bank's peers. Over time, a single bank can push the traditional frontier—the boundary $\mathcal{P}^{\partial t}$ of the production set—upward, but for the order-m frontier to change, something bigger has to happen. In particular, for the order-m frontier to shift upward, *many* banks must increase their outputs, as opposed to possibly a single bank; in other words, the performance of the industry as a whole must change. In this sense, our use of the order-m concept allows two improvements over traditional approaches: (i) we can use a $\sqrt{n}$-consistent estimator, and (ii) the frontier we estimate reveals more information about the behavior or performance of the industry as a whole, rather than perhaps only a few banks observed near the boundary of the production set.

Our empirical results reveal a substantial increase in productivity across banks of all sizes between 1993 and 2002, with productivity gains the largest for banks in the larger asset-size deciles. Over the entire period 1984–2002, productivity gains were more modest, and generally not statistically significant. The sources of productivity gains during 1993–2002 varied across size deciles. However, technological progress, as indicated by outward expansion of the order-m frontier, and improvement in order-m scale efficiency largely account for productivity gains by the larger banks that had the biggest overall gains in productivity. Being, to our knowledge, the first study to estimate efficiency and changes in productivity relative to the order-m frontier, it is difficult to compare our findings with those of other studies. However, our results do seem consistent with the performance of the U.S. banking industry since the early 1990s and the changes in the size distribution of in favor of larger banks.

REFERENCES

Alam, I. (2001), A nonparametric approach for assessing productivity dynamics of large U.S. banks, *Journal of Money, Credit, and Banking* 33, 121–139.

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

Berger, A.N. and L.J. Mester (2003), Explaining the dramatic changes in performance of US banks: technological change, deregulation, and dynamic changes in competition, *Journal of Financial Intermediation* 12, 57–95.

Cazals, C., J.P. Florens, and L. Simar (2002), Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* 106, 1–25.

Clark, J.A. and T.F. Siems (2002), X-efficiency in banking: Looking beyond the balance sheet, *Journal of Money, Credit and Banking* 34, 987–1013.

Deprins, D., L. Simar, and H. Tulkens (1984), "Measuring Labor Inefficiency in Post Offices," in *The Performance of Public Enterprises: Concepts and Measurements*, ed. by M. Marchand, P. Pestieau and H. Tulkens. Amsterdam: North-Holland, 243–267.

Efromovich, S. (1999), *Nonparametric Curve Estimation: Methods, Theory, and Applications,* New York: Springer-Verlag New York, Inc.

Färe, Rolf. *Fundamentals of Production Theory.* Berlin: Springer Verlag, 1988.

Färe, R. and S. Grosskopf (1996), *Intertemporal Production Frontiers: With Dynamic DEA,* Boston: Kluwer Academic Publishers.

Färe, R. and S. Grosskopf (1998), "Malmquist productivity indexes: A survey of theory and practice", in R. Färe, S. Grosskopf and R. Russell (eds.), *Essays in Honor of Sten Malmquist,* Dordrecht: Kluwer Academic Publishers.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1992), Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach, *Journal of Productivity Analysis* 3, 85–101.

Färe, R., S. Grosskopf, B. Lindgren and P. Roos (1994), "Productivity Developments in Swedish Hospitals: A Malmquist Output Approach" in A. Charnes, W. Cooper, A. Lewin and L. Seiford (eds.), *Data Envelopment Analysis: Theory, Methodology and Applications,* Dordrecht: Kluwer Academic Publishers.

Färe, R., S. Grosskopf, M. Norris and Z. Zhang (1994), Productivity growth, technical progress, and efficiency change in industrialized countries, *American Economic Review* 84, 66–83.

Gallant, A.R. (1981), On the bias in flexible functional forms and an essentially unbiased form, *Journal of Econometrics* 15,211–245.

Gallant, A.R. (1982), Unbiased determination of production technologies, *Journal of Econometrics* 20, 285–323.

Gijbels, I., E. Mammen, B.U. Park, and L. Simar (1999), On estimation of monotone and concave frontier functions, *Journal of the American Statistical Association* 94, 220–228.

Kneip, A., B. Park and L. Simar (1998), A note on the convergence of nonparametric DEA efficiency measures, *Econometric Theory* 14, 783–793.

Kneip, A., L. Simar, and P.W. Wilson (2003), "Asymptotics for DEA Estimators in Non-parametric Frontier Models," unpublished working paper, Department of Economics, University of Texas at Austin, Austin, Texas 78712.

Korostelev, A., L. Simar, and A.B. Tsybakov (1995a), Efficient estimation of monotone boundaries, *The Annals of Statistics* 23, 476–489.

Korostelev, A., L. Simar, and A.B. Tsybakov (1995b), On estimation of monotone and convex boundaries, *Publications de l'Institut de Statistique des Universités de Paris XXXIX* 1, 3–18.

McAllister, P.H., and D. McManus, 1993, Resolving the scale efficiency puzzle in banking, *Journal of Banking and Finance* 17, 389–405.

Park, B., L. Simar, and C. Weiner (2000), FDH efficiency scores from a stochastic point of view, *Econometric Theory* 16, 855–877.

Scott, D. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization,* New York: John Wiley & Sons, Inc.

Sealey, C. and J. Lindley (1977), Inputs, outputs, and a theory of production and cost at depository financial institutions, *Journal of Finance* 32, 1251–1266.

Sheather, S.J., and M.C. Jones (1991), A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society* B, 53, 684–690.

Shephard, R. (1970), *Theory of Cost and production Functions,* Princeton: Princeton University Press.

Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis,* Chapman and Hall, London.

Simar, L. (2003), Detecting outliers in frontier models: A simple approach, *Journal of Productivity Analysis,* forthcoming.

Simar, L. and P.W. Wilson (1998), "Productivity Growth in Industrialized Countries", Discussion Paper #9810, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L., and P.W. Wilson (1999), Estimating and bootstrapping Malmquist indices, *European Journal of Operational Research* 115, 459–471.

Simar, L. and Wilson, P.W. (2000), Statistical inference in nonparametric frontier models: The state of the art, *J. Productivity Anal.* 13, 49–78.

Shephard, Ronald W. *Theory of Cost and Production Functions.* Princeton: Princeton University Press, 1970.

Wheelock, D.C. and P.W. Wilson (1999), Technical progress, inefficiency, and productivity change in U.S. banking, 1984–1993, *Journal of Money, Credit, and Banking* 31, 212–234.

Wheelock, D.C. and P.W. Wilson (2000), Why do banks disappear? The determinants of US Bank failures and acquisitions, *Review of Economics and Statistics* 82, 127–138.

Wheelock, D.C. and P.W. Wilson (2001), New evidence on returns to scale and product mix among U.S. commercial banks, *Journal of Monetary Economics* 47, 653–674.

Wilson, P.W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics* 11, 319–323.

Wilson, P.W. (1995), Detecting influential observations in Data Envelopment Analysis, *Journal of Productivity Analysis* 6, 27–45.

Zofio, J.L. and C.A.K. Lovell (1998), "Yet Another Malmquist Productivity index Decomposition", unpublished working paper, Department of Economics,

TABLE 1
Summary Statistics for Inputs and Outputs

Variable	Mean	Std. Dev.	Minimum	Maximum
1984 (13,845 observations):				
x1	98437.67	1844286.26	2.77	125372370.00
x2	131835.34	903581.66	397.29	63983942.41
x3	107.36	971.78	2.00	70608.00
x4	3760.55	35118.39	0.00	2459856.04
x5	14882.39	133900.34	78.90	8745847.18
y1	25221.80	192967.57	2.77	12506921.37
y2	85759.26	1400146.19	1.38	85366832.78
y3	36745.39	327376.55	2.77	27765780.73
y4	95529.18	969083.83	424.97	64878183.83
y5	1901.62	29304.71	0.00	1852159.47
1993 (10,661 observations):				
x1	116220.80	2117563.50	6.33	143345290.00
x2	216090.34	1452941.59	65.41	83460280.62
x3	135.37	1148.68	2.00	69994.00
x4	5389.80	55765.51	0.00	3208144.32
x5	28182.94	245455.96	55.91	12342019.20
y1	34766.61	278353.22	6.33	11340858.74
y2	83878.00	1348162.04	3.16	101816647.00
y3	87450.86	669762.41	2.11	44684038.40
y4	149637.49	1497002.20	588.67	71500261.63
y5	5454.34	92930.05	0.00	5943664.94
2002 (7,561 observations):				
x1	335768.51	6977635.87	24.28	414110242.00
x2	391020.01	4852819.65	530.53	286055211.00
x3	214.25	2734.63	2.00	129545.00
x4	9034.71	115750.55	0.00	5646974.19
x5	69662.55	958245.76	244.58	46050714.86
y1	61435.03	908139.33	2.70	39638521.72
y2	170457.94	3361045.89	0.90	218484848.00
y3	234260.34	2714990.27	13.49	156106465.00
y4	321655.97	6029899.35	782.30	391690495.00
y5	12813.80	233588.92	0.00	13121122.20

NOTE: Dollar quantities are measured in 1000s of 1996 dollars.

TABLE 2
Summary Statistics for Contemporaneous Efficiency Estimates

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FDH, 1984:					
1.000	1.000	1.000	1.000	1.000	1.000
FDH, 1983:					
1.000	1.000	1.000	1.000	1.000	1.000
FDH, 2002:					
1.000	1.000	1.000	1.000	1.000	1.000
DEA, 1984:					
0.525	0.817	0.873	0.868	0.926	1.000
DEA, 1983:					
0.794	0.925	0.948	0.947	0.971	1.000
DEA, 2002:					
0.859	0.966	0.975	0.975	0.985	1.000

TABLE 3
Contemporaneous Order-m Efficiency Estimates
for Median Banks, $m = 150$

Decile	$\widehat{D}_{m,n_t}(\boldsymbol{x}, \boldsymbol{y})$	— 95% CI —	
1984:			
1	1.294	1.139	1.376
2	1.404	1.314	1.452
3	1.399	1.320	1.464
4	1.503	1.445	1.555
5	1.541	1.468	1.609
6	1.559	1.500	1.608
7	1.584	1.537	1.643
8	1.603	1.558	1.660
9	1.693	1.590	1.728
10	2.243	2.102	2.311
1993:			
1	1.311	1.221	1.328
2	1.242	1.051	1.319
3	1.413	1.317	1.467
4	1.434	1.321	1.473
5	1.450	1.415	1.531
6	1.472	1.382	1.501
7	1.503	1.440	1.557
8	1.599	1.542	1.660
9	1.778	1.700	1.861
10	2.448	2.291	2.594
2002:			
1	1.574	1.415	1.590
2	1.357	1.199	1.470
3	1.403	1.305	1.486
4	1.499	1.395	1.593
5	1.499	1.426	1.575
6	1.584	1.517	1.651
7	1.578	1.519	1.653
8	1.687	1.600	1.761
9	1.755	1.652	1.821
10	2.377	2.235	2.545

TABLE 4

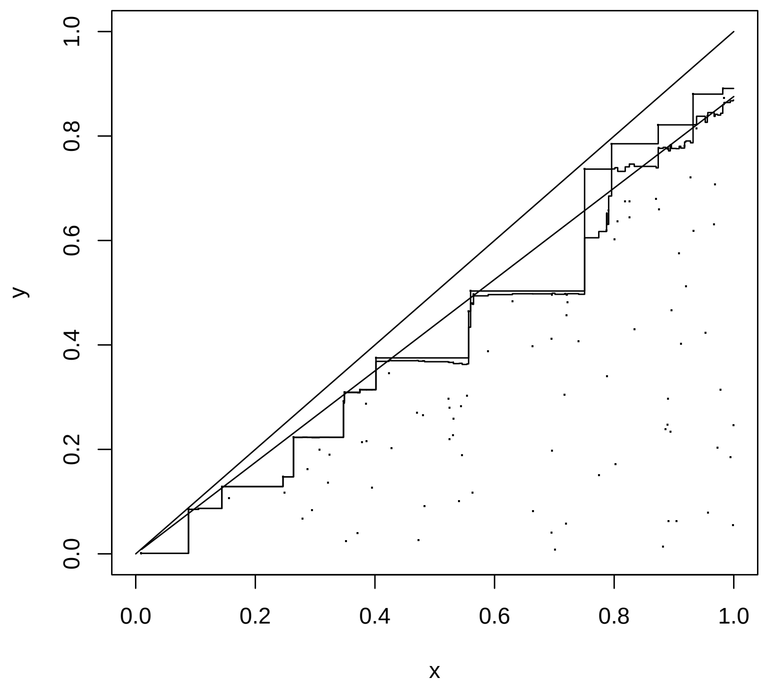
Estimates of Changes in Order-m Technical Efficiency (ΔM_Eff) and Order-m Scale Efficiency (ΔM_SEff) for Median Banks ($m = 150$)

Decile	$\widehat{\mathcal{M}}_{m,n_{t_1},n_{t_2}}$	$\Delta\widehat{M\_Eff}$	$\Delta\widehat{M\_SEff}$	$\Delta\widehat{M\_Fron}$	$\Delta\widehat{M\_SFron}$
1984–1993:					
1	0.994	1.013	1.015	0.972	0.995
2	1.007***	0.884***	1.173***	1.117***	0.869***
3	1.021***	1.011	1.038	1.036	0.941**
4	1.001	0.954**	1.077***	1.020	0.955*
5	1.003	0.941	1.089**	1.011	0.968
6	1.003	0.944***	1.082***	1.016	0.966
7	0.996	0.949**	1.070***	1.034*	0.950**
8	1.007	0.998	1.028	1.008	0.975
9	0.984**	1.050**	0.953*	1.008	0.976
10	0.987	1.091***	0.919**	1.026	0.959*
1993–2002:					
1	1.037***	1.216***	0.875*	0.895	1.089
2	1.032**	0.966	1.085	1.026	0.960
3	1.031*	1.003	1.028	1.066**	0.937**
4	1.072***	0.998	1.048	0.973	1.054
5	1.087***	0.973	1.057	1.046	1.010
6	1.131***	1.016	1.021	0.986	1.106***
7	1.157***	0.996	1.029	1.071***	1.054
8	1.197***	1.052	0.997	1.041**	1.097***
9	1.174***	1.036	0.986	1.072***	1.071
10	1.187***	1.060*	0.984	1.093*	1.041
1984–2002:					
1	1.032***	1.200***	0.862***	0.817**	1.220**
2	1.008***	1.092	0.925	1.014	0.984
3	0.982***	0.993	0.991	0.957	1.043
4	1.014*	1.046	0.973	0.977	1.020
5	1.000	1.034	0.971	0.964	1.034
6	1.010	1.076***	0.944**	0.960*	1.037
7	1.004	1.050*	0.962	0.973	1.022
8	1.017	1.055	0.970	0.968	1.027
9	1.015	0.987	1.035	0.975	1.018
10	1.032	0.971	1.071	0.968	1.025

NOTE: one, two, or three asterisks indicates the difference between an estimate and one (which would indicate no change) is statistically significant at .10, .05, or .01, respectively.

Figure 1
Example with Uniform DGP, $p = q = 1$, $m = 50$

$n = 100$



$n = 1000$

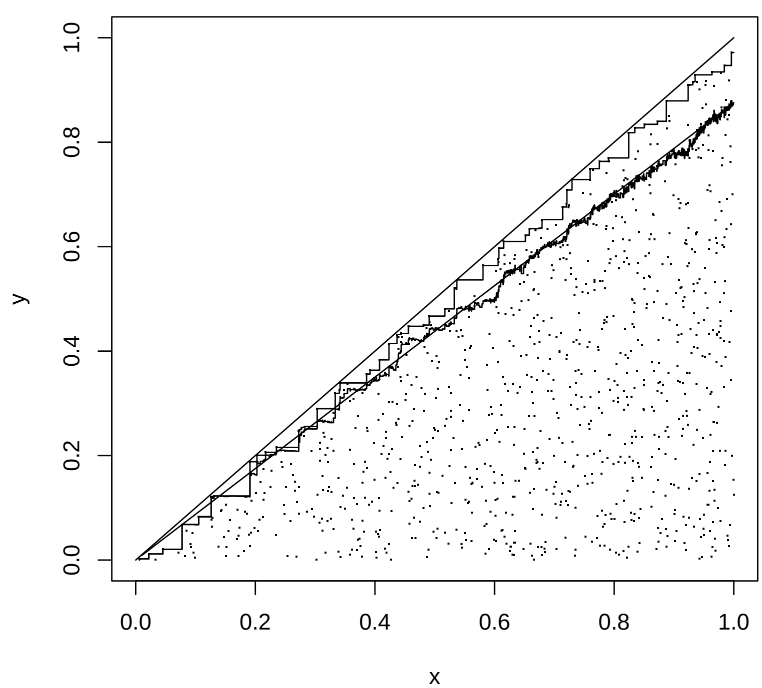


Figure 2
Example with Uniform DGP, $p = q = 1$, $m \in \{50, 150, 1500\}$

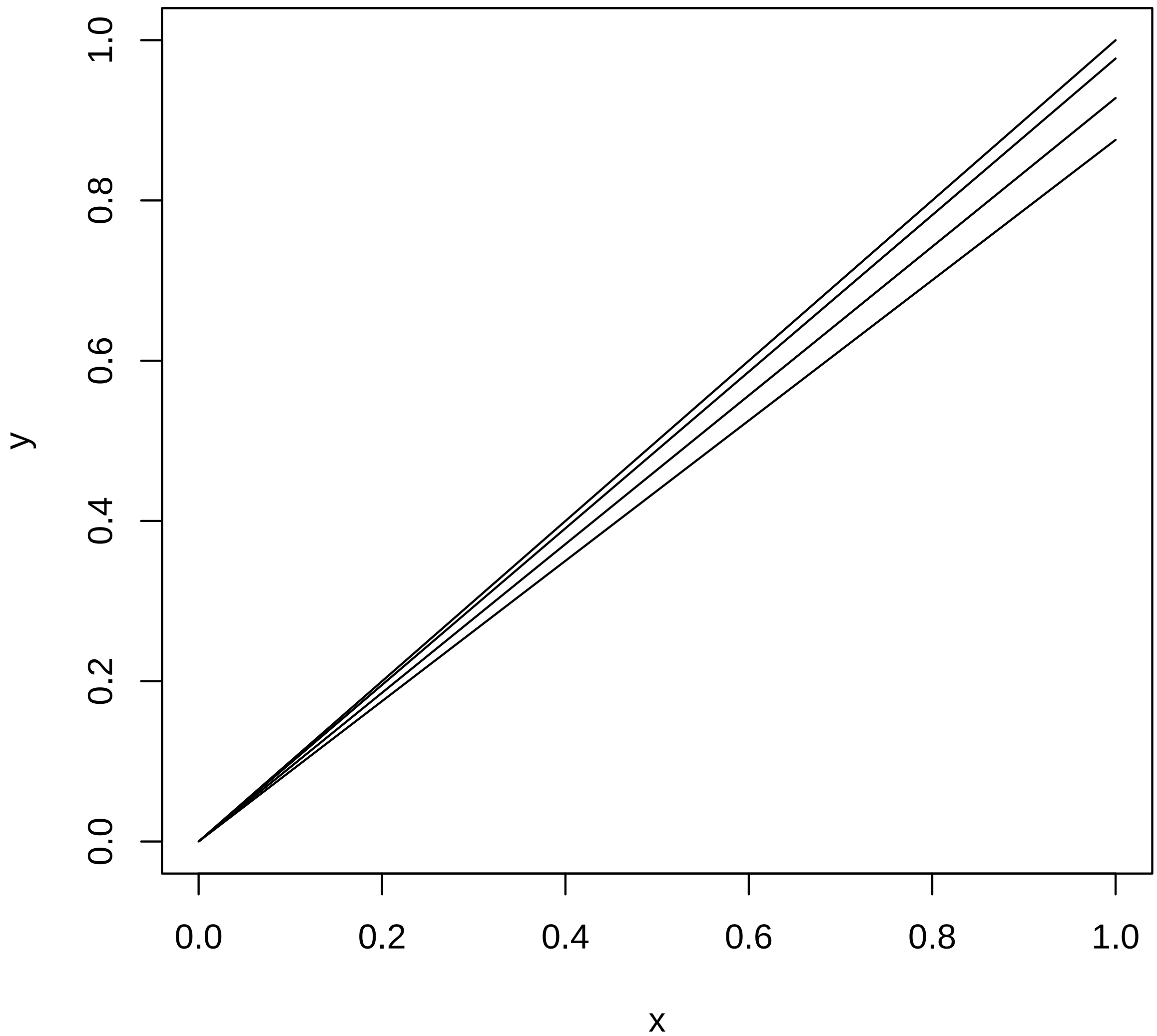
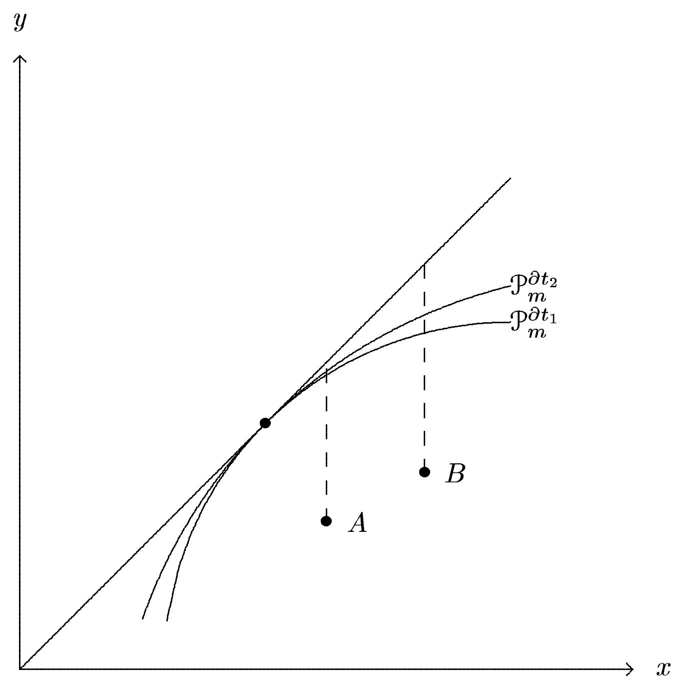


Figure 3
Changes in Scale of the Order-m Frontier

— A —



— B —

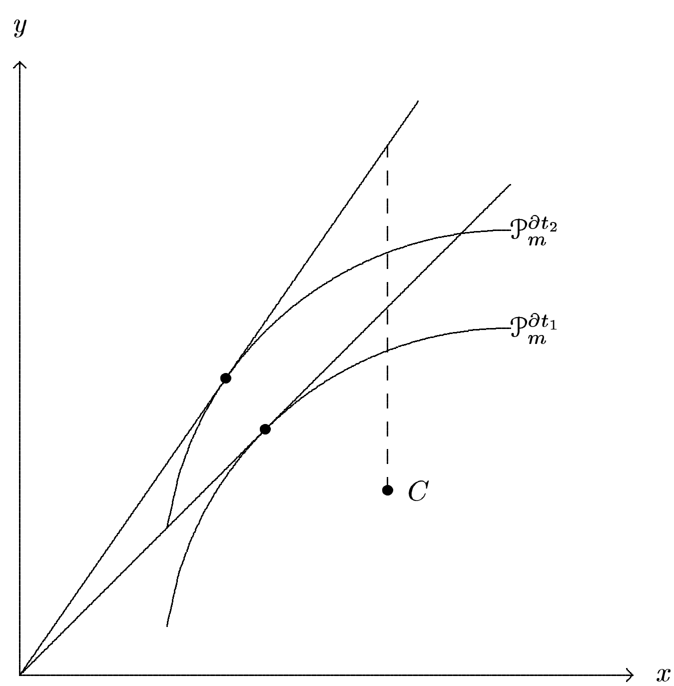


Figure 4
Kernel Estimates of Density of Order-m Efficiency
Estimates for 2002, by Quintile, with $m = 75$, 150, 300, and 1500