Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

The Evolution of Cost-Productivity and Efficiency Among U.S. Credit Unions

David C. Wheelock
and
Paul W. Wilson

Working Paper 2009-008D
http://research.stlouisfed.org/wp/2009/2009-008.pdf

March 2009
Revised January 2012

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The Evolution of Cost-Productivity and Efficiency Among U.S. Credit Unions

David C. Wheelock Paul W. Wilson*

January 2012

Abstract

Advances in information-processing technology have significantly eroded the advantages of small scale and proximity to customers that traditionally enabled community banks and other small-scale lenders to thrive. Nonetheless, U.S. credit unions have experienced increasing membership and market share, though consolidation has reduced the number of credit unions and increased their average size. We investigate the evolution of the efficiency and productivity of U.S. credit unions between 1989 and 2006 using a new methodology that benchmarks the performance of individual firms against an estimated order-α quantile lying "near" the efficient frontier. We construct a cost analog of the widely-used Malmquist productivity index, and decompose the index to estimate changes in cost and scale efficiency, and changes in technology, that explain changes in cost-productivity. We find that cost-productivity fell on average across all credit unions but especially among smaller credit unions. Smaller credit unions confronted an unfavorable shift in technology that increased the minimum cost required to produce given amounts of output. In addition, all but the largest credit unions became less scale efficient over time.

JEL classification nos.: C13, C14, G21

Keywords: efficiency, productivity, quantile estimation, credit unions

*Wheelock: Research Department, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166; email david.c.wheelock@stls.frb.org. Wilson: The John E. Walker Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, SC 29634–1309; email pww@clemson.edu. Research described in this paper was conducted while Wilson was a visiting scholar in the Research Department of the Federal Reserve Bank of St. Louis. The authors thank an anonymous referee for comments on a previous version of this paper, and Craig Aubuchon for capable assistance. Any remaining errors are solely our responsibility. The views expressed in this paper do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

1 Introduction

Technological advances and changes in regulation have profoundly altered the landscape of banking in the United States and elsewhere. For example, the relaxation of restrictions on branching, both within and across state borders, precipitated a consolidated wave that has halved the number commercial banks in the United States since the mid-1980s. Over the same years, advances in information processing technologies lowered the cost of obtaining quantitative and other "hard" information about potential borrowers, and thereby reduced the advantages of small scale, close proximity and local ties that gave small, "community" banks a competitive advantage in lending to small businesses and other "informationally-opaque" borrowers (Petersen and Rajan, 2002; Berger, 2003; Bernanke, 2006). Besides promoting consolidation among banks, regulatory and technological changes have spurred growth in the size of banks (Berger et al., 1999). Large banks have tended to be more profitable than small banks in recent years, and exhibit larger increases in productivity and efficiency (Wheelock and Wilson, 2009).

Credit unions, like community banks, traditionally have served small retail customers. Credit unions are mutual organizations that provide deposit, lending, and other services to a membership defined by an occupational, fraternal, or other common bond. A common bond is advantageous because it can reduce the cost of assessing the credit-worthiness of potential borrowers and thereby facilitate unsecured lending on reasonable terms to a credit union's members. The advances in information technology that have eroded the advantages of close customer relationships in business lending, however, have likely also eroded the advantages of small scale and common bond that traditionally have enabled credit unions to provide financial services to their members at low cost (Walter, 2006). Thus far, credit unions seem to be adapting to the new environment. Since 1985, the share of U.S. depository institution assets held by credit unions has nearly doubled, from 3.3 percent to 6.0 percent, and credit union membership has increased faster than U.S. population, from 52 million members in 1985 to 93 million members in 2009. The Credit Union Membership Access Act of 1998 may have facilitated the increase in membership by affirming the right of credit unions to accept members from unrelated groups. Since then, the number and size of credit unions characterized by multiple common bonds has increased rapidly (Walter, 2006). Credit unions now hold about 10 percent of U.S. household savings deposits, 9 percent of all consumer loans, and 13.2 percent of non-revolving consumer loans. Credit unions are also increasingly a source of business loans, and legislation pending in Congress would permit credit unions to offer even more business loans by increasing the cap for such loans from 12.25 percent of a credit union's total assets to 25 percent.[1]

As with commercial banks, the evolving competitive environment appears to favor larger credit unions, which have tended to grow more rapidly than smaller credit unions (Goddard et al., 2002). Between 1985 and 2006, the average, inflation-adjusted total assets of U.S. credit unions increased by more than 600 percent. U.S. credit unions held an average of \$84.6 million of assets in 2006 (\$50.6 million in constant 1985 dollars) versus \$7.8 million in 1985. Consolidation has also sharply reduced the total number of credit unions from a peak of 23,866 in 1969 to just 8,662 in 2006. Further increases in scale seem likely because even the largest credit unions appear to operate under increasing returns to scale (Wheelock and Wilson, 2011). It remains an open question, however, whether credit unions, as a group, will continue to gain market share. Much of their recent increase in market share has come at the expense of savings and loan associations and savings banks, which saw a decline in market share from 30.1 percent to 15.9 percent between 1985 and 2006. By contrast, the share of industry assets held by commercial banks rose from 66.1 percent to 78.1 percent over the same years. Credit unions are likely to continue to fill a niche, but as an industry may not thrive unless they can exploit new technologies to become more productive and scale efficient.

This paper investigates productivity growth among U.S. credit unions to assess how successfully credit unions have contained costs while fulfilling the desire of their members for favorable terms on loans and deposits. In this framework, we examine changes in cost-productivity, i.e., the extent to which the cost of producing given levels of output has changed over time. Credit unions become more cost-productive if the cost they incur to produce given levels of outputs declines over time or, equivalently, if the levels of outputs they produce for a given level of cost rises. We also estimate changes in cost and scale efficiency for credit unions.

[1] H.R. 3380, the Promoting Lending to America's Small Business Act was introduced in Congress during July 2009 by Representative Paul Kanjorski. S. 2919, which would amend the Federal Credit Union Act, was introduced by Sen. Mark Udall on December 21, 2009. Data on credit union membership, deposits and loans are available from the Credit Union National Association: http//www.cuna.org/.

Credit unions become more cost efficient if they move closer to the efficient frontier, and more scale efficient if they move closer to a region of the underlying technology characterized by constant returns to scale. A credit union could become more cost or scale efficient without becoming more cost-productive as a result of an unfavorable shift in the technology that increases the minimum feasible cost of producing given levels of outputs.

We specify a cost relationship for credit unions that takes account of the unique objectives of the owners of mutually-owned depository institutions for high deposit interest rates and low loan interest rates. We estimate the cost relationship non-parametrically using a suitably adapted version of the "order-α quantile" frontier estimators developed by Daouia (2003), Daouia and Simar (2007), and Wheelock and Wilson (2008). By using a nonparametric estimator, we avoid the problem of specifying and estimating a potentially incorrect parametric cost function.[2] Further, unlike traditional nonparametric estimators, such as data envelopment analysis (DEA), our nonparametric order-α quantile estimator has a relatively rapid, root-n convergence rate (similar to parametric estimators) and is robust to data outliers.[3]

We construct the cost analog of the familiar Malmquist productivity index, defined in terms of our nonparametric estimator, and decompose the index to allocate changes in cost-productivity to changes in cost efficiency, technology and scale efficiency. In addition, we decompose a residual term to gain insight into the sources of changes in scale efficiency. Our results indicate that, in general, credit unions became less cost-productive between 1989 and 2006, indicating that they incurred higher (inflation-adjusted) operating costs to produce given levels of output in 2006 than in 1989. We also find that smaller credit unions tended to experience larger declines in cost-productivity than large credit unions. Small credit unions appear to have faced a shift in the cost frontier that increased the minimum cost of producing given amounts of output. Although small credit unions, on average, became more cost efficient over time, they also became less scale efficient. By contrast, the largest credit unions became marginally less cost efficient on average, but somewhat more scale efficient. Thus, our results are consistent with the conjecture that recent advances in technology and

[2] Many studies have found that even relatively flexible functional forms, such as the translog function, are mis-specifications of cost relationships for banks and other depository institutions (e.g., McAllister and McManus, 1993; Wheelock and Wilson, 2001; Wheelock and Wilson, 2011).

[3] The root-n convergence rate obtains only if the estimator is used to estimate a partial frontier lying close to the full frontier, which is the approach we take here.

changes in regulation have favored larger credit unions.

The rest of the paper unfolds as follows: Section 2 discusses recent literature on credit union performance. Section 3 describes the variables in a credit union cost relationship and presents our statistical model for estimation method. Section 4 defines measures of changes in cost-productivity, efficiency, etc., and Section 5 reports and discusses the estimation results. The final section presents our conclusions.

2 Literature Review

The performance of U.S. credit unions has been evaluated on several dimensions. Most studies assume that credit unions seek to minimize operating cost while maximizing member benefits in terms of the prices or variety of services they offer.[4] Fried et al. (1993), for example, estimate the productive efficiency of credit unions in the context of a model in which credit unions seek to maximize member benefits in terms of the price, quantity and variety of services offered to members subject to resource availability and the operating environment. The study employs a nonparametric free disposal hull (FDH) estimator and data from 1990, and obtains a mean inefficiency estimate of 9.2 percent. That is, they find that, on average, credit unions are capable of producing 9.2 percent more service with the amounts of variable resources available. Notably, the study also finds that larger credit unions, measured in terms of total assets, are more efficient than small credit unions.

Frame et al. (2003) also examine efficiency in the context of a model that assumes that credit unions seek to minimize non-interest costs subject to input prices, the level and types of output they produce, and the prevailing production technology. Based on estimation of a parametric translog cost function using data from 1998 for credit unions with more than $50 million of total assets, Frame et al. (2003) find significant differences in the performance of large credit unions with different types of common bonds. Specifically, they find that credit unions with residential common bonds have higher costs than those with occupational or associational bonds.

Studies have also examined the effects of mergers on credit union performance. For

[4] See Smith et al. (1981), Smith (1984), Fried et al. (1993), Fried et al. (1999), Frame et al. (2003), and Bauer (2008). A few studies have found some evidence of agency problems at credit unions to the detriment of their members (Emmons and Schmid, 1999b; Frame et al., 2003; and Leggett and Strand, 2002). However, we make no attempt here to distinguish between the interests of credit union managers and members.

example, Fried et al. (1999) investigate the impact of mergers on credit union efficiency in the context of a model in which credit unions seek to minimize cost while maximizing the services provided to members. The study uses data envelopment analysis (DEA) to estimate efficiency relative to a "member service performance" frontier, and finds that, on average, credit unions that engage in acquisitions are more efficient than those that are acquired. Further, the study finds that members suffer no deterioration in service when their credit union acquires another credit union, whereas members of acquired credit unions tend to experience improved service.

Bauer et al. (2009) also examine the impact of mergers on credit union performance, using the event study methodology for detecting changes in credit union performance of Bauer (2008). Based on data for 1994–2004, the study finds that members of acquired credit unions benefit from higher deposit interest rates and lower loan interest rates compared with expected rates based on pre-merger information. Further, the study finds that the interest rates offered by credit unions that make acquisitions are not significantly affected by mergers.

Several studies investigate the relationship between credit union costs and firm size. Many find that average operating expenses decline as credit unions become larger (Emmons and Schmid, 1999a; Leggett and Strand, 2002; Wilcox, 2006). Further, Goddard et al. (2008) find that larger credit unions are better able to diversify into non-traditional product lines, such as business loans, credit cards and mutual funds, and that doing so reduces the volatility of their earnings.

Wheelock and Wilson (2011) evaluate alternative measures of returns to scale for credit unions using a cost relationship that takes account of the benefit of high deposit interest rates and low loan interest rates to credit union members, similar to the models of Fried et al. (1993), Fried et al. (1999), and Frame et al. (2003). Wheelock and Wilson (2011) find that nearly all U.S. credit unions operate under increasing returns to scale, which seems consistent with the faster average growth rates of total assets, membership and earnings observed among larger U.S. credit unions (Goddard et al., 2002).[5]

[5] Wheelock and Wilson (2011) estimate returns to scale both along a ray from the origin through the median observed vector of credit union outputs ("ray-scale economies") and along rays from the origin through the observed output vector for each credit union ("expansion-path economies"). The study finds evidence of rapidly increasing ray-scale economies below the median total assets of U.S. credit unions, but near constant returns for larger credit unions. However, estimates of expansion-path scale economies, which may better reflect scale economies near the combinations of inputs and outputs in actual credit union

Although prior research has found some evidence that larger credit unions are more efficient and have lower average costs than small credit unions, studies have not investigated changes in credit union performance over time except to compare performance before and after mergers. Here we investigate changes in the cost-productivity and efficiency (both cost and scale efficiency) of credit unions. We focus particularly on whether changes in performance varied systematically across credit unions of different sizes, as predicted by studies discussing the effects of recent changes in regulation and information technology.

3 Variable Specification and Statistical Model

3.1 Model

Credit unions are mutual organizations that make loans, accept deposits, and provide other services to their members. Whereas commercial banks attempt to maximize the spread between interest paid on deposits and interest charged for loans, credit unions presumably attempt to provide favorable terms on both loans and deposits to their members. Following previous studies (e.g., Smith et al., 1981; Smith, 1984; Fried et al., 1993; Fried et al., 1999; Frame et al., 2003; Bauer, 2008), we model credit unions as service providers that seek to minimize non-interest costs ($COST$) subject to the prices of labor and capital input, the prevailing production technology, and the level and types of output they produce. We specify two variable output quantities, namely total loans ($LOANS$) (the sum of real estate loans, commercial loans, and consumer loans) and other investments ($INVEST$). In addition, we specify two quasi-fixed outputs that reflect benefits to members: savings pricing ($PRSAV$) and loan pricing ($PRLOAN$). Following Frame et al. (2003), we specify the price dimension of service to credit union members as the average interest rates on deposits and loans. Also like Frame et al. (2003), our model includes as inputs financial capital (CAP) and labor (LAB) and the corresponding input-prices ($WCAP$ and $WLAB$) faced by each credit union; variable cost ($COST$) equals $(WCAP \times CAP) + (WLAB \times LAB)$. Table 1 lists the variables in our model and reports how each is defined in terms of call report items.[6]

production, indicate that even the largest credit unions operate under increasing returns to scale.

[6] Call report data for individual credit unions are available from the National Credit Union Administration (http://www.ncua.gov).

We define

$$\boldsymbol{x} = \begin{bmatrix} CAP & LAB \end{bmatrix}', \tag{3.1}$$

$$\boldsymbol{w} = \begin{bmatrix} WCAP & WLAB \end{bmatrix}', \tag{3.2}$$

$$\boldsymbol{y}_1 = \begin{bmatrix} LOANS & INVEST \end{bmatrix}', \tag{3.3}$$

and

$$\boldsymbol{y}_2 = \begin{bmatrix} PRSAV & 1/PRLOAN \end{bmatrix}', \tag{3.4}$$

where $\boldsymbol{x}$ is a vector of production inputs, $\boldsymbol{w}$ is the corresponding vector of input prices, $\boldsymbol{y}_1$ is a vector of variable outputs, and $\boldsymbol{y}_2$ is a vector of quasi-fixed outputs. Using the reciprocal of the loan pricing variable $PRLOAN$ in (3.4) maintains increasing costs with respect to output quantities.

We specify a cost relationship in terms of the output quantities and input prices defined above. We estimate the relationship non-parametrically using a modified version of the order-α quantile frontier estimators of Daouia (2003), Aragon et al. (2005), Daouia and Simar (2007), and Wheelock and Wilson (2008). Nonparametric estimation avoids the problem of specification error. As noted previously, even fairly flexible function forms, such as the translog function, have been found to mis-specify cost relationships for banks and credit unions (e.g., Wheelock and Wilson, 2011). Of course, any efficiency estimates based on estimation of a mis-specified model would be suspect.

Unlike DEA and similar nonparametric frontier estimators, which measure a firm's performance relative to an estimate of the efficient frontier, order-α quantile estimators measure efficiency in terms of a quantile lying "near" the efficient frontier. The advantage of using order-α quantile estimators is that, unlike DEA and other traditional nonparametric frontier estimators, order-α quantile estimators are both robust to data outliers and have the rapid root-n convergence rate of linear parametric estimators.

Because order-α quantile estimators are fairly new and have been used less frequently than DEA and similar nonparametric methods, we next describe in some detail our statistical model and distance function measures.

3.2 Inefficiency Measurement

Given vectors $\boldsymbol{x} \in \mathbb{R}_+^p$ of p input quantities and $\boldsymbol{y} \in \mathbb{R}_+^q$ of q output quantities, standard microeconomic theory of the firm posits a production set at time t represented by

$$\mathcal{P}^t \equiv \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y} \text{ at time } t\}. \tag{3.5}$$

This set represents the set of feasible combinations of inputs and outputs at a given point in time, and may change with the passage of time. We assume throughout that the production set $\mathcal{P}^t$ is free-disposal, i.e., if $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t$, then $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \in \mathcal{P}^t \ \forall\ \widetilde{\boldsymbol{x}} \geq \boldsymbol{x}$ and $\mathbf{0} \leq \widetilde{\boldsymbol{y}} \leq \boldsymbol{y}$. We also assume throughout that all production requires the use of strictly positive levels of some inputs; i.e., $(\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}^t$ if $\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{y} \geq \mathbf{0},\ \boldsymbol{y} \neq \mathbf{0}$. These assumptions are standard in microeconomic theory; e.g., see Shephard (1970) or Färe (1988).[7]

It is often assumed in addition that $\mathcal{P}$ is closed. Then the upper boundary of $\mathcal{P}^t$, denoted $\mathcal{P}^{\partial^t}$, is referred to as the *technology* or *production frontier*, and is defined by

$$\mathcal{P}^{\partial^t} = \{(\boldsymbol{x}, \boldsymbol{y}) \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t,\ (\gamma^{-1}\boldsymbol{x}, \gamma\boldsymbol{y}) \notin \mathcal{P}\ \forall\ \gamma > 1\}. \tag{3.6}$$

Free disposability of the production set $\mathcal{P}^t$ implies monotonicity of the frontier $\mathcal{P}^{\partial^t}$.

Firms face input prices $\boldsymbol{w} \in \mathbb{R}_{++}^p$ corresponding to the inputs represented in $\boldsymbol{x}$. For a given input-price vector $\boldsymbol{w}$,

$$\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t) = \{(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}) \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\} \tag{3.7}$$

is the set of feasible combinations of cost and outputs. Note that $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t) \subset \mathbb{R}_+^{q+1}$. The set $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$ is completely determined by $\mathcal{P}^t$ and the given input price vector $\boldsymbol{w}$. Free disposability of $\mathcal{P}^t$ implies a type of free disposability for $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$; i.e., if $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$, then $(\boldsymbol{w}'\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)\ \forall\ \widetilde{\boldsymbol{x}} \geq \boldsymbol{x}$ and $\mathbf{0} \leq \widetilde{\boldsymbol{y}} \leq \boldsymbol{y}$. If $\mathcal{P}^t$ is assumed closed and convex, then $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$ must also be closed and convex. If $\mathcal{P}^t$ is closed, then we can replace $\mathcal{P}^t$ in (3.7) with $\mathcal{P}^{\partial^t}$ to obtain the cost frontier

$$\mathcal{C}^{\partial}(\boldsymbol{w}, \mathcal{P}^t) = \mathcal{C}(\boldsymbol{w}, \mathcal{P}^{\partial^t}) \subset \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t), \tag{3.8}$$

which forms the lower boundary of the set $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$.

[7] Throughout, we define inequalities involving vectors on an element-by-element basis; e.g., for $\widetilde{\boldsymbol{x}},\ \boldsymbol{x} \in \mathbb{R}_+^p$, $\widetilde{\boldsymbol{x}} \geq \boldsymbol{x}$ means that some number $\ell \in \{0,\ 1,\ \ldots,\ p\}$ of the corresponding elements of $\widetilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ are equal, while $(p - \ell)$ of the elements of $\widetilde{\boldsymbol{x}}$ are greater than the corresponding elements of $\boldsymbol{x}$.

The minimum cost of producing output levels $\boldsymbol{y}$ from quantities of inputs $\boldsymbol{x}$ at prices $\boldsymbol{w} \in \mathbb{R}^p_{++}$ is given by the cost function

$$C(\boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) = \inf_{\boldsymbol{x} \geq 0} \{\boldsymbol{w}'\boldsymbol{x} \mid \boldsymbol{w},\ (\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)\}. \tag{3.9}$$

A standard measure of cost efficiency for a firm facing input prices $\boldsymbol{w}$ and using input quantities $\boldsymbol{x}$ to produce output quantities $\boldsymbol{y}$ with $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t$ is

$$\begin{aligned} \rho(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) &= \frac{\boldsymbol{w}'\boldsymbol{x}}{C(\boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t))} \\ &= \sup \left\{\rho > 0 \mid (\rho^{-1}\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)\right\}. \end{aligned} \tag{3.10}$$

Färe et al. (1985) refer to the inverse of $\rho(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{w}, \mathcal{P}^t)$ as "overall" efficiency.

The standard measure of cost inefficiency defined in (3.10) holds output levels fixed. However, just as technical efficiency can be measured in various directions, cost efficiency can also be measured in different directions. For example, one might use

$$\tau(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) = \sup \left\{\tau > 0 \mid (\boldsymbol{w}'\boldsymbol{x}, \tau\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)\right\} \tag{3.11}$$

or

$$\kappa(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) = \sup \left\{\kappa > 0 \mid (\kappa^{-1}\boldsymbol{w}'\boldsymbol{x}, \kappa\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)\right\} \tag{3.12}$$

to measure cost efficiency. These measures are illustrated in Figure 1 for a given, constant input-price vector $\boldsymbol{w}$ and a single output quantity. The curve passing through points B, D, and E is the (minimum) cost function $C(\boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t))$ defined in (3.9). For the cost-inefficient firm operating at point A, the standard measure of cost efficiency defined in (3.10) gives a cost inefficiency measure $\rho = \frac{AC}{AB} > 1$. The output cost efficiency measure defined in (3.11) gives $\tau = \frac{FE}{FA} > 1$, while the hyperbolic cost efficiency measure defined in (3.12) yields $\kappa = \frac{AC}{AG} = \frac{FH}{FA}$. Here, τ is the feasible proportion by which output could be expanded while holding costs constant, whereas κ is the feasible equi-proportionate reduction in costs and simultaneous increase in output.

The hyperbolic measure defined in (3.12) is the cost analog of the hyperbolic technical efficiency measures considered by Färe et al. (1985), Wheelock and Wilson (2008), and Wilson (2009). The advantages of the hyperbolic cost efficiency measure in (3.12) are similar to those of its hyperbolic technical efficiency counterpart; see Wilson (2011) for discussion.

Our model of credit union activities described in Section 3.1 treats some credit union outputs as variable and some as quasi-fixed. Hence, we partition the output vector $\boldsymbol{y}$ by writing

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1' & \boldsymbol{y}_2' \end{bmatrix}' \tag{3.13}$$

where $\boldsymbol{y}_1 \in \mathbb{R}_+^{q-r}$, $\boldsymbol{y}_2 \in \mathbb{R}_+^{r}$, and $r \in \{0,\ 1,\ \ldots,\ q\}$ is the number of quasi-fixed outputs, $r \leq q$. Now consider a credit union facing input prices $\boldsymbol{w}$ and operating at $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^t$. In order to deal with quasi-fixed outputs while retaining some of the flavor of the hyperbolic distance measure defined in (3.12), we can define the *conditional* distance function

$$\gamma(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) \equiv \sup\left\{\gamma > 0 \mid (\gamma^{-1}\boldsymbol{w}'\boldsymbol{x}, \gamma\boldsymbol{y}_1, \boldsymbol{y}_2) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)\right\} \tag{3.14}$$

to measure cost efficiency. The measure $\gamma(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t))$ gives the feasible, simultaneous, proportionate reduction in cost $\boldsymbol{w}'\boldsymbol{x}$ and increase in variable outputs $\boldsymbol{y}_1$ holding quasi-fixed outputs $\boldsymbol{y}_2$ constant.[8]

The various measures that have been introduced so far are unobserved, and must be estimated from a random sample $\mathcal{S}_n^t = \{\boldsymbol{X}_i, \boldsymbol{Y}_i\}_{i=1}^n$ of input-output vectors at time t. We assume $\Pr((\boldsymbol{X}_i, \boldsymbol{Y}_i) \in \mathcal{P}^t) = 1$ for each $(\boldsymbol{X}_i, \boldsymbol{Y}_i) \in \mathcal{S}_n^t$. It is well-known that $\mathcal{P}^t$, and hence $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$, can be estimated by the free disposal hull (FDH) estimator proposed by Deprins et al. (1984) or the data envelopment analysis (DEA) estimator proposed by Farrell (1957). These estimators envelop all of the sample observations; consequently, the corresponding efficiency estimators are sensitive to outliers or extreme values in the data. In addition, both FDH and DEA estimators of $\mathcal{P}^t$, as well as the corresponding efficiency estimators, suffer from the well-known curse of dimensionality.[9] Two alternatives to FDH and DEA estimators have been developed. Cazals et al. (2002) introduced the notion of order-m

[8] Unlike the distance function defined in (3.12), in which all input prices and output quantities are variable, the distance function in (3.14) holds $\boldsymbol{y}_2$ fixed. This distance function is similar to the directional distance function introduced by Chambers et al. (1996, 1998), except that efficiency is measured along a hyperbolic path in the $(q+1)$-dimensional subspace spanned by $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}_1)$ instead of along a linear path as in the case of directional distance functions.

[9] See Park et al. (2000) for assumptions required for consistency of the corresponding FDH efficiency estimator and its asymptotic properties. See Kneip et al. (2008) and Park et al. (2010) for assumptions required for consistency of the corresponding DEA efficiency estimator and its asymptotic properties under variable returns to scale and constant returns to scale. Park et al. (2000) establish a convergence rate of $n^{-1/(p+q)}$ for the FDH efficiency estimator, while Kneip et al. (2008) establish a rate of $n^{-2/(p+q+1)}$ under variable returns to scale. Under constant returns to scale, the convex-cone version of the DEA estimator is shown by Park et al. (2010) to converge at rate $n^{-2/(p+q)}$.

partial frontiers, while Daouia (2003), Aragon et al. (2005), and Daouia and Simar (2007) introduced the concept of order-α *partial* frontiers that envelop most, but not all, sample observations and consequently avoid the extreme sensitivity to outliers encountered with FDH and DEA estimators. Aragon et al. (2005) and Daouia and Ruiz-Gazen (2006) show that partial frontiers based on α-quantile estimators have robustness properties superior to those of partial frontiers based on order-m estimators.

In order to adapt the order-α idea to our setting, we use the probabilistic framework introduced by Cazals et al. (2002) and extended by Daraio and Simar (2005b). In particular, for random input-output vectors $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{S}_n^t$, we posit a joint probability measure leading to the distribution function

$$H^t(\boldsymbol{x}, \boldsymbol{y}) = \Pr(\boldsymbol{X} \leq \boldsymbol{x}, \boldsymbol{Y} \geq \boldsymbol{y}) \tag{3.15}$$

at time t. As noted by Daouia and Simar (2007), this distribution function completely characterizes the data-generating process. Given the assumption $\Pr((\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{P}^t) = 1$ introduced earlier, $H^t(\boldsymbol{x}, \boldsymbol{y})$ has support over $\mathcal{D}^t \subseteq \mathcal{P}^t$. As noted by Kneip et al. (2008), in most situations $\mathcal{D}^t$ will equal $\mathcal{P}^t$, but we allow for the possibility that $\mathcal{D}^t$ is a strict subset of $\mathcal{P}^t$.[10]

Given a vector of input prices $\boldsymbol{w}$, H^t induces a distribution function

$$G^t(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}) = \Pr(\boldsymbol{w}'\boldsymbol{X} \leq \boldsymbol{w}'\boldsymbol{x}, \boldsymbol{Y} \geq \boldsymbol{y}) \tag{3.16}$$

with support over the set $\mathcal{K}^t = \{(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}^t\} \subseteq \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$. Clearly, H^t is monotone, non-decreasing in $\boldsymbol{x}$ and monotone, non-increasing in $\boldsymbol{y}$; therefore, G^t is monotone, non-decreasing in $\boldsymbol{w}'\boldsymbol{x}$ and monotone, non-increasing in $\boldsymbol{y}$. The domain of $G^t(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y})$ has $(q+1)$ dimensions, whereas the domain of $H^t(\boldsymbol{x}, \boldsymbol{y})$ has $(p+q)$ dimensions. Although the α-quantile estimator we use is root-n consistent, decreasing the number of dimensions should reduce its variance.

The distribution function $G^t(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y})$ defined in (3.16) can be decomposed by writing

$$G^t(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}) = \underbrace{\Pr\left(\boldsymbol{w}'\boldsymbol{X} \leq \boldsymbol{w}'\boldsymbol{x},\ \boldsymbol{Y}_1 \geq \boldsymbol{y}_1 \mid \boldsymbol{Y}_2 \geq \boldsymbol{y}_2\right)}_{=G^t(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}_1 \mid \boldsymbol{y}_2)} \underbrace{\Pr\left(\boldsymbol{Y}_2 \geq \boldsymbol{y}_2\right)}_{=S^t(\boldsymbol{y}_2)}, \tag{3.17}$$

[10] In particular, the proof given by Park et al. (2000) of consistency of the FDH efficiency estimator requires that $\mathcal{D}^t$ be a compact set, implying $\mathcal{D}^t \subset \mathcal{P}^t$. For similar reasons, compactness of $\mathcal{D}^t$ is also needed to link order-α estimators to FDH estimators (as $\alpha \to 1$) as in Daouia and Simar (2007) and Wheelock and Wilson (2008).

where $\boldsymbol{Y}$ has been partitioned into components $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ as in (3.13). The first term on the right-hand side of (3.17) is a conditional distribution function, while the second term is the joint survivor function for the quasi-fixed outputs $\boldsymbol{y}_2$. Then for all $\boldsymbol{y}$ such that $S^t(\boldsymbol{y}_2) > 0$, (3.14) can be rewritten as

$$\gamma(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) = \sup\left\{\gamma > 0 \mid G^t(\gamma^{-1}\boldsymbol{w}'\boldsymbol{x}, \begin{bmatrix}\gamma\boldsymbol{y}_1 & \boldsymbol{y}_2\end{bmatrix}) > 0\right\}. \tag{3.18}$$

since $G^t(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y})$ has bounded support on $\mathcal{K}^t \subseteq \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$.[11] More importantly, for all $\boldsymbol{y}$ such that $S^t(\boldsymbol{y}_2) > 0$ and for $\alpha \in (0, 1]$, a *conditional* α-quantile hyperbolic distance function is defined by

$$\gamma_\alpha(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) \equiv \sup\left\{\gamma > 0 \mid G^t(\gamma^{-1}\boldsymbol{w}'\boldsymbol{x}, \gamma\boldsymbol{y}_1 \mid \boldsymbol{y}_2) > (1-\alpha)\right\}. \tag{3.19}$$

This distance function blends features of the conditional input- and output-oriented order-α quantile efficiency measures proposed by Daouia (2003), Aragon et al. (2005) and Daouia and Simar (2007), and the *unconditional* hyperbolic order-α quantile distance function defined by Wheelock and Wilson (2008). For a unit operating at $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$, and for $\gamma_\alpha = \gamma_\alpha(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t))$ $(>, <)$ 1, the distance function defined by (3.19) gives the simultaneous, proportionate (decrease, increase) in cost $\boldsymbol{w}'\boldsymbol{x}$ and (increase, decrease) in variable outputs $\boldsymbol{y}_1$ (holding quasi-fixed outputs $\boldsymbol{y}_2$ constant) that would result in the firm being dominated by units incurring (weakly) less cost than $\gamma_\alpha\boldsymbol{w}'\boldsymbol{x}$ and producing (weakly) more than $\gamma_\alpha\boldsymbol{y}_1$ variable output while producing $\boldsymbol{y}_2$ quasi-fixed output with probability $(1-\alpha)$.[12] If $\alpha = 1$ and $\mathcal{P}^t$ is closed, then $(\gamma_\alpha^{-1}\boldsymbol{w}'\boldsymbol{x}, \begin{bmatrix}\gamma_\alpha\boldsymbol{y}_1 & \boldsymbol{y}_2\end{bmatrix}) \in \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$ by the free-disposability assumption.[13]

Along the lines of Daouia and Simar (2007), for all $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}_1)$ such that $G^t(\boldsymbol{w}'\boldsymbol{x}, \begin{bmatrix}\boldsymbol{y}_1 & \boldsymbol{0}\end{bmatrix}) >$

[11] Similar representations of input, output, hyperbolic, and directional distance functions have been made by Daraio and Simar (2005a), Daouia and Simar (2007), Wilson (2011), and Simar and Vanhems (2012).

[12] Deprins et al. (1984) introduced the concept of dominance in the production setting.

[13] It might be tempting to define an *unconditional* measure along the lines of Wheelock and Wilson (2008), with efficiency measured along a hyperbolic path in the subspace spanned by $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}_1)$. For example, one could write

$$\delta_\alpha(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) \equiv \sup\left\{\delta > 0 \mid G^t(\delta^{-1}\boldsymbol{w}'\boldsymbol{x}, \delta\boldsymbol{y}_1, \boldsymbol{y}_2) > (1-\alpha)\right\}$$

where again $\alpha \in (0, 1]$. However, this distance function may not exist. In particular, for some points, it is possible that $G^t(\delta^{-1}\boldsymbol{w}'\boldsymbol{x}, \delta\boldsymbol{y}_1, \boldsymbol{y}_2) \leq (1-\alpha)\ \forall\ \delta > 0$; in other words, there might be insufficient probability mass in some regions of the support of G^t to allow *any* scaling of the point $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2)$ along the path $(\delta^{-1}\boldsymbol{w}'\boldsymbol{x}, \delta\boldsymbol{y}_1, \boldsymbol{y}_2)$, $\delta > 0$, such that the probability given by $G^t(\delta^{-1}\boldsymbol{w}'\boldsymbol{x}, \delta\boldsymbol{y}_1, \boldsymbol{y}_2)$ can be made greater than $(1-\alpha)$. Consequently, we use the conditional measure defined in (3.19).

0 we can now define the conditional, hyperbolic α-quantile efficient frontier as the set

$$\begin{aligned}\mathcal{C}_\alpha^\partial(\boldsymbol{w},\mathcal{P}^t) = \big\{(\gamma_\alpha(\boldsymbol{w}'\boldsymbol{x},\boldsymbol{y}\mid\mathcal{C}(\boldsymbol{w},\mathcal{P}^t))^{-1}\boldsymbol{w}'\boldsymbol{x},\\ \big[\gamma_\alpha(\boldsymbol{w}'\boldsymbol{x},\boldsymbol{y}\mid\mathcal{C}(\boldsymbol{w},\mathcal{P}^t))\boldsymbol{y}_1 \quad \boldsymbol{y}_2\big]) \mid (\boldsymbol{w}'\boldsymbol{x},\boldsymbol{y})\in\mathcal{C}(\boldsymbol{w},\mathcal{P}^t)\big\}\,,\end{aligned} \tag{3.20}$$

In the language of Daouia and Simar (2007), $\mathcal{C}_\alpha^\partial(\boldsymbol{w},\mathcal{P}^t)$ is the set of efficient cost-output combinations at the level $(\alpha\times 100)$-percent. Points $(\boldsymbol{w}'\boldsymbol{x},\boldsymbol{y})\in\mathcal{C}_\alpha^\partial(\boldsymbol{w},\mathcal{P}^t)$ have probability $G^t(\boldsymbol{w}'\boldsymbol{x},\boldsymbol{y})\leq 1-\alpha$ of being dominated if $G^t(\cdot,\cdot)$ is continuous on $\mathcal{K}^t$.

The set $\mathcal{C}_\alpha^\partial(\boldsymbol{w},\mathcal{P}^t)$ defined in (3.20) forms the boundary of the closed set

$$\mathcal{C}_\alpha(\boldsymbol{w},\mathcal{P}^t) = \big\{(\boldsymbol{w}'\boldsymbol{x},\boldsymbol{y})\mid(\boldsymbol{w}'\boldsymbol{x},\boldsymbol{y})\in\mathcal{C}(\boldsymbol{w},\mathcal{P}^t),\ \big(\gamma^{-1}\boldsymbol{w}'\boldsymbol{x},\gamma\boldsymbol{y}_1,\boldsymbol{y}_2\big)\in\mathcal{C}_\alpha^\partial(\boldsymbol{w},\mathcal{P}^t),\ \gamma\geq 1\big\}\,. \tag{3.21}$$

This set is the order-α analog of the cost set introduced in (3.7); hence we call $\mathcal{C}_\alpha(\boldsymbol{w},\mathcal{P}^t)$ the conditional, hyperbolic cost set of order-α. Clearly, $\mathcal{C}_\alpha(\boldsymbol{w},\mathcal{P}^t)\subseteq\mathcal{C}(\boldsymbol{w},\mathcal{P}^t)$, with $\mathcal{C}_\alpha(\boldsymbol{w},\mathcal{P}^t)\subset\mathcal{C}(\boldsymbol{w},\mathcal{P}^t))\ \forall\ \alpha\in(0,1)$.

Although cost efficiency is typically measured relative to the frontier $\mathcal{C}^\partial(\boldsymbol{w},\mathcal{P}^t)$, other benchmarks, such as $\mathcal{C}_\alpha^\partial(\boldsymbol{w},\mathcal{P}^t)$, can be used. Using the hyperbolic order-α quantile $\mathcal{C}_\alpha^\partial(\boldsymbol{w},\mathcal{P}^t)$ as a benchmark offers several advantages; in particular, convergence is faster and, as explained below, inference is much simpler than when we use the cost frontier $\mathcal{C}^\partial(\boldsymbol{w},\mathcal{P}^t)$ as the benchmark.

3.3 Estimation

Since $\mathcal{P}^t$ and hence the other quantities that have been discussed here are unknown, they must be estimated from an observed sample $\mathcal{S}_n^t = \{(\boldsymbol{X}_i,\boldsymbol{Y}_i)\}_{i=1}^n$ of data on firms' input and output quantities at time t. We assume that the observations in $\mathcal{S}_n^t$ are identically, independently distributed random draws from the distribution function $H^t(\cdot,\cdot)$ introduced above in Section 3.2.

Wilson (2011) and Simar and Vanhems (2012) use simple transformations to prove consistency and other results for hyperbolic and directional distance measures. The idea is to transform the coordinate space in such a way that the problem becomes identical, in the new coordinate system, to one for which estimators with nice statistical properties have been developed. We use a similar device here.

Define a mapping $\phi\colon \mathbb{R}_+^{q+1} \to \mathbb{R}_+^{q+1}$ such that $(\boldsymbol{w'x}, \boldsymbol{y}_1, \boldsymbol{y}_2) \mapsto (\boldsymbol{u}, \boldsymbol{v})$ where

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{w'x} & \boldsymbol{y}_1^{-1} \end{bmatrix}', \tag{3.22}$$

$$\boldsymbol{v} = \boldsymbol{y}_2, \tag{3.23}$$

and $\boldsymbol{y}_1^{-1}$ denotes the vector of length $(q-r)$ containing the inverses of the elements of $\boldsymbol{y}_1$. Let $\mathcal{C}_\phi(\boldsymbol{w}, \mathcal{P}^t)) \in \mathbb{R}_+^{q+1}$ denote the set of points given by applying the transformation ϕ to every point in $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$. This transformation also induces a distribution function $G_\phi^t(\boldsymbol{u}, \boldsymbol{v}) = \Pr(\boldsymbol{U} \le \boldsymbol{u}, \boldsymbol{V} \ge \boldsymbol{v})$, which is a transformation of $G^t(\boldsymbol{w'x}, \boldsymbol{y})$ defined in (3.16) and which has support on $\mathcal{K}_\phi^t$, the set of points obtained by applying the transformation ϕ to every point in $\mathcal{K}^t$. Moreover, analogous to (3.17), $G_\phi^t(\boldsymbol{u}, \boldsymbol{v})$ can be decomposed by writing

$$G_\phi^t(\boldsymbol{u}, \boldsymbol{v}) = \underbrace{\Pr(\boldsymbol{U} \le \boldsymbol{u} \mid \boldsymbol{V} \ge \boldsymbol{v})}_{=G_\phi^t(\boldsymbol{u}|\boldsymbol{v})} \underbrace{\Pr(\boldsymbol{V} \ge \boldsymbol{v})}_{=S_\phi^t(\boldsymbol{v})}. \tag{3.24}$$

Since the transformation ϕ is monotonic, $G^t(\boldsymbol{w'x}, \boldsymbol{y}) = G_\phi^t(\boldsymbol{u}, \boldsymbol{v})$ for $\boldsymbol{u}$, $\boldsymbol{v}$ defined as in (3.22)–(3.23), and quantiles are stable with respect to monotonic transformations such as ϕ. Given free disposability of $\mathcal{P}^t$, it is easy to verify that $\mathcal{C}_\phi(\boldsymbol{w}, \Psi^t)$ is free-disposal in the sense that if $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}_\phi(\boldsymbol{w}, \Psi^t)$, then $(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}}) \in \mathcal{C}_\phi(\boldsymbol{w}, \Psi^t)$ $\forall$ $\widetilde{\boldsymbol{u}} \ge \boldsymbol{u}$ and $\boldsymbol{0} \le \widetilde{\boldsymbol{v}} \le \boldsymbol{v}$. In addition, the decomposition in (3.24) is identical to the decomposition that appears in Daouia and Simar (2007, p. 378, first line of second equation). Consequently, the distance function $\gamma_\alpha(\boldsymbol{w'x}, \boldsymbol{y}) \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$ defined in (3.19) can be written equivalently in terms of the $(\boldsymbol{u}, \boldsymbol{v})$-coordinates as

$$\begin{aligned} \gamma_\alpha(\boldsymbol{w'x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)) &= \sup\left\{\gamma > 0 \mid G_\phi^t(\gamma^{-1}\boldsymbol{u} \mid \boldsymbol{v}) > (1-\alpha)\right\} \\ &= \left[\inf\left\{\theta \mid G_\phi^t(\theta\boldsymbol{u} \mid \boldsymbol{v}) > (1-\alpha)\right\}\right]^{-1} \\ &= \theta_\alpha(\boldsymbol{u}, \boldsymbol{v})^{-1} \end{aligned} \tag{3.25}$$

where $\theta_\alpha(\boldsymbol{u}, \boldsymbol{v})$ is the α-quantile input efficiency score defined in Daouia and Simar ((2007), p. 379, Definition 2.1). In other words, the conditional α-quantile hyperbolic distance function defined in (3.19) for a unit operating at $(\boldsymbol{w'x}, \boldsymbol{y}) \in \mathcal{C}^t(\boldsymbol{w}, \mathcal{P}^t)$ is equivalent to the reciprocal of the α-quantile input efficiency score defined by Daouia and Simar when evaluated at $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}_\phi(\boldsymbol{w}, \mathcal{P}^t)$.

Note that in the $(\boldsymbol{u}, \boldsymbol{v})$-coordinates, $\mathcal{C}_\phi(\boldsymbol{w}, \mathcal{P}^t)$ satisfies the same property as the production set $\mathcal{P}^t$ in $(\boldsymbol{x}, \boldsymbol{y})$-coordinates. Hence $\gamma_\alpha(\boldsymbol{w'x}, \boldsymbol{y} \mid \mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$ can be estimated by the reciprocal of the non-parametric estimator of the α-quantile input efficiency score discussed in Daouia and Simar (2007, Section 3) after transforming the data as described above. For fixed values of α strictly less than one, the estimator is strongly consistent, asymptotically normal, and converges at rate $n^{1/2}$ under the assumptions introduced in Section 3.2 and additional mild assumptions given by Daouia and Simar.

4 Measuring Changes in Performance

Although cost efficiency is measured at a point in time, it is often interesting to evaluate how efficiency, productivity, and other measures of performance evolve over time. In competitive industries, one would expect inefficient firms to be driven from the market, although this does not happen instantaneously and firms that are inefficient today might become more efficient tomorrow and vice-versa. In the case of non-profit organizations, such as credit unions, competitive pressures that would encourage efficient operation may be absent or operate differently than they do for profit-seeking firms.

In a production framework with only one input and one output, *average product* is defined as the ratio of output to input quantities. If $\mathcal{P}^{\partial t}$ exhibits constant returns to scale everywhere, then productivity and technical efficiency are equivalent, although they might be measured differently. With variable returns to scale, however, technically efficient firms operating along $\mathcal{P}^{\partial t}$ in regions of either increasing or decreasing returns to scale will be less productive than technically efficient firms operating along the constant-returns region of $\mathcal{P}^{\partial t}$; they might also be less productive than some technically inefficient firms.

Simple ratios are not useful for measuring productivity in cases of multiple inputs and multiple outputs. Instead, Malmquist productivity (MP) indices are often used to measure *changes* in productivity. Recent examples in the banking literature include Alam (2001), Berg et al. (1992), Portela and Thanassoulis (2010), Wheelock and Wilson (1999), and Wheelock and Wilson (2009).

In a cost framework with only one output, *average cost* is simply cost divided by the single output quantity, and it is trivial to see how average cost changes over time. With multiple outputs, however, an index similar to the MP index is needed to provide a measure of change in cost-productivity, which is the multivariate analog of average cost. Recently, Ball et al. (2005) introduced a Malmquist cost-productivity (MCP) index to measure productivity growth within a cost framework similar to the framework developed above in Section 3.[14] They note that "since the cost structure of an industry is a fundamental determinant of cost-effective production decisions, a cost framework as used in MCP is a desirable foundation for representing production patterns and analyzing the productive contributions of... outputs and inputs to production."

Here we extend the ideas of Ball et al. (2005) to define a MCP index in terms of the partial cost frontier $\mathcal{C}_{\alpha}^{\partial}(\boldsymbol{w}, \mathcal{P}^t)$ defined in (3.20), as opposed to the full cost frontier $\mathcal{C}^{\partial}(\boldsymbol{w}, \mathcal{P}^t)$ defined in (3.8). We also define our MCP index in terms of hyperbolic efficiency measures, rather than in terms of the usual input- or output-oriented efficiency measures. Hyperbolic efficiency measures have an important advantage in that the hyperbolic-based index is always defined, whereas an MP index defined using input- or output- measures may not exist for some points when the technology shifts or rotates over time.[15]

Our approach to measuring cost-productivity allows both cost and some outputs to vary. In so doing, we avoid problems of infeasibility (i.e., existence) in the cross-period distance functions used to define components of the Malmquist cost-productivity index. Infeasibilities occur when all outputs are held constant and the cost frontier shifts so that a ray parallel to the cost axis from a credit union's location in one period does not intersect the estimated frontier in the other period. This problem was noted by Ball et al. (2005, p. 380), who worked in the cost direction while holding output quantities fixed. Given that infeasibilities often occur in cross-period studies such as ours when output quantities are held fixed, researchers in other applications should find our approach useful.

MP and MCP indices must be defined in terms of constant returns to scale to properly measure productivity changes in either the production or cost framework, and to allow the index to be interpreted in terms of total factor productivity.[16] Let $\mathcal{V}(\mathcal{P}^t)$ denote the convex cone (with vortex at the origin) of the production set $\mathcal{P}^t$ so that $\mathcal{P} \subseteq \mathcal{V}(\mathcal{P}^t)$. Then $\mathcal{P}^{\partial t}$ displays globally constant returns to scale if and only if $\mathcal{P} = \mathcal{V}(\mathcal{P}^t)$. Otherwise, $\mathcal{P} \subset \mathcal{V}(\mathcal{P}^t)$ and $\mathcal{P}^t$ is characterized by variable returns to scale. Using notation introduced in Section 3, we can write the MCP index introduced by Ball et al. (2005) for firm i defined in terms of

[14] A similar index was used by Edvardsen et al. (2006).

[15] See Wheelock and Wilson (2009) for additional discussion and an example of the use of hyperbolic efficiency measures to define an MP index in terms of distances to partial frontiers.

[16] See Färe and Grosskopf (1996) and Ball et al. (2005) for discussion.

the full cost frontier $\mathcal{C}^{\partial}(\boldsymbol{w}, \mathcal{P}^t)$ as

$$\mathscr{M}_{\rho,i}^{t_1,t_2} = \left[\frac{\rho(\boldsymbol{w}_i^{t_1 \prime} \boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{V}(\mathcal{C}(\boldsymbol{w}^{t_1}, \mathcal{P}^{t_1})))}{\rho(\boldsymbol{w}_i^{t_2 \prime} \boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{V}(\mathcal{C}(\boldsymbol{w}^{t_1}, \mathcal{P}^{t_1})))} \times \frac{\rho(\boldsymbol{w}_i^{t_1 \prime} \boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{V}(\mathcal{C}(\boldsymbol{w}^{t_2}, \mathcal{P}^{t_2})))}{\rho(\boldsymbol{w}_i^{t_2 \prime} \boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{V}(\mathcal{C}(\boldsymbol{w}^{t_2}, \mathcal{P}^{t_2})))} \right]^{1/2}, \quad (4.26)$$

where superscripts have been added to denote quantities at times t_1 and t_2. Note that we use the efficiency measure defined in (3.10) to define $\mathscr{M}_{\rho,i}^{t_1,t_2}$, but with $\mathcal{V}(\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t))$ replacing $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$ in (3.10).

MCP (and MP) indices can be defined in terms of various benchmarks, including either the full frontier $\mathcal{P}^{\partial t}$ or the partial frontier $\mathcal{C}_{\alpha}^{\partial}(\boldsymbol{w}, \mathcal{P}^t)$. Working in a hyperbolic direction while conditioning on quasi-fixed outputs, we define an MCP index in terms of the partial frontier defined in (3.20) by writing

$$\mathscr{M}_{i}^{t_1,t_2}(\alpha) = \left[\frac{\gamma(\boldsymbol{w}_i^{t_1 \prime} \boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{V}(\mathcal{C}_{\alpha}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1})))}{\gamma(\boldsymbol{w}_i^{t_2 \prime} \boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{V}(\mathcal{C}_{\alpha}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1})))} \times \frac{\gamma(\boldsymbol{w}_i^{t_1 \prime} \boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{V}(\mathcal{C}_{\alpha}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2})))}{\gamma(\boldsymbol{w}_i^{t_2 \prime} \boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{V}(\mathcal{C}_{\alpha}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2})))} \right]^{1/2}, \quad (4.27)$$

where $\gamma(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{V}(\mathcal{C}_{\alpha}(\boldsymbol{w}, \mathcal{P}^t)))$ is defined by replacing $\mathcal{C}(\boldsymbol{w}, \mathcal{P}^t)$ in (3.14) with the convex cone of the conditional, hyperbolic order-α cost set $\mathcal{C}_{\alpha}(\boldsymbol{w}, \mathcal{P}^t)$ defined in (3.21).

Both $\mathscr{M}_{\rho,i}^{t_1,t_2}$ and $\mathscr{M}_{i}^{t_1,t_2}(\alpha)$ measure changes in cost-productivity between times t_1 and t_2, and both consist of the geometric mean of two ratios of distance functions. They differ only in terms of the specific distance functions used to define the index. The distance functions used to define $\mathscr{M}_{\rho,i}^{t_1,t_2}$ measure cost efficiency holding output levels fixed, while the distance functions used to define $\mathscr{M}_{i}^{t_1,t_2}(\alpha)$ allow outputs to vary. Use of the convex hull operator $\mathcal{V}(\cdot)$ in both indices means that efficiency is measured relative to constant returns to scale technologies in all cases, which is necessary for proper measurement of changes in productivity.

In the first ratio on the RHS of (4.27), the lower boundary of $\mathcal{V}\left(\mathcal{C}_{\alpha}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1})\right)$ provides a fixed benchmark against which to measure any change in the cost-productivity of a firm operating at $(\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1}) \in \mathcal{P}^{t_1}$ in the first period and at $(\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2}) \in \mathcal{P}^{t_2}$ in the second period. Similarly, in the second ratio on the RHS of (4.27), the lower boundary of $\mathcal{V}\left(\mathcal{C}_{\alpha}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2})\right)$ provides a second fixed benchmark against which to measure changes in cost-productivity. If $\mathscr{M}_{i}^{t_1,t_2}(\alpha) > 1$, then the cost-productivity of firm i has increased between time t_1 and t_2. However, if $\mathscr{M}_{i}^{t_1,t_2}(\alpha) < 1$, then cost-productivity has decreased, and if $\mathscr{M}_{i}^{t_1,t_2}(\alpha) = 1$, then cost-productivity has not changed. Similar reasoning applies in the case of the index in (4.26).

In the literature, MP indices have been decomposed in a variety of ways to examine changes in efficiency, technology, and other aspects of performance. Here, we decompose the MCP index defined in (4.27) along the lines of the MP-index decomposition in Wheelock and Wilson (1999). In particular, using the cost-efficiency measure defined in (3.19), the change in cost efficiency experienced by firm i between times t_1 and t_2 is given by the ratio

$$\mathscr{E}_i^{t_1,t_2}(\alpha) \equiv \frac{\gamma_\alpha(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{C}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1}))}{\gamma_\alpha(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{C}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2}))}. \tag{4.28}$$

A value greater than (equal to, less than) 1 indicates an increase (no change, a decrease) in cost efficiency over the period from t_1 to t_2.

Analogous to changes in technology (i.e., changes in the full frontier $\mathcal{P}^{t\partial}$), changes in the order-α cost quantile are measured by the index

$$\mathscr{T}_i^{t_1,t_2}(\alpha) \equiv \left[\frac{\gamma_\alpha(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{C}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2}))}{\gamma_\alpha(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{C}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1}))} \times \frac{\gamma_\alpha(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{C}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2}))}{\gamma_\alpha(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{C}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1}))}\right]^{1/2}. \tag{4.29}$$

This index also consists of a geometric mean of two ratios. The first ratio on the RHS of (4.29) measures the extent to which the order-α cost frontier defined in (3.20) shifted between t_1 and t_2 along the hyperbolic path through the ith firm's position at time t_1. Similarly, the second ratio measures the shift in the cost frontier along the hyperbolic path through the same firm's location at time t_2. If $\mathscr{T}_i^{t_1,t_2}(\alpha) > 1$, then the cost frontier shifts downward between t_1 and t_2, creating the potential for a reduction in cost. If $\mathscr{T}_i^{t_1,t_2}(\alpha) < 1$, then the cost frontier shifts upward, reflecting an increase in the minimum cost of producing given amounts of output. No change is indicated by $\mathscr{T}_i^{t_1,t_2}(\alpha) = 1$.

The MCP index can be decomposed still further. For example, the decomposition used by Wheelock and Wilson (1999) contains two additional terms, analogous to

$$\mathscr{S}_i^{t_1,t_2}(\alpha) = \frac{\gamma(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1})))/\gamma_\alpha(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{C}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1}))}{\gamma(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2})))/\gamma_\alpha(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{C}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2}))} \tag{4.30}$$

and $\mathscr{W}_i^{t_1,t_2}(\alpha) = \mathscr{U}_i^{t_1,t_2}(\alpha) \times \mathscr{V}_i^{t_1,t_2}(\alpha)$, where

$$\mathscr{U}_i^{t_1,t_2}(\alpha) = \left[\frac{\gamma(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2})))/\gamma_\alpha(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{C}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2}))}{\gamma(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1})))/\gamma_\alpha(\boldsymbol{w}^{t_1\prime}\boldsymbol{x}_i^{t_1}, \boldsymbol{y}_i^{t_1} \mid \mathcal{C}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1}))}\right]^{1/2} \tag{4.31}$$

and

$$\mathscr{V}_i^{t_1,t_2}(\alpha) = \left[\frac{\gamma(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2})))/\gamma_\alpha(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{C}(\boldsymbol{w}_i^{t_2}, \mathcal{P}^{t_2}))}{\gamma(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1})))/\gamma_\alpha(\boldsymbol{w}^{t_2\prime}\boldsymbol{x}_i^{t_2}, \boldsymbol{y}_i^{t_2} \mid \mathcal{C}(\boldsymbol{w}_i^{t_1}, \mathcal{P}^{t_1}))}\right]^{1/2}. \tag{4.32}$$

Some algebra confirms that $\mathscr{M}_i^{t_1,t_2}(\alpha) = \mathscr{E}_i^{t_1,t_2}(\alpha) \times \mathscr{T}_i^{t_1,t_2}(\alpha) \times \mathscr{S}_i^{t_1,t_2}(\alpha) \times \mathscr{U}_i^{t_1,t_2}(\alpha) \times \mathscr{V}_i^{t_1,t_2}(\alpha)$.

The sub-index $\mathscr{S}_i^{t_1,t_2}(\alpha)$ defined in (4.30) consists of a ratio of two scale efficiency measures; the numerator is a measure of scale efficiency for firm i at time t_1, and the denominator measures scale efficiency for firm i at time t_2. By construction, both scale efficiency measures are necessarily weakly greater than 1, with larger (smaller) values indicating greater (less) scale *in*efficiency. The ratio in (4.30) provides a measure of the change in scale efficiency experienced by firm i between time t_1 and time t_2. Values of $\mathscr{S}_i^{t_1,t_2}(\alpha)$ greater than (less than) 1 indicate that firm i has become more (less) scale efficient between times t_1 and t_2, while a value of 1 indicates no change in scale efficiency. Note that a firm could become more scale efficient (i) by moving closer to the most-productive scale size (MPSS), i.e., the region where returns to scale are constant, or (ii) if the MPSS moves closer to the firm's location in the cost set due to changes in the shape or position of the technology (or, as defined here, the order-α quantile). Of course, any change in scale efficiency would reflect the net effect of any movement of the firm and changes in technology.

The residual terms $\mathscr{U}_i^{t_1,t_2}(\alpha)$ and $\mathscr{V}_i^{t_1,t_2}(\alpha)$ defined in (4.31) and (4.32) are useful for identifying the reasons for any changes in scale efficiency. The expression inside the square brackets on the RHS of (4.31) is a ratio of two scale efficiency measures, similar to (4.30). The numerator of (4.31) is a measure firm i's scale efficiency at time t_1 based on the technology available at time t_1. The denominator is also a measure based on firm i's location at time t_1, but relative to the technology available at time t_2. Hence, because the firm's position is held fixed, any change indicated by $\mathscr{U}_i^{t_1,t_2}(\alpha)$ reflects movement in the position or shape of the order-α cost quantile $\mathcal{C}_\alpha^\partial(\boldsymbol{w}, \mathcal{P}^t)$ defined in (3.20). As in (4.30), the scale efficiency measures in the numerator and denominator of (4.31) are by construction weakly greater than 1. Values of $\mathscr{U}_i^{t_1,t_2}(\alpha)$ greater than (equal to, less than) 1 imply that firm i would have experienced a decrease (no change, an increase) in scale efficiency between times t_1 and t_2 if it had been at its time t_1 location in $(\boldsymbol{w'x}, \boldsymbol{y})$-space at both times t_1 and t_2.

The term $\mathscr{V}_i^{t_1,t_2}(\alpha)$ in (4.32) is defined similarly to $\mathscr{U}_i^{t_1,t_2}(\alpha)$, except that the scale efficiency measures that define $\mathscr{V}_i^{t_1,t_2}(\alpha)$ hold firm i's position fixed at its location at time t_2 instead of t_1. Similar reasoning reveals that values of $\mathscr{V}_i^{t_1,t_2}(\alpha)$ greater than (equal to, less than) 1 imply that firm i would have experienced a decrease (no change, an increase) in scale efficiency between times t_1 and t_2 if it had been at its time t_2 location at *both* times t_1 and t_2.

The quantities of interest, namely the indices $\mathscr{M}_i^{t_1,t_2}(\alpha)$, $\mathscr{E}_i^{t_1,t_2}(\alpha)$, $\mathscr{T}_i^{t_1,t_2}(\alpha)$, $\mathscr{S}_i^{t_1,t_2}(\alpha)$, $\mathscr{U}_i^{t_1,t_2}(\alpha)$, and $\mathscr{V}_i^{t_1,t_2}(\alpha)$ defined in (4.27)–(4.32), are not observed and therefore must be estimated. Consistent estimates of the α-quantile distance functions based on $\mathcal{C}(\boldsymbol{w}^t, \mathcal{P}^t)$ used in (4.28)–(4.32) can be obtained as described earlier in Section 3.3, while the distance functions based on the convex hull of sets defined by (3.21) can be estimated as described in Appendix A.

5 Data and Estimation Results

We use data on credit unions that are observed in both 1989 and 2006. We screened for obvious data problems and omitted observations for credit unions that reported non-positive values for loans or investments, or where the calculated values for *WCAP, WLAB, PRSAV,* or *PRLOAN* are outside the interval $(0, 1)$ or those for *CAP* or *LAB* are non-positive, leaving 7,219 matched observations for 1989 and 2006. Table 2 reports summary statistics for the variables in our model.[17] The minimum and maximum values displayed in Table 2 are in some cases rather extreme. However, we were conservative in deleting implausible observations since screening data for implausible observations is always subjective to some extent and our estimators are robust with respect to outliers.

Empirical application of the order-α quantile estimator requires the choice of one or more values for α. Daouia and Simar (2007) and Wheelock and Wilson (2009) consider various values for α near, but less than, 1.0, and find that the choice makes little difference qualitatively in their empirical results. Similarly, here we computed contemporaneous efficiency estimates using values of $\alpha \in \{0.95,\ 0.96,\ \ldots,\ 0.99\}$ for our sample of U.S. credit unions. Figure 2 shows scatter plots of contemporaneous hyperbolic order-α quantile efficiency estimates for several values of α for the years 1989 and 2006. The efficiency estimates for different values

[17] Quarterly call report data for credit unions are available from the National Credit Union Administration: http://www.ncua.gov/DataServices/FOIA/5300CallReportData.aspx.

of α are highly correlated, as indicated by the fact that points in the scatter plots tend to lie near 45-degree lines, suggesting that our results will depend little, if at all, on the choice of α. Indeed, we find that qualitatively our results are largely invariant to the choice of α; hence, we report estimates based only on $\alpha = 0.95$.[18]

We computed estimates of the Malmquist cost-productivity index defined in (4.27) and its various components by replacing the unknown distance function values in (4.27)–(4.32) with estimates computed using the strategy described in Appendix A. Tables 3–8 report geometric means of estimates for the Malmquist index and its various components over observations grouped by quintiles of total assets in 1989 (rows) and 2006 (columns). Quintile Q1 consists of credit unions in the smallest-size quintile in a given year, Q2 consists of those in the next smallest size-quintile, etc.[19] In addition to the geometric means, the cells of each table report in parentheses the number of credit unions in each quintile across the two periods. For example, there were 1,119 credit unions with total assets among the smallest 20 percent of all credit unions in both 1989 and 2006, 279 credit unions with assets among the smallest 20 percent in 1989 (Q1) but among the second-smallest 20 percent in 2006 (Q2), and 39 credit unions with assets among the smallest 20 percent in 1989 (Q1) but among the third-smallest 20 percent in 2006 (Q3), etc.

Table 3 reports means of estimated changes in cost-productivity for credit unions between 1989 and 2006. We used standard bootstrapping methods to determine statistical significance of the means (i.e., significant differences from 1.0), which is indicated in the table by asterisks.[20] The results indicate that cost-productivity tended to fall among credit unions in all size groups, but on average, smaller credit unions experienced the largest de-

[18] Estimates corresponding to $\alpha = 0.96, \ldots, 0.99$ are available in a separate appendix (Appendix B) available from the authors on request.

[19] The quintiles of total assets consist of credit unions with total assets in the following ranges: for 1989, \$97,038–\$2,379,546; \$2,379,546–\$5,758,445; \$5,758,445–\$12,341,856; \$12,341,856–\$33,085,534; \$33,085,534–\$5,504,010,913; for 2006, \$115,848–\$3,512,487; \$3,512,487–\$9,204,064; \$9,204,064–\$22,949,767; \$22,949,767–\$67,414,680; \$67,414,680–\$25,754,336,720 (all figures in constant year 2000 dollars).

[20] Specifically, we draw from the empirical distribution of our data by selecting credit unions (uniformly, independently, and with replacement) to build a bootstrap pseudo data set. On a particular draw, if the ith credit union is selected, then its observations in both 1989 and 2006 enter the pseudo data. Once the pseudo data have been constructed, distances are measured from observations in the original data to quantiles estimated from the pseudo data, analogous to the original estimation. After 2,000 bootstrap replications, we use the various bootstrap values to estimate confidence intervals using the bias-correction described by Efron and Tibshirani (1993). Significant difference from 1 is determined by whether an estimated confidence interval includes 1.

clines. For example, cost-productivity fell by an average 30.3 percent $((1.0 - 0.6969) \times 100)$ for credit unions in the smallest-size quintile (Q1) in both 1989 and 2006, but by just 8.2 percent $((1.0 - 0.9176) \times 100)$ for credit unions located in the largest-size quintile (Q5) in both periods. With the exception of the two credit unions that moved from Q2 in 1989 to Q5 in 2006, credit unions moving to a larger-size quintile between 1989 and 2006 tended to exhibit smaller declines in cost-productivity than those remaining in the same quintile or moving to a smaller-size quintile. For example, productivity fell by an average of 18.2 percent among the 914 credit unions in Q4 in both 1989 and 2006, but by just 7.7 percent among those moving from Q4 to Q5. Our results are thus consistent with the view that environmental changes, such as improvements in information-processing and credit-monitoring technologies and changes in regulation, have tended to favor larger financial firms over their smaller competitors.

Next we examine changes in efficiency and technology for further insights into changes in the performance of credit unions of different sizes. Table 4 reports means of estimated changes in cost efficiency for credit unions in each asset-size quintile. Here we find that the largest credit unions—those in Q5 in both 1989 and 2006—experienced an average *decline* in cost efficiency of 5.3 percent $((1.0 - 0.9474) \times 100)$. By contrast, smaller credit unions tended to become more cost efficient, except those moving from Q1 in 1989 to either Q2 or Q3 in 2006. Moreover, the improvements in cost efficiency tended to be large, ranging from an average of 4.3 percent for credit unions moving from Q5 to Q4, to 43.2 percent for the one credit union moving from Q1 to Q5.

Results for the mean changes in technology, reported in Table 5, help to reconcile the results for changes in productivity and efficiency. The mean changes in technology for the largest credit unions—those located in Q5 in 2006—are not statistically different from 1.0 (except for the one credit union that moved from Q1 in the 1989 to Q5 in 2006), although the point estimates suggest some improvement in technology between 1989 and 2006. However, the mean changes in technology are mostly less than 1.0 for smaller credit unions, indicating that an unfavorable shift in the cost frontier can explain the simultaneous declines in cost-productivity and increases in cost efficiency experienced by many smaller credit unions between 1989 and 2006. Although many small credit unions incurred higher costs in producing given amounts of output, and therefore became less cost-productive, shifts in technology reduced their distance from the cost frontier, which made them more cost efficient.

Those credit unions moving from one of the smaller quintiles in 1989 to Q5 in 2006 experienced (on average) increases in cost efficiency as indicated by the results in Table 4 as well as possible improvements in technology, but also declines (on average) in cost-productivity, as indicated by the results in Table 3. Our results for changes in scale efficiency and its sub-components, reported in Tables 6–8, help resolve this apparent paradox.

Table 6 reports estimates of changes in scale efficiency over 1989–2006. Our results indicate that most credit unions experienced either *declines* or no statistically significant changes in scale efficiency, indicated by mean values less than 1.0. Only two groups (the single credit union that moved from Q1 to Q5, and those that were in Q5 in both periods) show improvements in scale efficiency, and only one of these is statistically significant.

Recall from the discussion in Section 4 that a credit union would become more scale efficient if (i) it moved closer to a region of constant returns to scale or (ii) a region of constant returns to scale shifted closer to the credit union's location in the cost space. To help disentangle these two effects, we turn to the results in Tables 7 and 8. The results for the residual term $\mathscr{U}_i^{t_1,t_2}(\alpha)$ reported in Table 7 indicate that, on average, credit unions located in quintiles Q2–Q5 in 1989 would have experienced *declines* in scale efficiency had they remained at their 1989 locations in $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y})$-space. The estimates in Table 7 are large and statistically significant for the largest size-quintiles. For example, credit unions in Q5 in both years would have experienced an average decline in scale efficiency of about 246 percent if they had remained at their 1989 locations in $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y})$-space in 2006.

The results in Table 8 show what would have happened to scale efficiency if credit unions had been at their 2006 locations in the cost space in both 1989 and 2006. Credit unions in Q3–Q5 in both years would have experienced substantial and statistically significant *improvements* (on average) in scale efficiency had they been at their 2006 locations in $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y})$-space in 1989. These results indicate that many credit unions moved to regions in the cost-space where scale economies showed improvement between 1989 and 2006. Nonetheless, recalling the results for changes in scale efficiency shown in Table 6, it appears that, on average, only the largest credit unions became more scale efficient between 1989 and 2006 (and even then, mean improvement was not generally statistically significant). In other words, although larger credit unions tended to move to locations where scale efficiency improved, the extent of improvement at those locations was insufficient to produce statistically significant gains in scale efficiency for most credit unions.

The smallest credit unions, i.e., those in Q1 in 2006, experienced the largest average declines in scale efficiency over time. Interestingly, our results indicate that credit unions in Q1 in both 1989 and 2006 would have experienced no change, on average, in scale efficiency if they had merely remained at their 1989 location in $(\boldsymbol{w}'\boldsymbol{x}, \boldsymbol{y})$-space in both years. However, they would have experienced an average decline in cost efficiency of about 11 percent if they had been at the 2006 locations in both periods, and because of their movement within the cost space, these credit unions experienced an average decline in cost efficiency of about 20 percent. Taken as a whole, our results indicate that smaller credit unions, and those which dropped from larger- to smaller-size quintiles, fared less well than their larger and faster-growing competitors in terms of scale efficiency, as well as cost-productivity and changes in technology.

Our estimation results reveal substantial changes in the cost productivity, efficiency, technology, and scale efficiency of credit unions over the 17 year period 1989–2006. We investigated whether these changes occurred primarily in either the early or later years of the period by re-estimating the indices $\mathscr{M}_i^{t_1,t_2}(\alpha)$, $\mathscr{E}_i^{t_1,t_2}(\alpha)$, $\mathscr{T}_i^{t_1,t_2}(\alpha)$, $\mathscr{S}_i^{t_1,t_2}(\alpha)$, $\mathscr{U}_i^{t_1,t_2}(\alpha)$, and $\mathscr{V}_i^{t_1,t_2}(\alpha)$ separately for 1989–1997 and for 1997–2006. We found that both the patterns and magnitudes of changes in productivity, efficiency, etc., were broadly similar over the two periods. As for the entire period 1989–2006, in both subperiods we found that larger credit unions tended to experience smaller declines in productivity and larger increases in scale efficiency than small credit unions. These results suggest that the Credit Union Membership Access Act of 1998, which permitted federally-insured credit unions to accept members from unrelated groups, did not substantially affect long-run patterns of productivity or efficiency change.[21]

We also investigated whether changes in productivity, efficiency, etc., differ systematically by credit union regulator or insurance agency. Federally-chartered credit unions are regulated and supervised by the National Credit Union Administration (NCUA), and the deposits of federally-chartered credit unions are insured by the NCUA. State-chartered credit unions,

[21] This is perhaps not surprising since the Credit Union Membership Access Act of 1998 confirmed a practice permitted by the National Credit Union Administration that a court ruling had overturned. Estimation results for the two subperiods are available from the authors upon request.

however, are subject to state credit union regulations and supervision. Many state-chartered credit unions carry NCUA insurance and they are required to comply with federal regulations covering all federally-insured credit unions. However, some state-chartered credit unions are insured by state insurance systems, and those credit unions are not subject to NCUA regulation or supervision. We reestimated our models separately for 1) federally-chartered credit unions; 2) all state-chartered credit unions; and 3) all federally-insured state-chartered credit unions (the number of non-federally insured state credit unions is too small to obtain meaningful estimates). Further, we estimated our models separately for each group for each subperiod 1989–97, 1997-2006, as well as for 1989–2006 as a whole. The results for each group and each subperiod were again remarkably similar, with no qualitative differences in performance between federally-chartered and state-chartered credit unions.[22]

6 Conclusions

The regulatory and technological environment in which credit unions and other depository institutions operate has changed dramatically since the mid-1980s. For example, the requirement that members of a credit union be linked by a common bond has been eased, as have limits on business lending by credit unions. The recently enacted Dodd-Frank Wall Street Reform and Consumer Protection Act subjects credit unions to similar consumer protection, disclosure and reporting rules as commercial banks, while legislation pending in Congress would further ease limits on business lending by credit unions.

Alongside the evolving regulatory environment, advances in information-processing technology have lowered the cost of obtaining and evaluating information about potential borrowers, and thereby eroded some of the advantages of small scale and a common membership bond in serving consumers and small business customers. As the regulatory and technological environment have evolved, the average size of credit unions, like commercial banks, has increased and the industry has consolidated through a wave of mergers. Whether credit unions will remain viable in the long run depends, in part, on how well they continue to adapt to the changing environment in which they operate. Our research indicates that between 1989 and 2006, credit unions, on average, became less cost-productive, i.e., they incurred greater

[22] Results for the various periods and subsamples are available from the authors upon request.

operating costs in producing given levels of outputs. In addition, we find that productivity declines were largest among the smaller credit unions. Although small credit unions tended to become more cost efficient, adverse movement in the location and shape of the cost frontier at the left end of the size distribution resulted in significant declines in cost-productivity among small credit unions. Our results are thus consistent with the view that recent changes in regulation and technology have tended to favor larger depository institutions over their smaller competitors. They are also consistent with findings for commercial banks.

Methodologically, this paper provides innovation on several fronts. First, we adapted the hyperbolic unconditional quantile estimator of Wheelock and Wilson (2008) to the estimation of cost efficiency where input prices and some outputs are fixed, but other outputs and cost are variable. Second, we derived the corresponding cost analog of the Malmquist productivity index, which we decomposed to allocate changes in cost-productivity to changes in cost efficiency, technology, and scale efficiency. Further, we decomposed a residual term to gain insights about the sources of changes in scale efficiency. Finally, we developed an estimator of our new distance function, from which we derived estimators of our new Malmquist index and components. These techniques could be used in a variety of settings to examine changes in the performance of financial institutions and other types of firms.

A Appendix: Nonparametric Estimation Strategy

In order to estimate $\gamma(\boldsymbol{w}_i'\boldsymbol{x}_i, \boldsymbol{y}_i \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i, \mathcal{P}^t)))$ corresponding to firm i, first use the Daouia and Simar (2007) estimator to obtain an estimate $\widehat{\gamma}_i = \widehat{\gamma}_\alpha(\boldsymbol{w}_i'\boldsymbol{x}_i, \boldsymbol{y}_i \mid \mathcal{C}(\boldsymbol{w}_i \mid \mathcal{P}^t))$ for firm i as described above in Section 3.3. Next, for firms $j = 1, \ldots, n$, $j \neq i$, use the same approach to compute estimates $\widehat{\gamma}_j = \widehat{\gamma}_\alpha(\boldsymbol{w}_i'\boldsymbol{x}_j, \boldsymbol{y}_j \mid \mathcal{C}(\boldsymbol{w}_i \mid \mathcal{P}^t)$ (note that here, the price vector $\boldsymbol{w}_i$ faced by firm i is used). Then project observations onto the estimated order-α cost frontier by computing $(\widehat{\gamma}_j^{-1}\boldsymbol{w}_i'\boldsymbol{x}_j, \widehat{\gamma}_j\boldsymbol{y}_{1j}, \boldsymbol{y}_{2j})$ for each $j = 1, \ldots, n$ to form a set $\mathcal{A}_{i,n} = \{(\widehat{\gamma}_j^{-1}\boldsymbol{w}_i'\boldsymbol{x}_j, \widehat{\gamma}_j\boldsymbol{y}_{1j}, \boldsymbol{y}_{2j})\}_{j=1}^n$. By construction, the points in $\mathcal{A}_{i,n}$ form an estimate of the frontier $\mathcal{C}_\alpha^\partial(\boldsymbol{w}_i, \mathcal{P}^t)$ defined by (3.20). Given the convergence of $\widehat{\gamma}_\alpha(\boldsymbol{w}_i'\boldsymbol{x}_i, \boldsymbol{y}_i \mid \mathcal{C}(\boldsymbol{w}_i \mid \mathcal{P}^t))$, we can reasonably conjecture that $\mathcal{A}_{i,n}$ converges pointwise to the true frontier $\mathcal{C}_\alpha^\partial(\boldsymbol{w}_i, \mathcal{P}^t)$. Also by construction, the conical hull of this estimate is $\mathcal{V}(\mathcal{A}_{i,n})$. Distance to $\mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i, \mathcal{P}^t))$ is thus estimated by distance to $\mathcal{V}(\mathcal{A}_{i,n})$, which can be computed to an arbitrary degree using the bisection method along the lines of Wheelock and Wilson (2008).

The DEA estimator of the Shephard (1970) input distance function (e.g., see Simar and Wilson, 2000) can be used to assess convergence of the bisection method. Denote (for a particular firm i) the $(q+1)$-tuples in $\mathcal{A}_{i,n}$ by $(z_{x,j}, \boldsymbol{z}_{y,j})$ where $z_{x,j} = \widehat{\gamma}_j^{-1}\boldsymbol{w}_i'\boldsymbol{x}_j$ and $z_{y,j} = \begin{bmatrix}\widehat{\gamma}_j\boldsymbol{y}_{1j}' & \boldsymbol{y}_{2j}'\end{bmatrix}'$. Then for an arbitrary point $(z_x, \boldsymbol{z}_y) \in \mathbb{R}_+1 \times \mathbb{R}_+q$, the constant returns to scale version of the **?** input distance function estimator $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{V}(\mathcal{A}_{i,n}))$ can be computed as

$$\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{V}(\mathcal{A}_{i,n})) = \min_{\theta,\omega_1,\ \ldots,\ n} \{\theta > 0 \mid \boldsymbol{z}_y \leq \sum_{j=1}^{n} \omega_j \boldsymbol{z}_{yj}, \theta z_x \geq \sum_{j=1}^{n} \omega_j z_{xj}, \\ \omega_j \geq 0 \ \forall\ j = 1,\ \ldots,\ n\}. \tag{A.1}$$

The complete algorithm for computing an estimate $\widehat{\gamma}(\boldsymbol{w}_i'\boldsymbol{x}_i, \boldsymbol{y}_i \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i, \mathcal{P}^t)))$ of $\gamma(\boldsymbol{w}_i'\boldsymbol{x}_i, \boldsymbol{y}_i \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i, \mathcal{P}^t)))$ appears below.

Algorithm #1

[1] Set $\gamma_a = 1$, $\gamma_b = 1$.

[2] Set $z_x = \gamma_a^{-1}\boldsymbol{w}_i'\boldsymbol{x}_i$ and $\boldsymbol{z}_y = \begin{bmatrix}\gamma_a\boldsymbol{y}_{1i}' & \boldsymbol{y}_{2i}'\end{bmatrix}'$.

[3] Compute $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n})$; if $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n}) \geq 1$ then set $\gamma_a = 2 \times \gamma_a$.

[4] Repeat steps [2]–[3] until $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n}) < 1$.

[5] Set $z_x = \gamma_b^{-1}\boldsymbol{w}_i'\boldsymbol{x}_i$ and $\boldsymbol{z}_y = \begin{bmatrix}\gamma_b\boldsymbol{y}_{1i}' & \boldsymbol{y}_{2i}'\end{bmatrix}'$.

[6] Compute $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n})$; if $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n}) < 1$ then set $\gamma_b = 0.5 \times \gamma_b$.

[7] Repeat steps [5]–[6] until $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n}) >= 1$.

[8] Set $\gamma_c := (\gamma_a + \gamma_b)/2$, $z_x = \gamma_c^{-1}\boldsymbol{w}_i'\boldsymbol{x}_i$, $\boldsymbol{z}_y = \begin{bmatrix}\gamma_c\boldsymbol{y}_{1i}' & \boldsymbol{y}_{2i}'\end{bmatrix}'$, and compute $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n})$.

[9] If $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n}) >= 1$, then set $\gamma_a := \gamma_c$; otherwise, set $\gamma_b := \gamma_c$.

[10] If $(\gamma_b - \gamma_a) > \epsilon$, where ϵ is a suitably small tolerance value, repeat steps [8]–[9].

[11] If $\widehat{\theta}(z_x, \boldsymbol{z}_y \mid \mathcal{A}_{i,n}) >= 1$, then set $\widehat{\gamma}(\boldsymbol{w}_i'\boldsymbol{x}_i, \boldsymbol{y}_i \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i, \mathcal{P}^t))) = \gamma_c$; otherwise, set $\widehat{\gamma}(\boldsymbol{w}_i'\boldsymbol{x}_i, \boldsymbol{y}_i \mid \mathcal{V}(\mathcal{C}_\alpha(\boldsymbol{w}_i, \mathcal{P}^t))) = \gamma_b$.

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.

Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005), Nonparametric frontier estimation: A conditional quantile-based approach, *Econometric Theory* 21, 358–389.

Ball, E., R. Färe, S. Grosskopf, and O. Zaim (2005), Accounting for externalities in the measurement of productivity growth: the Malmquist cost productivity measure, *Structural Change and Economic Dynamics* 16, 374–394.

Bauer, K. (2008), Detecting abnormal credit union performance, *Journal of Banking and Finance* 32, 573–586.

Bauer, K. J., L. L. Miles, and T. Nishikawa (2009), The effect of mergers on credit union performance, *Journal of Banking and Finance* 33, 2267–2274.

Berg, S. A., F. R. Førsund, and E. S. Jansen (1992), Malmquist indices of productivity growth during the deregulation of norwegian banking, 1980–89, *Scandanavian Journal of Economics* 94, S211–S228.

Berger, A. N. (2003), The economic effects of technological progress: Evidence from the banking industry, *Journal of Money, Credit, and Banking* 35, 141–76.

Berger, A. N., R. S. Demsetz, and P. E. Strahan (1999), The consolidation of the financial services industry: Causes, consequence, and implications for the future, *Journal of Banking and Finance* 23, 135–194.

Bernanke, B. S. (2006), Community banking and community bank supervision in the twenty-first century. Remarks at the Independent Community Bankers of America National Convention and Techworld, Las Vegas, Nevada, March 8, 2006.

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

Chambers, R. G., Y. Chung, and R. Färe (1996), Benefit and distance functions, *Journal of Economic Theory* 70, 407–419.

— (1998), Profit, directional distance functions, and nerlovian efficiency, *Journal of Optimization Theory and Applications* 98, 351–364.

Daouia, A. (2003), *Nonparametric Analysis of Frontier Production Functions and Efficiency Measurement using Nonstandard Conditional Quantiles*, Ph.D. thesis, Groupe de Recherche en Economie Mathématique et Quantititative, Université des Sciences Sociales, Toulouse I, et Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse III.

Daouia, A. and A. Ruiz-Gazen (2006), Robust nonparametric frontier estimators: Qualitative robustness and influence function, *Statistica Sinica* 16, 1233–1253.

Daouia, A. and L. Simar (2007), Nonparametric efficiency analysis: A multivariate conditional quantile approach, *Journal of Econometrics* 140, 375–400.

Daraio, C. and L. Simar (2005a), Conditional nonparametric frontier models for convex and non-convex technologies: A unifying approach. Discussion paper #0502, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

— (2005b), Introducing environmental variables in nonparametric frontier models: A probabilistic approach, *Journal of Productivity Analysis* 24, 93–121.

Deprins, D., L. Simar, and H. Tulkens (1984), Measuring labor inefficiency in post offices, in M. M. P. Pestieau and H. Tulkens, eds., *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland, pp. 243–267.

Edvardsen, D. F., F. R. Førsund, W. Hansen, S. A. C. Kittelsen, and T. Neurauter (2006), Productivity and regulatory reform of norwegian electricity distribution utilities, in T. Coelli and D. Lawrence, eds., *Performance Measurement and Regulation of Network Utilities*, chapter 5, Glos, Unitied Kingdom: Edward Elgar Publishing, Ltd., pp. 97–131.

Efron, B. and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, London: Chapman and Hall.

Emmons, W. R. and F. A. Schmid (1999a), Credit unions and the common bond, *Federal Reserve Bank of St. Louis Review* 81, 41–64.

— (1999b), Wages and risk-taking in occupational credit unions: Theory and evidence, *Federal Reserve Bank of St. Louis Review* 81, 13–31.

Färe, R. (1988), *Fundamentals of Production Theory*, Berlin: Springer-Verlag.

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

Färe, R., S. Grosskopf, and C. A. K. Lovell (1985), *The Measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff Publishing.

Farrell, M. J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society A* 120, 253–281.

Frame, W. S., G. V. Karels, and C. A. McClatchey (2003), Do credit unions use their tax advantage to benefit members? evidence from a cost function, *Review of Financial Economics* 12, 35–47.

Fried, H. O., C. A. K. Lovell, and P. V. Eeckaut (1993), Evaluating the performance of U.S. credit unions, *Journal of Banking and Finance* 17, 251–265.

Fried, H. O., C. A. K. Lovell, and S. Yaisawarng (1999), The impact of mergers on credit union service provision, *Journal of Banking and Finance* 23, 367–386.

Goddard, J. A., D. G. McKillop, and J. O. S. Wilson (2002), The growth of U.S. credit unions, *Journal of Banking and Finance* 26, 2327–2356.

— (2008), The diversification and financial performance of U.S. credit unions, *Journal of Banking and Finance* 32, 1836–1849.

Kneip, A., L. Simar, and P. W. Wilson (2008), Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory* 24, 1663–1697.

Leggett, K. J. and R. W. Strand (2002), Membership growth, multiple membership groups and agency control at credit unions, *Review of Financial Economics* 11, 37–46.

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

Park, B. U., S.-O. Jeong, and L. Simar (2010), Asymptotic distribution of conical-hull estimators of directional edges, *Annals of Statistics* 38, 1320–1340.

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

Petersen, M. A. and R. G. Rajan (2002), Does distance still matter? the information revolution in small business lending, *The Journal of Finance* 57, 2533–2570.

Portela, M. C. A. S. and E. Thanassoulis (2010), Malmquist-type indices in the presence of negative data: An application to bank branches, *Journal of Banking and Finance* 34, 1472–1483.

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

Simar, L. and A. Vanhems (2012), Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* 166, 342–354.

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

Smith, D. (1984), A theoretic framework for the analysis of credit union decision making, *Journal of Finance* 69, 1155–1168.

Smith, D. J., T. F. Cargill, and R. A. Meyer (1981), An economic theory of a credit union, *Journal of Finance* 36, 519–528.

Walter, J. R. (2006), Not your father's credit union, *Federal Reserve Bank of Richmond Economic Quarterly* 92, 353–377.

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.

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

— (2008), Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations, *Journal of Econometrics* 145, 209–225.

— (2009), Robust nonparametric quantile estimation of efficiency and productivity change in U. S. commercial banking, 1985–2004, *Journal of Business and Economic Statistics* 27, 354–368.

— (2011), Are credit unions too small?, *Review of Economics and Statistics* 93, 1343–1359.

Wilcox, J. A. (2006), Performance divergence of large and small credit unions, *Federal Reserve Bank of San Francisco Economic Letter* 19.

Wilson, P. W. (2009), Asymptotic properties of some non-parametric hyperbolic efficiency estimators. Unpublished working paper. The John E. Walker Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, South Carolina 29634, USA.

— (2011), Asymptotic properties of some non-parametric hyperbolic efficiency estimators, in I. van Keilegom and P. W. Wilson, eds., *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Berlin: Springer-Verlag, pp. 115–150.

Table 1: Variable Definitions

LOANS — total loans and leases (CUSA1263).

INVEST — Investments: for 1989, total investments (less derivatives contracts) (CUSA4577); for 2006, balances due from depository institutions in the US (CUSA0082) + investments eligible for liquidity (CUSA0851) + membership capital at corporate credit unions (CUSAB158) + deposits in commercial banks, S&Ls, savings banks (total amount) (CUSA8632) + paid in capital at corporate credit unions (CUSAB148) + all other investments in corporate credit unions (CUSA1110) + US Treasury securities—book value (excluding trading accounts) (CUSA0400) + US Government agency and corporation obligations—book value (excluding trading accounts) (CUSA0600) + mutual funds (CUSA8628) + shares, deposits, and certificates in other credit unions, total amount (CUSA1116).

PRSAV — Savings pricing: [dividends on shares (CUSA4278) + interest on deposits (CUSA4279)] / total shares and deposits (CUSA2197).

PRLOAN — Loan pricing: interest and fee income on loans, total (CUSA(4010) / amount of total loans and leases (CUSA1263).

CAP — Financial capital: total shares and deposits (CUSA2196).

LAB — Labor: number of full-time credit union employees (CUSA6047) + (1/2 times) number of part-time credit union employees (CUSA6048).

WCAP — Price of financial capital: capital expenses, i.e. gross occupancy expense (CUSA4210) + office operations expense (CUSA4209) + advertising expense (CUSA4143) + travel and conference expense (CUSA4207) + loan expenses (CUSA4152) + operating expenses fees, professional and outside services (CUSA4211) + other operating expenses (CUSA4240) + miscellaneous operating expenses (CUSA4526), divided CAP

WLAB — Price of labor: labor expenses, i.e. officers and employee compensation (CUSA4137), divided by LAB.

COST — Variable cost: capital expenses + labor expenses, i.e., $(WCAP \times CAP) + (WLAB \times LAB)$.

Table 2: Summary Statistics for Variables Used in Cost Function Specification

	Min	Quartile 1	Median	Mean	Quartile 3	Max
			1989			
LOANS	19.69	2196.0	5697.0	2.151E+04	1.665E+04	4.161E+06
INVEST	2.813	529.5	1868.0	9523.0	6554.0	1.733E+06
PRSAV	1.200E-05	0.05645	0.06204	0.06186	0.06779	0.1797
PRLOAN	0.01058	0.1060	0.1131	0.1144	0.1213	0.3319
CAP	78.76	2674.0	7489.0	2.964E+04	2.319E+04	4.630E+06
LAB	0.5000	2.000	4.000	14.30	11.50	2160.0
WCAP	0.0003270	0.01106	0.01589	0.01677	0.02136	0.1081
WLAB	0.2599	23.91	30.24	30.26	36.47	720.1
COST	2.813	87.19	243.3	949.6	761.5	1.411E+05
			2006			
LOANS	1.899	2761.0	8509.0	6.237E+04	3.166E+04	1.954E+07
INVEST	7.597	1206.0	3965.0	2.081E+04	1.263E+04	4.653E+06
PRSAV	0.0003210	0.01518	0.02012	0.02045	0.02502	0.1190
PRLOAN	0.02128	0.06192	0.06775	0.07098	0.07576	0.5000
CAP	57.92	3955.0	1.224E+04	7.582E+04	4.256E+04	1.966E+07
LAB	0.5000	2.500	6.500	29.69	22.00	5282.0
WCAP	0.001388	0.01552	0.02148	0.02245	0.02780	0.1097
WLAB	0.4368	32.83	40.64	40.85	48.85	148.6
COST	2.849	170.0	547.0	2887.0	1962.0	6.744E+05

NOTE: *COST, LOANS,* and *INVEST* are measured in thousands of year-2000 dollars. *PRSAV. PRLOAN,* and *WCAP* are dimensionless quantities; *WLAB* is measured in thousands of year-2000 dollars per full-time equivalent employee.

Table 3: Mean Estimates of Changes in Cost Productivity $\mathscr{M}_i^{t_1,t_2}(\alpha)$ by Size Quintile ($\alpha = 0.95$)

		2006				
		Q1	Q2	Q3	Q4	Q5
	Q1	0.6969***	0.7311***	0.7546***	0.8420***	0.7961***
		(1119)	(279)	(39)	(5)	(1)
	Q2	0.7184***	0.7352***	0.7909***	0.8826***	0.6639***
		(300)	(785)	(320)	(38)	(2)
1989	Q3	0.7373***	0.7240***	0.7761***	0.8611***	0.9459***
		(24)	(367)	(756)	(278)	(18)
	Q4	—	0.7039***	0.7411***	0.8179***	0.9226***
		(0)	(14)	(323)	(914)	(193)
	Q5	—	—	0.6942***	0.7973***	0.9176***
		(0)	(0)	(5)	(209)	(1230)

NOTE: Each cell contains two entries; the top entry gives the mean, and the bottom entry gives the number of observations. One, two, or three asterisks indicate significance at 90, 95, or 99-percent, respectively.

Table 4: Mean Estimates of Changes in Cost Efficiency $\mathscr{E}_i^{t_1,t_2}(\alpha)$ by Size Quintile ($\alpha = 0.95$)

		2006				
		Q1	Q2	Q3	Q4	Q5
	Q1	1.0761***	0.9762*	0.9523*	1.2037***	1.4318***
		(1119)	(279)	(39)	(5)	(1)
	Q2	1.1218***	1.0793***	1.0776***	1.1472***	1.2275***
		(300)	(785)	(320)	(38)	(2)
1989	Q3	1.1835***	1.0853***	1.0784***	1.1117***	1.1689***
		(24)	(367)	(756)	(278)	(18)
	Q4	—	1.1644***	1.0730***	1.0947***	1.1282***
		(0)	(14)	(323)	(914)	(193)
	Q5	—	—	1.0526	1.0426**	0.9474**
		(0)	(0)	(5)	(209)	(1230)

NOTE: Each cell contains two entries; the top entry gives the mean, and the bottom entry gives the number of observations. One, two, or three asterisks indicate significance at 90, 95, or 99-percent, respectively.

Table 5: Mean Estimates of Change in Technology $\mathscr{T}_i^{t_1,t_2}(\alpha)$ by Size Quintile ($\alpha = 0.95$)

		2006 Q1	Q2	Q3	Q4	Q5
	Q1	0.7625** (1119)	0.8646 (279)	0.9662 (39)	1.0210 (5)	1.9231*** (1)
	Q2	0.7722** (300)	0.7369 (785)	0.8385 (320)	0.9592 (38)	1.0374 (2)
1989	Q3	0.6904*** (24)	0.7311 (367)	0.8960 (756)	1.0159 (278)	1.8575 (18)
	Q4	— (0)	0.6850* (14)	0.7382* (323)	0.8455 (914)	1.3347 (193)
	Q5	— (0)	— (0)	0.3939** (5)	0.6407* (209)	1.1613 (1230)

NOTE: Each cell contains two entries; the top entry gives the mean, and the bottom entry gives the number of observations. One, two, or three asterisks indicate significance at 90, 95, or 99-percent, respectively.

Table 6: Mean Estimates of Changes in Scale Efficiency $\mathscr{S}_i^{t_1,t_2}(\alpha)$ by Size Quintile ($\alpha = 0.95$)

		2006				
		Q1	Q2	Q3	Q4	Q5
	Q1	0.7917*** (1119)	0.8840*** (279)	0.9143* (39)	0.8484*** (5)	1.1090** (1)
	Q2	0.7856*** (300)	0.8213*** (785)	0.8531*** (320)	0.8496*** (38)	0.7127*** (2)
1989	Q3	0.7914*** (24)	0.8388*** (367)	0.8589*** (756)	0.8760*** (278)	0.8461*** (18)
	Q4	— (0)	0.8375*** (14)	0.8751*** (323)	0.8791*** (914)	0.8940* (193)
	Q5	— (0)	— (0)	0.9801 (5)	0.9496 (209)	1.0818 (1230)

NOTE: Each cell contains two entries; the top entry gives the mean, and the bottom entry gives the number of observations. One, two, or three asterisks indicate significance at 90, 95, or 99-percent, respectively.

Table 7: Mean Estimates of $\mathscr{U}_i^{t_1,t_2}(\alpha)$ by Size Quintile ($\alpha = 0.95$)

		2006 Q1	Q2	Q3	Q4	Q5
	Q1	0.9669	0.9716	0.8903**	0.8773***	0.8903**
		(1119)	(279)	(39)	(5)	(1)
	Q2	1.0443	1.0984	1.0820	1.0266	1.1850
		(300)	(785)	(320)	(38)	(2)
1989	Q3	1.0891	1.0743	1.0437	1.0137	1.0401*
		(24)	(367)	(756)	(278)	(18)
	Q4	—	1.1354	1.1835*	1.2851**	1.3680***
		(0)	(14)	(323)	(914)	(193)
	Q5	—	—	1.7649**	1.7834***	2.4619***
		(0)	(0)	(5)	(209)	(1230)

NOTE: Each cell contains two entries; the top entry gives the mean, and the bottom entry gives the number of observations. One, two, or three asterisks indicate significance at 90, 95, or 99-percent, respectively.

Table 8: Mean Estimates of $\mathscr{V}_i^{t_1,t_2}(\alpha)$ by Size Quintile ($\alpha = 0.95$)

		2006 Q1	Q2	Q3	Q4	Q5
	Q1	1.1094***	1.0086	1.0074	0.9205	0.2928***
		(1119)	(279)	(39)	(5)	(1)
	Q2	1.0109	1.0247	0.9482	0.9197***	0.6173***
		(300)	(785)	(320)	(38)	(2)
1989	Q3	1.0469*	1.0124	0.8961***	0.8587***	0.4951***
		(24)	(367)	(756)	(278)	(18)
	Q4	—	0.9280	0.9034***	0.7821***	0.5010***
		(0)	(14)	(323)	(914)	(193)
	Q5	—	—	0.9678	0.7048***	0.3132***
		(0)	(0)	(5)	(209)	(1230)

NOTE: Each cell contains two entries; the top entry gives the mean, and the bottom entry gives the number of observations. One, two, or three asterisks indicate significance at 90, 95, or 99-percent, respectively.

Figure 1: Alternative Measures of Cost Efficiency

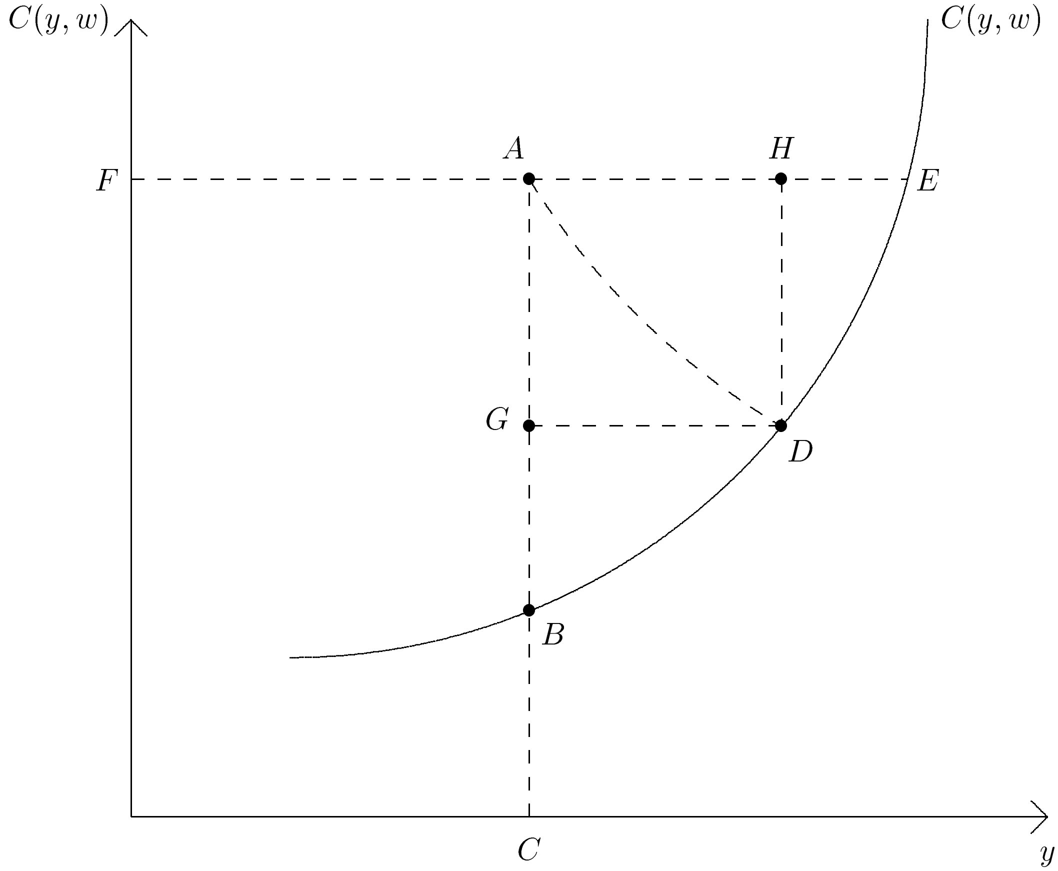
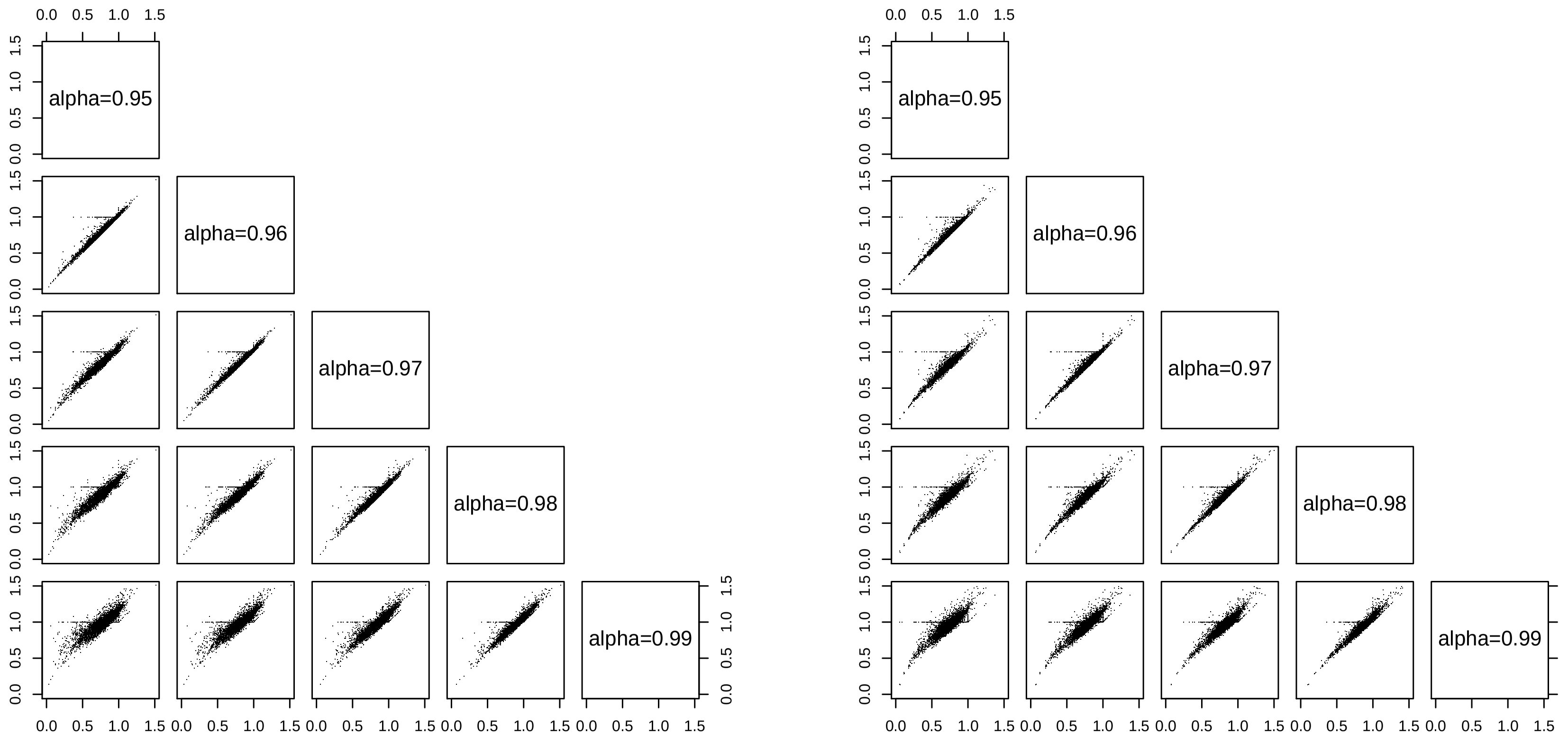


Figure 2: Contemporaneous Hyperbolic α-Quantile Efficiency Estimates for 1989 and 2006



1989 **2006**