



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



Are Credit Unions Too Small?

David C. Wheelock
And
Paul W. Wilson

Working Paper 2008-033A
<http://research.stlouisfed.org/wp/2008/2008-033.pdf>

September 2008

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

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Are Credit Unions Too Small?

DAVID C. WHEELOCK PAUL W. WILSON*

September 2008

Abstract

Since 1985, the share of U.S. depository institution assets held by credit unions has nearly doubled, and the average (inflation-adjusted) size of credit unions has increased over 600 percent. We use a non-parametric local-linear estimator to estimate a cost relationship for credit unions and derive estimates of ray-scale and expansion-path scale economies. We employ a dimension-reduction technique to reduce estimation error, and bootstrap methods for inference. We find substantial evidence of increasing returns to scale across the range of sizes observed among credit unions, suggesting that an easing of regulations on credit union membership or activities would lead to further increases in the size of credit unions.

*Wheelock: Research Department, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166–0442; wheelock@stls.frb.org. Wilson: The John E. Walker Department of Economics, 222 Sistine Hall, Clemson University, Clemson, South Carolina 29634–1309, USA; email pww@clemson.edu. This research was conducted while Wilson was a visiting scholar in the Research Department of the Federal Reserve Bank of St. Louis. We thank the CyberInfrastructure Technology Integration group at Clemson University for providing the Condor Pool used for our computations; we are especially grateful to Sebastien Goasguen for technical assistance. We have benefited from research assistance by Craig Aubuchon and Heidi Beyer; of course, any remaining errors are ours, not theirs. 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. *JEL* classification nos.: C12, C13, C14, L11, G21. Keywords: credit unions, scale economies, non-parametric, regression.

1 Introduction

Over the past three decades, advances in information-processing and communications technology (IT) and changes in regulation have had a profound impact on the environment in which commercial banks and other depository institutions operate. IT advances have enabled the development of new bank services (from automated teller machines to internet banking), financial instruments (such as various types of derivative securities), payments instruments (such as debit cards and automated clearinghouse payments), and credit evaluation and monitoring platforms.¹ The same period saw the deregulation of deposit interest rates and branch banking, the imposition of risk-based capital requirements, and numerous other regulatory changes affecting depository institutions.²

On balance, the recent changes in technology and regulation appear to have favored large institutions. The growth rates of larger banks, savings institutions and credit unions have typically exceeded those of their smaller competitors. For example, adjusted for inflation, the average U.S. commercial bank was 4.3 times larger in 2006 than the average U.S. bank in 1985.³ The average size of savings institutions and credit unions increased similarly.

Information technology has tended to favor larger institutions both because of the relatively high fixed cost of information processing equipment and software, and because these technologies have eroded some of the traditional benefits of small scale and close proximity to borrowers that enabled small lenders to out-compete larger institutions for some customers. For example, small business lending traditionally has been dominated by small, “community” banks, where close proximity and personal relationships have been important for obtaining information about the creditworthiness of potential borrowers. However, Petersen and Rajan (2002) argue that advances in IT have reduced the value of “soft” information in small business lending by making quantifiable information about potential borrowers more readily available, implying that close proximity between borrowers and lenders has become less important than in the past.

Like community banks, credit unions traditionally have operated at small scale and spe-

¹See Berger (2003) for details and analysis of the effects of technological progress on productivity growth in the banking industry and on the structure of the banking industry.

²See Spong (2000) for a summary of current U.S. banking regulations.

³In 1985, U.S. banks held an average of \$189.5 million of assets. In 2006, banks held an average of \$1,363 billion of assets (\$815.5 million in constant 1985 dollars).

cialized in “relationship” lending. Credit unions are mutual organizations that provide deposit, lending, and other financial services to a membership defined by an occupational, fraternal or other bond. A common bond is advantageous because it can reduce the cost of assessing the creditworthiness of potential borrowers and thereby facilitate unsecured lending on reasonable terms to the credit union’s members. However, as with other lenders, recent advances in information processing and communications technology have lowered the cost of acquiring “hard” information about potential borrowers, and thereby have eroded some of the advantages of small scale and common bond that traditionally enabled credit unions to provide financial services at low cost to their memberships.⁴

Despite changes that seem to favor larger depository institutions, the share of total industry assets held by credit unions—which traditionally have been much smaller in scale than other depository institutions—has increased rapidly since the early 1980s. For example, the share of industry assets held by credit unions nearly doubled between 1985 and 2006, from 3.3 percent to 6.0 percent. Much of this gain came at the expense of savings and loan associations and savings banks, which saw a decline in share from 30.1 percent to 15.9 percent. Over the same period, the share of industry assets held by commercial banks rose from 66.1 percent to 78.1 percent. As with banks and savings institutions, large credit unions have experienced faster growth in total assets than small credit unions (Goddard et al., 2002). Adjusted for inflation, the average credit union held 6.5 times more assets in 2006 than the average credit union in 1985.⁵ And, also like banks and savings institutions, the number of credit unions has declined sharply as the industry has consolidated. From a peak of 23,866 in 1969, the number of credit unions had fallen to just 8,662 in 2006. The Credit Union Membership Access Act of 1998 facilitated this consolidation by affirming the right of credit unions to accept members from unrelated groups. The number of credit unions characterized by multiple common bonds has since increased rapidly.⁶

⁴Walter (2006) notes that advances in information processing technology facilitated the emergence and expansion of national credit-reporting agencies in the 1970s, the increased use of credit cards, and the development of home-equity lines of credit. Further, Walter (2006) argues that the extension of federal deposit insurance to credit unions in the 1970s also reduced the benefits of a common bond by weakening the incentive for credit union depositors to monitor and discipline borrowers.

⁵U.S. credit unions held an average of \$84.6 million in assets in 2006 (\$50.6 million in constant 1985 dollars) versus \$7.8 million in 1985.

⁶As the advantages of a common bond were eroded, credit unions began to press for authority to expand their membership base. In 1982, the National Credit Union Administration (NCUA), the regulator of federal credit unions, ruled that a single credit union could serve employees of multiple employers even when not

The rapid consolidation and increasing average scale of credit unions have implications for U.S. banking market structure and for assessing competition in banking markets.⁷ Some research finds that agency problems are greater at larger credit unions, suggesting that credit union members may be harmed by continued growth in the average size of credit unions (e.g., Leggett and Strand, 2002).⁸ However, several studies have noted an inverse relationship between average operating expenses and credit union size (e.g., Emmons and Schmid, 1999a; Leggett and Strand, 2002; and Wilcox, 2005), and Wilcox (2006) finds that the cost advantage of large credit unions has been increasing over time. Further, Goddard et al. (2008) find that larger credit unions have more opportunities for diversification into non-traditional product lines, such as business loans, credit cards and mutual funds and that doing so has reduced the volatility of their earnings while providing their members with additional services.

This paper presents estimates of returns to scale for U.S. credit unions. We evaluate returns to scale in the context of a model of credit union cost, and unlike prior studies of credit unions, investigate whether scale economies have expanded over time in line with the industry's consolidation and the increasing average size of credit unions.⁹ Our data consist of more than 180,000 annual observations for 1989–2006 on all non-corporate credit unions (except those with missing or implausible data). We use a non-parametric, local-linear estimator to estimate our model, from which we derive estimates of returns to scale.¹⁰

all employers were engaged in the same industrial activity. Commercial banks challenged the NCUA ruling and in 1998 the Supreme Court ruled that the NCUA's interpretation was in violation of the Federal Credit Union Act, which limited membership in federally-chartered credit unions to groups having a common bond of occupation or other association. Congress responded by enacting the Credit Union Membership Access Act of 1998. See Walter (2006) for more about the early history and regulation of credit unions in the United States.

⁷See Gilbert and Zaretsky (2003) for a discussion of competitive analysis and anti-trust policy as applied to commercial banks, including the use of information about credit unions in assessing competition in banking markets. See Fried et al. (1999) and Goddard et al. (2007) for evidence on the determinants and effects of credit union mergers.

⁸Other papers investigating agency problems in credit unions include Emmons and Schmid (1999b) and Frame et al. (2003).

⁹Wheelock and Wilson (2001) report evidence of an increase in the minimum efficient scale of commercial banks between 1985 and 1994, whereas Wheelock and Wilson (1999) and Wheelock and Wilson (2008) find that on average, large banks have experienced larger increases in productivity over the past 20 years than small banks.

¹⁰Many studies derive estimates of scale economies by fitting a translog cost function across all firms in an industry. However, the translog function has been found to mis-specify many cost relationships, especially when firms are of widely varying sizes. In the appendix to this paper, we report results showing that the translog function also mis-specifies cost relationships for U.S. credit unions.

We augment the local-linear estimator with two additional kernel functions to (i) handle discrete dummy variables that indicate whether particular credit unions make commercial or real estate loans, and (ii) to incorporate a discrete time variable. Our augmentation is similar to that of Racine and Li (2004), who use a Nadarya-Watson-type kernel estimator to smooth continuous covariates. However, the local linear estimator we use to smooth along continuous dimensions has (asymptotically) less bias, with no more variance, than the Nadarya-Watson estimator.

We employ three different bandwidth parameters in our estimation—one for the continuous covariates (after pre-whitening and using a dimension reduction technique to mitigate the effects of the well-known curse of dimensionality), another for the two binary dummy variables, and a third for the discrete time variable. In addition, we use a bias-corrected bootstrap for inference. Only recently has bootstrapping and optimization of three bandwidths with more than 180,000 observations using least-squares cross validation become computationally feasible at low cost.¹¹

Estimating both ray scale economies and expansion-path scale economies, we find substantial evidence of increasing returns to scale throughout the range of sizes observed among credit unions, except among the smallest credit unions. The smallest credit unions are considerably smaller than the smallest commercial banks operating in the United States, and may experience constant or even decreasing returns before they become large enough to justify purchase and use of IT systems and perhaps other equipment, facilities, etc. Nevertheless, despite a large increase in the average size of U.S. credit unions, we find that a majority of even the largest 20 percent of credit unions operate under increasing returns to scale. Hence, if regulations on credit union membership or activities were eased further, as some have proposed, the average size of credit unions would seem likely to increase further

¹¹We use a high-throughput Condor pool operated by Clemson University for our computations. Condor systems are a form of grid computing, and consist of a scheduler that sends jobs to machines in the pool when they are idle, thereby harvesting (or scavenging; hence the name “Condor”) otherwise unused CPU cycles. Our bandwidth optimization and other computational problems are ideally suited for high-throughput computing systems (as opposed to high-performance computing systems such as vector machines and massively parallel machines, which involve considerable expense) since the problems are easily divided into independent pieces that can run on different machines and which need no communication between tasks until the very end. Additional details on the development of Condor systems are available at <http://www.cs.wisc.edu/condor/>; details on the Condor system operated by Clemson University are available at <http://ccit.clemson.edu/support/research/>.

as credit unions sought to exploit scale economies.¹²

The remainder of the paper unfolds as follows. In the next section, we describe a model of credit union costs. Section 3 presents details of our estimation strategy. Results are presented in Section 4, and our conclusions are discussed in Section 5.

2 A Model of Credit Union Costs

To estimate scale economies, we must first specify a model of credit union costs. Credit unions use a number of inputs to produce a wide range of services; in studies of credit union performance, limited data and, in the case of non-parametric approaches, limits on the number of dimensions that can reasonably be examined, force researchers to employ simplified models.

Following Frame et al. (2003) and Frame and Coelli (2001), we model credit unions as service providers that seek to minimize non-interest costs subject to the prices of labor and capital inputs, the prevailing production technology, and the level and types of output they produce.¹³ Table 1 lists the variables in our model and how each is defined in terms of call report items.¹⁴ We specify four output quantities: real estate loans (Y_1), commercial loans (Y_2), consumer loans (Y_3), and investments (Y_4). Further, following Frame et al. (2003), we treat the average interest rates on deposits (Y_5) and loans (Y_6) as additional outputs to capture the price dimension of service to credit union members. Also like Frame et al. (2003), our model includes the price of capital (W_1) and the price of labor (W_2) faced by each credit union. Finally, our model includes a discrete time variable (T) for each year in our data, and two dummy variables, D_1 and D_2 , that identify individual credit unions that make real estate or commercial loans, with

$$D_1 = \begin{cases} 1 & \text{if } Y_1 > 0; \\ 0 & \text{otherwise,} \end{cases} \quad (2.1)$$

¹²Pending legislation in Congress (HR 5519) would enable credit unions to lend to nonmembers in “underserved” areas and grant credit unions more flexibility in offering member business loans.

¹³See also Bauer (2008), Fried et al. (1993), Fried et al. (1999) and Smith (1984).

¹⁴Call report data for individual credit unions are available from the National Credit Union Administration (www.ncua.gov). We obtained our data from the Federal Reserve.

and

$$D_2 = \begin{cases} 1 & \text{if } Y_2 > 0; \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

Precise definitions of all variables are given in Table 1.

Table 2 reports summary statistics for the variables in our model, as well as for total operating cost (expenditures on physical capital and labor inputs) and total assets. Our data consist of 184,279 annual (year-end) observations for all state- and federally-chartered credit unions during 1989–2006; numbers of observations for each of 18 years are given in Table 3.¹⁵

The variables defined above and listed in Table 1 suggest a mapping

$$(Y_1, \dots, Y_6, W_1, W_2, T) \rightarrow C. \quad (2.3)$$

However, in order to impose homogeneity of the cost function with respect to input prices, it is convenient to divide both C and W_1 by W_2 . In addition, large numbers of observations on Y_2 and Y_3 are equal to zero as indicated by the summary statistics in Table 2. Consequently, we combine these outputs with Y_1 by using the sum $(Y_1 + Y_2 + Y_3)$ and the dummy variables D_1 and D_2 in our estimation. Then the mapping in (2.3) suggests a regression function

$$\left(\frac{C}{W_2} \right) = C(\mathbf{y}, \mathbf{w}) + \varepsilon, \quad (2.4)$$

where $\mathbf{y} = [(Y_1 + Y_2 + Y_3) \ Y_4]$, $\mathbf{w} = [Y_5 \ Y_6 \ W_1/W_2 \ T \ D_1 \ D_2]$, and ε is a random error term with $E(\varepsilon) = 0$. Given that the expectation of ε equals 0, $C(\mathbf{y}, \mathbf{w}) = E(C/W_2 \mid \mathbf{y}, \mathbf{w})$ is a conditional mean function that can be estimated by various regression techniques.

Now consider a particular point $(\mathbf{y}_0, \mathbf{w}_0)$ in the space of (\mathbf{y}, \mathbf{w}) . The set of points $\mathfrak{R}_0 = \{(\theta\mathbf{y}, \mathbf{w}) \mid \theta \in (0, \infty)\}$ comprises a ray along which the outputs $(Y_1 + Y_2 + Y_3)$ and Y_4 are produced in constant proportion to each other. Ray scale economies can be evaluated by examining how expected cost varies along this ray, providing insight into returns to scale along the ray \mathfrak{R}_0 . Returns to scale are frequently measured in terms of elasticities; the

¹⁵Observations where either $Y_1 < 0$, $Y_2 < 0$, $Y_3 < 0$, $Y_4 < 0$, $(Y_1 + Y_2 + Y_3) = 0$, $Y_5 \notin (0, 1)$, $Y_6 \notin (0, 1)$, $W_1 \notin (0, 1)$, or $W_2 \leq 0$ were deleted from our analysis. Observations that were deleted either reported incorrect values for one or more variables, or make no loans. In addition, we omit 397 observations for corporate credit unions, and consider only retail credit unions. These criteria cause a total of 18,307 observations to be deleted; most deletions are made due to implausible observations for one or both of the price variables.

elasticity of cost (with respect to \mathbf{y}) at a particular point (\mathbf{y}, \mathbf{w}) along the ray \mathfrak{R}_0 is given by

$$\eta(\mathbf{y}, \mathbf{w}) \equiv \left. \frac{\partial \log C(\theta \mathbf{y}, \mathbf{w})}{\partial \log \theta} \right|_{\theta=1} = \sum_j \frac{\partial \log C(\mathbf{y}, \mathbf{w})}{\partial \log y_j}, \quad (2.5)$$

where j indexes the elements of \mathbf{y} . The elasticity in (2.5) is the multi-product analog of marginal cost divided by average cost on the ray \mathfrak{R}_0 , with $\eta(\mathbf{y}, \mathbf{w}) (<, =, >) 1$ implying (increasing, constant, decreasing) returns to scale as outputs in \mathbf{y} are expanded along the ray \mathfrak{R}_0 . Credit unions for which $\eta(\mathbf{y}, \mathbf{w}) \neq 1$ are not competitively viable; if credit unions were subject to the normal rules of competitive behavior, either a smaller or a larger firm could drive a credit union with $\eta(\mathbf{y}, \mathbf{w}) \neq 1$ from a competitive market.

The measure defined in (2.5) requires estimation of derivatives of the cost function. We employ fully non-parametric estimation methods in this paper as discussed below in Section 3. Since non-parametric estimates of derivatives of a function are typically noisier than estimates of the function itself,¹⁶ we define the ratio

$$\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0) \equiv \frac{C(\theta \mathbf{y}_0, \mathbf{w}_0)}{\theta C(\mathbf{y}_0, \mathbf{w}_0)}. \quad (2.6)$$

It is straightforward to show that

$$\frac{\partial \mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)}{\partial \theta} \begin{matrix} \leq 0 \\ \geq 0 \end{matrix} \iff \eta(\mathbf{y}_0, \mathbf{w}_0) \begin{matrix} \leq 1 \\ \geq 1 \end{matrix}; \quad (2.7)$$

i.e., $\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ is decreasing (constant, increasing) in θ if returns to scale are increasing (constant, decreasing) at $(\theta \mathbf{y}_0, \mathbf{w}_0)$ along the ray \mathfrak{R}_0 passing through $(\mathbf{y}_0, \mathbf{w}_0)$. In addition, $\mathcal{S}(1 \mid \mathbf{y}_0, \mathbf{w}_0) = 1$ by definition. Thus, *ray scale economies* (RSE) along a ray \mathfrak{R}_0 can be examined by estimating $C(\mathbf{y}_0, \mathbf{w}_0)$ and $C(\theta \mathbf{y}_0, \mathbf{w}_0)$ for various values of θ , and using confidence bands to determine whether $\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ is downward or upward sloping.

Of course, not all credit unions are located along the ray \mathfrak{R}_0 ; in fact, it is conceivable that *none* are located along \mathfrak{R}_0 . RSE is merely a convenient way to summarize results on scale economies, but may be misleading if most credit unions are located “far” from \mathfrak{R}_0 . As an alternative to RSE, we also consider scale economies along each credit union’s expansion path, holding the mix of outputs in \mathbf{y} constant for each credit union. Consider a credit union operating at the point $(\mathbf{y}_0, \mathbf{w}_0)$, with cost $C(\mathbf{y}_0, \mathbf{w}_0)$. Let γ be a small positive number, say

¹⁶This is particularly true for the present case where we would require derivatives in several dimensions; in addition, bandwidth selection becomes problematic when estimating derivatives in more than one dimension.

0.05 and consider how cost changes as we move from $((1 - \gamma)\mathbf{y}_0, \mathbf{w}_0)$ to $((1 + \gamma)\mathbf{y}_0, \mathbf{w}_0)$; along this path, the output mix remains constant in the sense that relative proportions are maintained. Now let $\theta(1 - \gamma)\mathbf{y}_0 = (1 + \gamma)\mathbf{y}_0$; then $\theta = (1 + \gamma)/(1 - \gamma)$.

Expansion-path scale economies (EPSE) for the credit union operating at $(\mathbf{y}_0, \mathbf{w}_0)$ are measured by

$$\begin{aligned} \mathcal{E}_0 &= \frac{C(\theta(1 - \gamma)\mathbf{y}_0, \mathbf{w}_0)}{\theta C((1 - \gamma)\mathbf{y}_0, \mathbf{w}_0)} \\ &= \frac{C((1 + \gamma)\mathbf{y}_0, \mathbf{w}_0)}{\left(\frac{1 + \gamma}{1 - \gamma}\right) C((1 - \gamma)\mathbf{y}_0, \mathbf{w}_0)}. \end{aligned} \tag{2.8}$$

It is easy to show that the credit union operating at $(\mathbf{y}_0, \mathbf{w}_0)$ experiences (decreasing, constant, increasing) returns to scale along the path from $((1 - \gamma)\mathbf{y}_0, \mathbf{w}_0)$ to $((1 + \gamma)\mathbf{y}_0, \mathbf{w}_0)$ as $\mathcal{E}_0(>, =, <)1$. Our measure \mathcal{E}_0 gives an indication of returns to scale faced by a particular credit union along the path from the origin through the credit union's *observed* output vector, starting at a level equal to 95-percent of the quantities in \mathbf{y}_0 and continuing to a level equal to 105-percent of the quantities in \mathbf{y}_0 .

Both our RSE and EPSE measures are defined in terms of a credit union's cost function. In the next section, we discuss a strategy for estimating the cost function non-parametrically, which in turn allows us to estimate, and make inference about, our measures of scale economies.

3 Estimation Strategy

Various approaches exist for estimating regression functions (i.e., conditional mean functions) such as the one defined above in (2.4). In parametric approaches, a translog specification is often used for the conditional mean function. It is important to note, however, that because the translog cost function is merely a quadratic specification in log-space, the variety of shapes the cost function is permitted to take is limited. Further, because the translog is derived from a Taylor expansion of the cost function around the mean of the data, it makes little sense to use the translog specification to attempt inference about returns to scale over units of widely varying size. We find that the translog specification is easily rejected by our data; see the Appendix for details.

Rejection of the translog functional form is hardly surprising. Several studies have noted the problem; see, for example, Guilkey et al. (1983) and Chalfant and Gallant (1985) for Monte Carlo evidence, and Cooper and McLaren (1996) and Banks et al. (1997) for empirical evidence involving consumer demand. Still others have found a similar problem while estimating cost functions for hospitals (Wilson and Carey, 2004) and for US commercial banks (e.g., McAllister and McManus, 1993; Mitchell and Onvural, 1996; and Wheelock and Wilson, 2001); both hospitals and banks vary widely in terms of size, as do credit unions. The problem points to the use of non-parametric estimation methods. Although non-parametric methods are less efficient in a statistical sense than parametric methods when the *true* functional form is known, non-parametric estimation (asymptotically) avoids the risk of specification error when the true functional form is unknown, which, to our knowledge, is the case here.

We use a fully non-parametric, local-linear estimator augmented to handle discrete covariates. Fan and Gijbels (1996, chapter 1) and Härdle and Linton (1999) give nice descriptions of non-parametric regression and the surrounding issues. Non-parametric regression models may be viewed as infinitely parameterized; as such, any parametric regression model (such as the translog cost function) is nested within a non-parametric regression model. Clearly, adding more parameters to a parametric model affords greater flexibility. Non-parametric regression models represent the limiting outcome of adding additional parameters.

Several possibilities for non-parametric regression exist. Orthogonal series estimators based on the ideas of Szegø (1959) and Gallant (1981, 1982) involve representing the conditional mean function by an infinite Fourier series and using orthogonal polynomials (e.g., Laguerre or Legendre polynomials) or other functions (e.g., transcendental functions or Muntz-Satz expansions) to represent the basis functions, and have been used in studies of bank costs and elsewhere. One must choose a truncation point for the Fourier series; cross-validation and other methods (e.g., Eastwood, 1991) may be used.¹⁷ As a practical matter, in the

¹⁷Published papers using this approach have typically not optimized the number of terms according to such criteria. In addition, Barnett et al. (1991) note that “the basis functions with which Gallant’s model seeks to span the neoclassical function space are sines and cosines, despite the fact that such trigonometric functions are periodic and hence are far from neoclassical. In other words, the basis functions, which should be dense in the space to be spanned do not themselves even lie within that space.” Instead of trigonometric functions, one could use as the basis functions members of a family of orthogonal polynomials (e.g., Laguerre or Legendre polynomials), but the problems of determining the optimal number of terms, and using these in a non-linear, maximum-likelihood framework, remain.

present multivariate setting with a large number of observations, this method would incur the numerically challenging problem of inverting very large moment matrices. Our local-linear estimator avoids these problems.

Most non-parametric regression methods suffer from the well-known curse of dimensionality, a phenomenon that causes rates of convergence to become slower, and estimation error to increase dramatically, as the number of continuous right-hand side variables increases (the presence of discrete dummy variables does not affect the rate of convergence of our estimator). To help mitigate this problem, we use a dimension-reduction technique based on principal components. The idea is to trade a relatively small amount of information in the data for a reduction in dimensionality that will have a large (and favorable) impact on estimation error.

For an $(n \times 1)$ vector \mathbf{U} define the function

$$\psi(\mathbf{U}) \equiv (\mathbf{U} - n^{-1}\mathbf{i}'\mathbf{U}) [n^{-1}\mathbf{U}'\mathbf{U} - n^{-2}\mathbf{U}'\mathbf{i}\mathbf{i}'\mathbf{U}]^{-1/2} \quad (3.1)$$

where \mathbf{i} denotes an $(n \times 1)$ vector of 1s. The function $\psi(\cdot)$ standardizes a variable by subtracting its sample mean and then dividing by its sample standard deviation. Next, let \mathbf{A} be an $(n \times 5)$ matrix with columns $\psi(\log(Y_1 + Y_2 + Y_3))$, $\psi(\log(1 + Y_4))$, $\psi(Y_5)$, $\psi(Y_6)$, and $\psi\left(\log\left(\frac{W_1}{W_2}\right)\right)$. Here, the three loan variables Y_1 , Y_2 , and Y_3 are summed since, as noted previously in Section 2, many credit unions make neither real estate nor commercial loans. The dummy variables D_1 and D_2 retain some information that would otherwise be lost by identifying those credit unions that are observed to hold either real estate or commercial loans.

Let \mathbf{E} be the (5×5) matrix whose columns are the eigenvectors of the (5×5) correlation matrix whose elements are the Pearson correlation coefficients for pairs of columns of \mathbf{A} . Let λ_k be the eigenvalue corresponding to the k th eigenvector in the k th column of \mathbf{E} , where the columns of \mathbf{E} , and hence the eigenvalues, have been sorted so that $\lambda_1 \geq \dots \geq \lambda_5$. Then set $\mathbf{P} = \mathbf{A}\mathbf{E}$. The matrix \mathbf{P} has dimensions $(n \times 5)$, and its columns are the principal components of \mathbf{A} . It is well-known that principal component vectors are orthogonal. Moreover, it is also well-known that for each $k \in \{1, 2, \dots, 5\}$, the quantity

$$\phi_k = \frac{\sum_{j=1}^k \lambda_j}{\sum_{\ell=1}^5 \lambda_\ell} \quad (3.2)$$

represents the proportion of the independent linear information in \mathbf{A} that is contained in the first k principal components, i.e., the columns of \mathbf{P} .

Using our data, we find $\phi_k = 0.5012, 0.7665, 0.8986, 0.9757,$ and 1.0 for $k = 1, \dots, 5$ respectively. Consequently, we use the first four principal components, omitting the last one, in our non-parametric estimation of the credit union cost function. In doing so, we sacrifice a relatively small amount of information—2.43 percent of the independent linear information in the sample—in order to reduce the dimensionality of our estimation problem by one dimension in the space of the continuous covariates. Given the curse of dimensionality, this seems a good tradeoff.

Let $\mathbf{P}_{\cdot k}$ denote the k th column of the principal component matrix \mathbf{P} and define

$$\varphi(\mathbf{P}_{\cdot k}) \equiv \mathbf{P}_{\cdot k} \left[n^{-1} \mathbf{P}'_{\cdot k} \mathbf{P}_{\cdot k} - n^{-2} \mathbf{P}'_{\cdot k} \mathbf{i} \mathbf{i}' \mathbf{P}_{\cdot k} \right]^{-1/2}. \quad (3.3)$$

The transformation $\varphi(\mathbf{P}_{\cdot k})$ has (constant) unit variance. Next, let \mathbf{z}_i represent the row vector containing the i th observations on $\varphi(\mathbf{P}_{\cdot 1}), \varphi(\mathbf{P}_{\cdot 2}), \varphi(\mathbf{P}_{\cdot 3}),$ and $\varphi(\mathbf{P}_{\cdot 4})$. We can now write our model as a regression equation given by

$$\mathcal{C}_i = m(\mathbf{z}_i, T_i, D_{i1}, D_{i2}) + \varepsilon_i \quad (3.4)$$

where the subscript i indexes observations, $\mathcal{C}_i = \psi \left(\log \left(\frac{C_i}{W_{i2}} \right) \right)$, ε_i is a random error term with $E(\varepsilon_i) = 0$, $\text{VAR}(\varepsilon_i) = \sigma^2(\mathbf{z}_i)$, T_i represents the i th observation on the time variable, and D_{i1} , and D_{i2} represent the i th observations on D_1 and D_2 . The function $m(\mathbf{z}_i, T_i, D_{i1}, D_{i2}) = E(\mathcal{C}_i \mid \mathbf{z}_i, T_i, D_{i1}, D_{i2})$ is a conditional mean function, and can be estimated by non-parametric methods. Moreover, since the transformation from (C/W_2) to \mathcal{C} can be inverted, given an estimated value $\widehat{m}(\mathbf{z}, T, D_1, D_2)$, straightforward algebra leads to an estimate

$$\widehat{C}(\mathbf{y}, \mathbf{w}) = \exp \left[\psi^{-1} \left(\widehat{m}(\mathbf{z}, T, D_1, D_2) \right) \right]. \quad (3.5)$$

To estimate returns to scale for credit unions, we need merely estimate the measure $\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ defined earlier by replacing $C(\mathbf{y}_0, \mathbf{w}_0)$ and $C(\theta \mathbf{y}_0, \mathbf{w}_0)$ on the right-hand side of (2.6) with estimates $\widehat{C}(\mathbf{y}_0, \mathbf{w}_0)$ and $\widehat{C}(\theta \mathbf{y}_0, \mathbf{w}_0)$ obtained from (3.5). Note that if $\widehat{m}(\mathbf{z}, T, D_1, D_2)$ were an unbiased estimator of $m(\mathbf{z}, T, D_1, D_2)$, then $\widehat{C}(\mathbf{y}, \mathbf{w})$ would not be unbiased due to the non-linear transformation in (3.5). Moreover, even if an unbiased estimator of $C(\mathbf{y}, \mathbf{w})$ were available, substituting unbiased estimators on the right-hand side of (2.6) would not

yield an unbiased estimator of $Sc(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ for similar reasons. Since our non-parametric local-linear estimator is (weakly) consistent, but asymptotically biased, there seems little cost in making the non-linear transformation in (3.5). In addition, we employ a bias-corrected bootstrap method when estimating confidence intervals for our returns-to-scale measures.

In order to estimate the conditional mean function in (3.4), suppose (for the moment) that the time variable T and the binary dummy variables D_1, D_2 are irrelevant (i.e., do not influence the value of the conditional mean function $m(\mathbf{z}, T, D_1, D_2)$), so that we can write the conditional mean function on the right-hand side of (3.4) as $m(\mathbf{z})$. Both the Nadarya-Watson (Nadarya, 1964; Watson, 1964) kernel estimator and the local linear estimator are special cases of local polynomial estimators; with the local linear estimator, the local polynomial is of order 1, while with the Nadarya-Watson estimator the local polynomial is of order 0. The local linear estimator has less asymptotic bias, but the same asymptotic variance, as the Nadarya-Watson estimator.

To illustrate the local-linear estimator, momentarily ignore the discrete covariates in (3.4) and write the conditional mean function as $m_*(\mathbf{z})$. The local linear estimator follows from a first-order Taylor expansion of $m_*(\mathbf{z})$ in a neighborhood of an arbitrary point \mathbf{z}_0 :

$$m_*(\mathbf{z}) \approx m_*(\mathbf{z}_0) + \frac{\partial m_*(\mathbf{z}_0)}{\partial \mathbf{z}}(\mathbf{z} - \mathbf{z}_0). \quad (3.6)$$

This suggests estimating the conditional mean function at \mathbf{z}_0 by solving the locally weighted least squares regression problem

$$[\hat{\alpha}_0 \quad \hat{\boldsymbol{\alpha}}]' = \underset{\alpha_0, \boldsymbol{\alpha}}{\operatorname{argmin}} \sum_{i=1}^n [\mathcal{C}_i - \alpha_0 - (\mathbf{z}_i - \mathbf{z}_0)\boldsymbol{\alpha}]^2 K(|\mathbf{H}|^{-1}(\mathbf{z}_i - \mathbf{z}_0)) \quad (3.7)$$

where $K(\cdot)$ is a piece-wise continuous multivariate kernel function satisfying $\int_{\mathbb{R}^\ell} K(\mathbf{u}) d\mathbf{u} = 1$ and $K(\mathbf{u}) = K(-\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^\ell$; \mathbf{H} is an $\ell \times \ell$ matrix of bandwidths; α_0 is a scalar, and $\boldsymbol{\alpha}$ is an ℓ -vector.

The solution to the least squares problem in (3.7) is

$$[\hat{\alpha}_0 \quad \hat{\boldsymbol{\alpha}}]' = (\mathbf{Z}'\boldsymbol{\Phi}\mathbf{Z})^{-1} \mathbf{Z}'\boldsymbol{\Phi}\mathcal{C}, \quad (3.8)$$

where $\mathcal{C} = [\mathcal{C}_1 \quad \dots \quad \mathcal{C}_n]'$, $\boldsymbol{\Phi} = \operatorname{diag}[K(|\mathbf{H}|^{-1}(\mathbf{z}_i - \mathbf{z}_0))]$, and \mathbf{Z} is an $n \times (\ell + 1)$ matrix with i th row given by $[1 \quad (\mathbf{z}_i - \mathbf{z}_0)']$. The fitted value $\hat{\alpha}_0$ provides an estimate $\hat{m}_*(\mathbf{z}_0)$ of

the conditional mean function $m_*(\mathbf{z}_0)$ at an arbitrary point \mathbf{z}_0 .¹⁸

Introduction of the binary dummy variables D_{i1} and D_{i2} into the analysis requires some modification. One possibility is to split the sample into $2^2 = 4$ subsamples depending on the values of the discrete variables, and then analyze each group separately while treating time as a continuous variable. Unfortunately, some of the resulting subsamples may be very small—as is the case in our application (see Table 1). Moreover, to the extent that each subsample may contain some information that would be useful in estimation on the other subsamples, this approach does not make efficient use of the data.

With the local linear estimator, the introduction of discrete variables can be accommodated by modifying the weights in Φ . The idea involves smoothing across time periods as well as over the 2^2 categories defined by the two binary dummy variables, and to let the data determine how much smoothing is appropriate. Aitchison and Aitken (1976) discuss the use of a discrete kernel for discrimination analysis. Bierens (1987) and Delgado and Mora (1995) suggest augmenting the Nadarya-Watson estimator with a discrete kernel, and prove that the estimator remains consistent and asymptotically normal. Racine and Li (2000) establish convergence rates for the Nadarya-Watson estimator with mixed continuous-discrete data; the rate with continuous and discrete covariates is the same as the rate with the same number of continuous variables, but no discrete variables. Thus, introduction of discrete covariates does not exacerbate the curse of dimensionality, at least in the limit.

It is straightforward to extend these ideas to the local linear estimator. Let \mathbf{u}_i represent a vector of observations on k binary dummy variables, and consider an arbitrary Bernoulli vector \mathbf{u}_0 of length k . Then let $\delta(\mathbf{u}_i, \mathbf{u}_0) = (\mathbf{u}_i - \mathbf{u}_0)'(\mathbf{u}_i - \mathbf{u}_0)$, and define the discrete kernel function

$$G_1(\mathbf{u}_i | \mathbf{u}_0, \lambda_1) = \lambda_1^{k-\delta(\mathbf{u}_i, \mathbf{u}_0)}(1 - \lambda_1)^{\delta(\mathbf{u}_i, \mathbf{u}_0)} \quad (3.9)$$

where $\lambda_1 \in [\frac{1}{2}, 1]$ is a bandwidth parameter.

Note that $\lim_{\lambda_1 \rightarrow 1} G_1(\mathbf{u}_0 | \mathbf{u}_i, \lambda_1)$ equals either 1 or 0, depending on whether $\mathbf{u}_0 = \mathbf{u}_i$ or

¹⁸The fitted values in $\hat{\alpha}$ provide estimates of elements of the vector $\partial m(\mathbf{z}_0)/\partial \mathbf{z}$. However, if the object is to estimate first derivatives, mean-square error of the estimates can be reduced by locally fitting a quadratic rather than a linear expression (see Fan and Gijbels, 1996 for discussion); this increases computational costs, which are already substantial for the local linear fit. Moreover, determining the optimal bandwidths becomes more difficult and computationally more burdensome for estimation of derivatives. See Härdle (1990, pp. 160–162) for discussion of some of the issues that are involved with bandwidth selection for derivative estimation.

$\mathbf{u}_0 \neq \mathbf{u}_i$, respectively. In this case, estimation yields the same results as would be the case if the sample were split into 2^2 groups suggested by the dummy variables, with estimation performed independently on each of the 2^2 subsamples. Alternatively, if $\lambda_1 = \frac{1}{2}$, then $G_1(\mathbf{u}_0 | \mathbf{u}_i, \lambda_1) = 1$ regardless of whether $\mathbf{u}_0 = \mathbf{u}_i$ or $\mathbf{u}_0 \neq \mathbf{u}_i$; in this case, there is complete smoothing over the 2^2 categories, and including the dummy variables has no effect relative to the case where they are ignored.

Next, consider the ordered, categorical variable T_i which takes values in the set $\mathbb{T} = \{1, 2, \dots, T_{\max}\}$, and let $T_0 \in \mathbb{T}$. Define a kernel function

$$G_2(T_i | T_0, \lambda_2) = \lambda_2^{|T_i - T_0|} \quad (3.10)$$

where $\lambda_2 \in [0, 1]$ is a third bandwidth parameter. For $\lambda_2 < 1$, as the difference $|T_i - T_0|$ increases, $G_2(T_i | T_0, \lambda_2)$ becomes smaller. In other words, for $\lambda_2 < 1$, observations from time periods farther from T_0 receive less weight than observations from time periods that are closer to T_0 .

Because the continuous data have been pre-whitened so that the variables in \mathbf{z} appearing in (3.4) are uncorrelated and have unit variance due to the principal-components transformation and the transformation in (3.3), it is natural to specify the kernel function $K(\cdot)$ as an ℓ -variate spherically symmetric Epanechnikov kernel with a single, scalar bandwidth h ; i.e.,

$$K(\mathbf{u}) = \frac{\ell(\ell + 2)}{2S_\ell} (1 - \mathbf{u}\mathbf{u}') I(\mathbf{u}\mathbf{u}' \leq 1) \quad (3.11)$$

where $I(\cdot)$ is the indicator function, $S_\ell = 2\pi^{\ell/2}/\Gamma(\ell/2)$, $\Gamma(\cdot)$ denotes the gamma function, and $\mathbf{u} = h^{-\ell}(\mathbf{z}_i - \mathbf{z}_0)$. The spherically symmetric Epanechnikov kernel is optimal in terms of asymptotic minimax risk; see Fan et al. (1997) for details and a proof.

Incorporating the discrete covariates, an estimate $\hat{m}(\mathbf{z}_0, T_0, D_{01}, D_{02})$ of the conditional mean function in (3.4) at an arbitrary point $(\mathbf{z}_0, T_0, D_{01}, D_{02}) \in \mathbb{R}^\ell \times \mathbb{T} \times \{0, 1\}^2$ is given by $\hat{\alpha}_0$ obtained from

$$\begin{aligned} [\hat{\alpha}_0 \ \hat{\boldsymbol{\alpha}}]' = \operatorname{argmin}_{\alpha_0, \boldsymbol{\alpha}} \sum_{i=1}^n [\mathcal{C}_i - \alpha_0 - (\mathbf{z}_i - \mathbf{z}_0)\boldsymbol{\alpha}]^2 K(|\mathbf{H}|^{-1}(\mathbf{z}_i - \mathbf{z}_0)) \\ G_1(\mathbf{w}_i | \mathbf{w}_0, \lambda_1) G_2(T_i | T_0, \lambda_2), \end{aligned} \quad (3.12)$$

where $T_0 \in \{1, 2, \dots, 18\}$ and \mathbf{w}_0 is a (2×1) Bernoulli vector. The solution to the

least-squares problem in (3.12) is found by using

$$\Phi = \text{diag} \left[K(h^{-\ell}(\mathbf{z}_i - \mathbf{z}_0)G_1(\mathbf{w}_i | \mathbf{w}_0, \lambda_1)G_2(T_i | T_0, \lambda_2)) \right], \quad (3.13)$$

in (3.8). Here, the determinant of the bandwidth matrix \mathbf{H} has been replaced by h^ℓ ; since our principal components transformation pre-whitens the data, and the principal components are orthogonal, we use the same bandwidth in each direction.

To implement our estimator, we must choose values for the bandwidths h , λ_1 , and λ_2 . For the discrete data, we employ (globally) constant bandwidths, while for the continuous data we use an adaptive, nearest-neighbor bandwidth. We define h for any particular point $\mathbf{z}_0 \in \mathbb{R}^\ell$ as the maximum Euclidean distance between \mathbf{z}_0 and the κ nearest points in the observed sample $\{\mathbf{z}_i\}_{i=1}^n$, $\kappa \in \{2, 3, 4, \dots\}$. Thus, the bandwidth h is determined by κ , and varies depending on the density of the continuous explanatory variables locally around the point $\mathbf{z}_0 \in \mathbb{R}^\ell$ at which the conditional mean function is estimated. This results in a relatively large value for h where the data are sparse (and where more smoothing is required), and smaller values of h in regions where the data are relatively dense (where less smoothing is needed). The discrete kernels in (3.13) in turn give more (or less) weight to observations among the κ nearest neighbors that are close (or far) away along the time dimension, or that are in the same (or different) category determined by the combination of binary dummy variables.

Note that we are not using a nearest-neighbor *estimator*, but rather a nearest-neighbor bandwidth. Our bandwidth is used inside a kernel function, and the kernel function integrates to unity. Loftsgaarden and Quesenberry (1965) use this approach in the density estimation context to avoid nearest-neighbor density estimates (as opposed to bandwidths) that do not integrate to unity (see Pagan and Ullah, 1999, pp. 11-12 for additional discussion). Fan and Gijbels (1994; 1996, pp. 151–152) discuss nearest neighbor bandwidths in the regression context.

As a practical matter, we set $\kappa = [\lambda_0 n]$, where n represents the sample size and $[a]$ denotes the integer part of a . We optimize the choice of values for the bandwidth parameters by minimizing the least-squares cross-validation function; i.e., we select values

$$[\lambda_0 \quad \lambda_1 \quad \lambda_2]' = \underset{\lambda_0, \lambda_1, \lambda_2}{\text{argmin}} \sum_{i=1}^n [C_i - \hat{m}_{-i}(\mathbf{z}_i, T_i, D_{i1}, D_{i2})]^2, \quad (3.14)$$

where $\widehat{m}_{-i}(\mathbf{z}_i, T_i, D_{i1}, D_{i2})$ is computed the same way as $\widehat{m}(\mathbf{z}_i, T_i, D_{i1}, D_{i2})$, except that the i th diagonal element of Ψ is replaced with zero. The least-squares cross validation function approximates the part of mean integrated square error that depends on the bandwidths.¹⁹

Once appropriate values of the bandwidth parameters have been selected, the conditional mean function can be estimated at any point $(\mathbf{z}_0, T_0, D_{01}, D_{02}) \in \mathbb{R}^\ell \times \mathbb{T} \times \{0, 1\}^k$. We then estimate the RSE and EPSE measures defined in (2.6) and (2.8) by replacing the cost terms with estimates obtained from the relation (3.5). To make inferences about RSE and EPSE, we use the wild bootstrap proposed by Härdle (1990) and Härdle and Mammen (1993).²⁰ The wild bootstrap is used to obtain bootstrap estimates $\widehat{m}_b^*(\cdot)$, which are substituted into (2.6) and (2.8) to obtain bootstrap values $\widehat{\mathcal{S}}_b^*$ and $\widehat{\mathcal{E}}_b^*$ for particular values of \mathbf{z} and D_1, D_2 , with $b = 1, \dots, B$.

To make inference about \mathcal{S} , we use the bias-correction method described by Efron and Tibshirani (1993). In particular, we estimate $(1 - \alpha) \times 100$ -percent confidence intervals by $(\widehat{\mathcal{S}}^{*(\alpha_1)}, \widehat{\mathcal{S}}^{*(\alpha_2)})$, where $\widehat{\mathcal{S}}^{*(\alpha)}$ denotes the α -quantile of the bootstrap values $\widehat{\mathcal{S}}_b^*$, $b = 1, \dots, B$, and

$$\alpha_1 = \Phi \left(\widehat{z}_0 + \frac{\widehat{z}_0 + z^{(\alpha/2)}}{1 - \widehat{z}_0 + z^{(\alpha/2)}} \right), \quad (3.15)$$

$$\alpha_2 = \Phi \left(\widehat{z}_0 + \frac{\widehat{z}_0 + z^{(1-\alpha/2)}}{1 - \widehat{z}_0 + z^{(1-\alpha/2)}} \right), \quad (3.16)$$

$\Phi(\cdot)$ denotes the standard normal distribution function, $z^{(\alpha)}$ is the $(\alpha \times 100)$ -th percentile of the standard normal distribution, and

$$\widehat{z}_0 = \Phi^{-1} \left(\frac{\#\{\widehat{\mathcal{S}}_b^* < \widehat{\mathcal{S}}\}}{B} \right), \quad (3.17)$$

with $\Phi^{-1}(\cdot)$ denoting the standard normal quantile function (e.g., $\Phi^{-1}(0.95) \approx 1.6449$).

For the case of RSE, we sort the values in $\left\{ \left(\widehat{\mathcal{S}}_b^* - \widehat{\mathcal{S}} \right) \right\}_{b=1}^B$ by algebraic value, delete $(\frac{\alpha}{2} \times 100)$ -percent of the elements at either end of this sorted array, and denote the lower and upper endpoints of the remaining, sorted array as $-b_\alpha^*$ and $-a_\alpha^*$, respectively. Then a bootstrap estimate of a $(1 - \alpha)$ -percent confidence interval for \mathcal{S} is

$$\widehat{\mathcal{S}} + a_\alpha^* \leq \mathcal{S} \leq \widehat{\mathcal{S}} + b_\alpha^*. \quad (3.18)$$

¹⁹Choice of κ by cross validation has been proposed by Fan and Gijbels (1996) and has been used by Wheelock and Wilson (2001) and Wilson and Carey (2004) and others.

²⁰Ordinary bootstrap methods are inconsistent in our context due to the asymptotic bias of the estimator; see Mammen (1992) for additional discussion.

The idea underlying (3.18) is that the empirical distribution of the bootstrap values $(\widehat{\mathcal{S}}_b^* - \widehat{\mathcal{S}})$ mimics the unknown distribution of $(\widehat{\mathcal{S}} - \mathcal{S})$, with the approximation improving as $n \rightarrow \infty$. As $B \rightarrow \infty$, the choices of $-b_\alpha^*$ and $-a_\alpha^*$ become increasingly accurate estimates of the percentiles of the distribution of $(\widehat{\mathcal{S}}_b^* - \widehat{\mathcal{S}})$ (we set $B = 1000$). Any bias in $\widehat{\mathcal{S}}$ relative to \mathcal{S} is reflected in bias of $\widehat{\mathcal{S}}^*$ relative to $\widehat{\mathcal{S}}$; in the case of large bias, it is conceivable that the estimated confidence interval may not include the original estimate $\widehat{\mathcal{S}}$, since the estimated confidence interval corrects for the bias in $\widehat{\mathcal{S}}$. Confidence intervals for the EPSE measures are estimated similarly.

4 Empirical Results

As discussed above in Section 3, values for the three bandwidth parameters λ_0 , λ_1 , and λ_2 are needed for estimation. Using the credit-union data to optimize the least-squares cross-validation function in (3.14) yields $\widehat{\lambda}_0 = 0.00394$ (and hence $\widehat{\kappa} = 726$), $\widehat{\lambda}_1 = 0.962$, and $\widehat{\lambda}_2 = 0.99$. Recalling the discussion following (3.9) and (3.10), the data indicate that little smoothing should be used across the categories determined by the dummy variables D_1 and D_2 . However, the data also indicate that a good deal of smoothing should be done across time periods.

We used a grid-search over three dimensions to minimize the cross-validation function given in (3.14). Figure 2 shows the cross-validation function plotted against pairs of the three bandwidth parameters. The plots appear smooth and well-behaved, suggesting that the selected values of the bandwidth parameters yield a global (as opposed to local) minimum for the cross-validation function.

We used the selected bandwidth values to estimate the RSE measure defined in (2.6) for $\theta \in \{0.05, 0.10, 0.15, \dots, 0.95, 1.0, 2.0, \dots, 25.0\}$, with $(\mathbf{y}_0, \mathbf{w}_0)$ given by the medians of each variable, setting T equal to 1, 9, or 18 (corresponding to the first, middle, and last years of our observation period, i.e., 1989, 1997, and 2006). We chose the range of values for θ after noting that total assets in our sample range from about 0.05 times median assets to about 25 times median assets.

Results for estimation of RSE are illustrated in Figures 3–6; the four different figures reflect the four combinations of values for the dummy variables D_1 and D_2 . Each figure

contains three panels corresponding to 1989, 1997, and 2006. In each panel, we plot estimates of $\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ as a function of θ , using log-scales for both axes, and connecting the plotted points with solid lines. In addition, we used the bias-corrected bootstrap describe above in Section 3 to estimate 95-percent confidence intervals corresponding to each estimate of $\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$; upper and lower bounds are indicated by the dashed curves in Figures 3–6.

Recalling the discussion in Section 2, downward slopes for the RSE measure as a function of θ indicate increasing returns to scale along the ray from the origin through the median point $(\mathbf{y}_0, \mathbf{w}_0)$. Figure 3 displays results for $D_1 = D_2 = 0$, i.e., for credit unions with zero values for both real estate and commercial loans (this group accounts for 35.7 percent of the 184,279 observations in our sample). The estimated confidence intervals illustrated by Figure 3 indicates that the RSE measure, as a function of θ , is downward sloping over the entire range of values of θ in all three years, except for very small values of θ in 1989 and very large values in 1997 and 2006, where the results suggest locally constant returns to scale.

Note that in each panel of Figure 3, estimates of $\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ (indicated by the solid curve) for the smallest and the largest values of θ lie outside corresponding estimated 95-percent confidence intervals (indicated by the dashed curves). This reflects the fact that the local-polynomial estimator used to estimate the cost function in (3.4) is only weakly consistent, and asymptotically biased. In addition, estimates of cost $\widehat{C}(\mathbf{y}, \mathbf{w})$ obtained from (3.5) involve a non-linear transformation of fitted values from estimates \widehat{C} of the dependent variable in (3.4). Furthermore, estimation of the RSE measure in (2.6) involve further non-linear transformations of estimates \widehat{C} . Thus, even if the local-polynomial regression estimator yielded unbiased estimates, estimates of the RSE measure would be biased. As discussed above in Section 3, our bootstrap method involves a bias correction and hence estimates of $\mathcal{S}(\theta \mid \mathbf{y}_0, \mathbf{w}_0)$ sometimes lie outside the corresponding estimated confidence intervals.

Results for RSE among credit unions that make no real estate loans ($Y_1 = 0$) but do make commercial loans ($Y_2 > 0$) appear similar, as shown in Figure 4, with estimated confidence intervals suggesting that RSE are increasing except perhaps for very small credit unions in 1989. However, this group of credit unions is quite small, representing just 0.8 percent of the entire sample.

Roughly half of credit unions are observed to make real estate loans ($Y_1 > 0$) but no commercial loans ($Y_2 = 0$). Estimated RSE for these credit unions are illustrated in Figure

5. Here again, the results suggest increasing returns to scale, except for the smallest credit unions in 1989 and in 2006.

Figure 6 shows estimated RSE for credit unions that are observed to make both real estate and commercial loans. These credit unions account for about 13.9 percent of the entire sample and 22.6 percent of the observations in 2006. The results shown in Figure 6 suggest increasing returns to scale among credit unions that are beyond median output levels. For $\theta < 1$, the estimated confidence intervals in each of the three years shown are wider than in the preceding figures, and are suggestive of decreasing returns to scale due to the apparent upward slopes. However, this finding is likely spurious and irrelevant because the overwhelming majority of credit unions that make both real estate and commercial loans are large organizations with assets that exceed the industry median.

Our results for RSE are broadly suggestive of increasing returns to scale among credit unions. However, as noted above in Section 2, RSE may provide a misleading view of scale economies for credit unions because few, if any, credit unions are actually located on the ray \mathfrak{R}_0 . Table 4 gives results from our estimation of EPSE (with $\gamma = 0.05$ in (2.8)). As in the figures showing results for RSE, in Table 4 we focus on the first, middle, and last year of our observation period, i.e., 1989, 1997, and 2006. For each year, we estimate EPSE for each credit union represented in our sample for that year, as well as corresponding (95 percent) confidence intervals using the bias-corrected bootstrap described near the end of Section 3. For each year, we divide credit unions into quintiles of total assets; in Table 4, we report, for each quintile in each year, the median value of the EPSE estimates across credit unions, the numbers of estimates significantly less than one (and hence indicating increasing returns to scale) in the column labeled “IRS”, the numbers of estimates insignificantly different from one (and hence failing to reject constant returns to scale) in the column labeled “CRS”, and the numbers of estimates significantly greater than one (indicative of decreasing returns to scale) in the column labeled “DRS”. The last column of Table 4 gives the number of observations in each quintile-year.

The results presented in Table 4 indicate that in each year shown, a large majority of credit unions in each size quintile face increasing returns to scale. Hence, our results for EPSE are consistent with our results for RSE. Nevertheless, in each of the three years shown in the table, the first size quintile, comprising the smallest 20 percent of credit unions, has

the largest proportion of credit unions facing *decreasing* returns to scale. Recalling that the smallest credit unions are quite small—much smaller than the smallest commercial banks operating in the United State—these results may indicate a threshold for small credit unions. In order to increase their asset-size, very small credit unions may incur decreasing returns to scale before reaching a size large enough to justify the purchase of computers, software, IT services, etc. that are likely to lead to increasing returns to scale.

Regardless of why more credit unions in the smallest size quintile face decreasing returns to scale than credit unions in the larger quintiles, our results indicate that a majority of credit unions in every quintile operate under increasing returns to scale. This finding suggests that restrictions on credit union membership and activities may prevent many credit unions from attaining the size necessary to realize fully economies of scale.

5 Conclusions

Credit unions hold a small, but growing share of total U.S. depository institution assets. Moreover, like commercial banks, the average size of credit unions has increased sharply during the past two decades, suggesting that changes in regulation and technology have favored larger organizations over their smaller competitors. Researchers have found evidence of expanding returns to scale for commercial banks, and that large banks have experienced larger increases in productivity than small banks. However, we are unaware of studies investigating returns to scale rigorously for credit unions.

This paper uses a non-parametric local-linear estimator to estimate a model of credit union costs, from which we derive estimates of returns to scale. As other studies have found using data on commercial banks and other types of firms, we test and reject as a misspecification even a comparatively flexible translog cost function for credit unions. Our non-parametric estimator avoids the difficulty of specifying and estimating a parametric cost function such as a translog function. Further, we employ a dimension-reduction technique to reduce estimation error that can arise when non-parametric estimators are used to estimate high-dimension models.

We use annual data on all U.S. credit unions (except corporate credit unions and those with missing or implausible data) for 1989-2006 to estimate both ray-scale and expansion-path scale economies. Although most studies focus on ray-scale economies, we also examine

expansion-path scale economies to better estimate scale economies near the combinations of inputs and outputs that reflect actual credit union production. We find that throughout the sample period, a majority of credit unions operated under increasing returns to scale, as reflected in both ray-scale and expansion-path estimates. Thus, despite considerable industry consolidation and growth in average credit union size, it appears that as recently as 2006 most credit unions were too small to exploit fully possible scale economies. Regulations that limit the field of membership that credit unions may serve and restrictions on their activities may impede the growth of credit unions. Although the Credit Union Membership Access Act of 1998 eased constraints on credit union membership and helped promote industry consolidation, the principal finding of this paper is that a majority of credit unions continue to operate under increasing returns to scale. Hence, a further easing of membership restrictions or boundaries on credit union activities would likely spur further industry consolidation and increasing average size as many credit unions sought to achieve economies of scale.

A Appendix

In order to test a translog specification for credit unions' cost function, for each year 1989, ..., 2006 represented in our sample we computed median total assets and created two subsamples of observations. In subsample 1 we include all observations for a particular year where total assets are less than or equal to median assets for that year, while in subsample 2 we include all observations for the given year where total assets are greater than median assets for that year. Next, we use each subset to estimate the translog cost model

$$\begin{aligned}
\log(C/W_2) = & \beta_1 + \beta_2 \log(Y_1 + Y_2 + Y_3) + \beta_3 \log(1 + Y_4) + \beta_4 \log Y_5 + \beta_5 \log Y_6 \\
& + \beta_6 \log(W_1/W_2) + \beta_7 D_1 + \beta_8 D_2 + \beta_9 \log(Y_1 + Y_2 + Y_3) \log(Y_1 + Y_2 + Y_3) \\
& + \beta_{10} \log(Y_1 + Y_2 + Y_3) \log(1 + Y_4) + \beta_{11} \log(Y_1 + Y_2 + Y_3) \log Y_5 \\
& + \beta_{12} \log(Y_1 + Y_2 + Y_3) \log Y_6 + \beta_{13} \log(Y_1 + Y_2 + Y_3) \log(W_1/W_2) \\
& + \beta_{14} \log(1 + Y_4) \log(1 + Y_4) + \beta_{15} \log(1 + Y_4) \log Y_5 + \beta_{16} \log(1 + Y_4) \log Y_6 \\
& + \beta_{17} \log(1 + Y_4) \log(W_1/W_2) + \beta_{18} \log Y_5 \log Y_5 + \beta_{19} \log Y_5 \log Y_6 \\
& + \beta_{20} \log Y_5 \log(W_1/W_2) + \beta_{21} \log Y_6 \log Y_6 + \beta_{22} \log Y_6 \log(W_1/W_2) \\
& + \beta_{23} \log(W_1/W_2) \log(W_1/W_2) + \beta_{24} D_1 \log(Y_1 + Y_2 + Y_3) + \beta_{25} D_1 \log(1 + Y_4) \\
& + \beta_{26} D_1 \log Y_5 + \beta_{27} D_1 \log Y_6 + \beta_{28} D_1 \log(W_1/W_2) + \beta_{29} D_2 \log(Y_1 + Y_2 + Y_3) \\
& + \beta_{30} D_2 \log(1 + Y_4) + \beta_{31} D_2 \log Y_5 + \beta_{32} D_2 \log Y_6 + \beta_{33} D_2 \log(W_1/W_2) \\
& + \varepsilon,
\end{aligned} \tag{A.1}$$

$E(\varepsilon) = 0$. Note that dividing cost (C) and the price of capital (W_1) by the price of labor (W_2) ensures homogeneity with respect to input prices. In addition, it is necessary to add a constant to Y_4 due to a small number of observations with zero values for this variable. Our treatment of Y_1 , Y_2 , and Y_3 is similar to that in our non-parametric estimation, and avoids taking logs of zero, due to the large number of observed zero-values for Y_2 and Y_3 .

For subsample j containing n_j observations in a given year, $j \in \{1, 2\}$, let $\beta_j = [\beta_1 \dots \beta_{33}]'$, and let \mathbf{X}_j be the $(n_j \times 33)$ matrix containing the right-hand side variables in (A.1); the first column of \mathbf{X}_j consists of a vector of 1s. In addition, let \mathbf{Y}_j represent the $(n_j \times 1)$ matrix containing the n_j observations on the left-hand side variable in (A.1), so

that the model can be written (for sub-sample j in a given year) as

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \quad (\text{A.2})$$

where $\boldsymbol{\varepsilon}_j$ is an $(n_j \times 1)$ matrix of disturbances with zero mean.

Using data for each subsample $j = 1, 2$ in a given year, we estimate (A.1) using ordinary least squares (OLS), yielding $\widehat{\boldsymbol{\beta}}_j$ and $\widehat{\boldsymbol{\varepsilon}}_j = \mathbf{Y}_j - \mathbf{X}_j\widehat{\boldsymbol{\beta}}_j$. Next, we compute White's (1980) heteroskedasticity-consistent covariance matrix estimator

$$\widehat{\boldsymbol{\Sigma}} = (\mathbf{X}'_j\mathbf{X}_j)^{-1}(\mathbf{X}'_j\mathbf{E}_j\mathbf{X}_j)(\mathbf{X}'_j\mathbf{X}_j)^{-1} \quad (\text{A.3})$$

for each subsample, where \mathbf{E}_j is the $(n_j \times n_j)$ diagonal matrix with elements of $\widehat{\boldsymbol{\varepsilon}}_j$ along the principal diagonal. Under the null hypothesis $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, asymptotic normality of OLS estimators ensures that the Wald statistic

$$\widehat{W} = \left(\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2\right)' \left(\widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_2\right)^{-1} \left(\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2\right) \xrightarrow{d} \chi^2(33). \quad (\text{A.4})$$

We computed the Wald statistic in (A.4) for each of the 18 years represented in our sample, obtaining values ranging from 536.39 to 1083.08, and corresponding p -values less than 10^{-6} in every case. Hence, the translog specification in (A.1) is rejected at any reasonable level of significance, for each year represented in our sample.

References

- Aitchison, J., and Aitken, C. G. G. (1976), “Multivariate binary discrimination by the kernel method,” *Biometrika*, 63, 413–420.
- Banks, J., Blundell, R., and Lewbel, A. (1997), “Quadratic engel curves and consumer demand,” *Review of Economics and Statistics*, 79, 527–539.
- Barnett, W. A., Geweke, J., and Wolfe, M. (1991), “Seminonparametric Bayesian estimation of the asymptotically ideal production model,” *Journal of Econometrics*, 49, 5–50.
- Bauer, K. (2008), “Detecting abnormal credit union performance,” *Journal of Banking and Finance*. In press.
- Berger, A. N. (2003), “The economic effects of technological progress: Evidence from the banking industry,” *Journal of Money, Credit, and Banking*, 35, 141–76.
- Bierens, H. J. (1987), “Kernel estimators of regression functions,” In *Advances in Econometrics, Fifth World Congress*, ed. T. F. Bewley, vol. 1, Cambridge: Cambridge University Press.
- Chalfant, J. A., and Gallant, A. R. (1985), “Estimating substitution elasticities with the Fourier cost function,” *Journal of Econometrics*, 28, 205–222.
- Cooper, R. J., and McLaren, K. R. (1996), “A system of demand equations satisfying effectively global regularity conditions,” *Review of Economics and Statistics*, 78, 359–364.
- Delgado, M. A., and Mora, J. (1995), “On asymptotic inferences in nonparametric and semi-parametric models with discrete and mixed regressors,” *Investigaciones Economicas*, 19, 435–467.
- Eastwood, B. J. (1991), “Asymptotic normality and consistency of semi-parametric regression estimators using an upwards F test truncation rule,” *Journal of Econometrics*, 48, 151–191.
- Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, London: Chapman and Hall.
- Emmons, W. R., and Schmid, F. A. (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.
- Fan, J., and Gijbels, I. (1994), “Censored regression: Local linear regression smoothers,” *Journal of the American Statistical Association*, 89, 560–570.
- (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.
- Fan, J., T. Gasser, I. Gijbels, M. Brockmann, and Engel, J. (1997), “Local polynomial regression: Optimal kernels and asymptotic minimax efficiency,” *Annals of the Institute for Statistical Mathematics*, 49, 79–99.
- Frame, W. S., and Coelli, T. (2001), “U.S. financial services consolidation: The case of corporate credit unions,” *Review of Industrial Organization*, 18, 229–242.

- Frame, W. S., Karels, Gordon V., and McClatchey, C. A. (2003), "Do credit unions use their tax advantage to benefit members? evidence from a cost function," *Review of Financial Economics*, 12, 35–47.
- Fried, Harold O., Lovell, C. A. Knox, and Eeckaut, Philippe Vanden (1993), "Evaluating the performance of u. s. credit unions," *Journal of Banking and Finance*, 17, 251–265.
- Fried, Harold O., Lovell, C. A. Knox, and Yaisawarng, Suthathip (1999), "The impact of mergers on credit union service provision," *Journal of Banking and Finance*, 23, 367–386.
- Gallant, A. R. (1981), "On the bias in flexible functional forms and an essentially unbiased form," *Journal of Econometrics*, 15, 211–245.
- (1982), "Unbiased determination of production technologies," *Journal of Econometrics*, 20, 285–323.
- Gilbert, R. A., and Zaretsky, A. M. (2003), "Banking antitrust: Are the assumptions still valid?," *Federal Reserve Bank of St. Louis Review*, 85, 29–52.
- Goddard, John A., McKillop, Donal G., and Wilson, John O. S. (2002), "The growth of US credit unions," *Journal of Banking and Finance*, 26, 2327–2356.
- (2007), "Consolidation in the US credit union sector: Determinants of failure and acquisition," Unpublished working paper, Bangor Business School, University of Wales, Bangor, United Kingdom.
- (2008), "The diversification and financial performance of US credit unions," *Journal of Banking and Finance*. Forthcoming.
- Guilkey, D. K., Lovell, C. A. K., and Sickles, R. C. (1983), "A comparison of the performance of three flexible functional forms," *International Economic Review*, 24, 591–616.
- Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.
- Härdle, W., and Linton, O. (1999), "Applied nonparametric methods," In *Handbook of Econometrics, Volume 4*, ed. D. L. McFadden and R. F. Engle, New York: Elsevier North-Holland, Inc.
- Härdle, W., and Mammen, E. (1993), "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 21, 1926–1947.
- Leggett, Keith J., and Strand, Robert W. (2002), "Membership growth, multiple membership groups and agency control at credit unions," *Review of Financial Economics*, 11, 37–46.
- Loftsgaarden, D. O., and Quesenberry, C. P. (1965), "A nonparametric estimate of a multivariate density function," *Annals of Mathematical Statistics*, 36, 1049–1051.
- Mammen, Enno (1992), *When Does Bootstrap Work? Asymptotic Results and Simulations*, Berlin: Springer-Verlag.
- McAllister, P. H., and McManus, D. (1993), "Resolving the scale efficiency puzzle in banking," *Journal of Banking and Finance*, 17, 389–405.

- Mitchell, K., and Onvural, N. M. (1996), “Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form,” *Journal of Money, Credit, and Banking*, 28, 178–199.
- Nadarya, E. A. (1964), “On estimating regression,” *Theory of Probability and its Applications*, 10, 186–190.
- Pagan, A., and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge: Cambridge University Press.
- Petersen, M. A., and Rajan, R. G. (2002), “Does distance still matter? the information revolution in small business lending,” *The Journal of Finance*, 57, 2533–2570.
- Racine, J., and Li, Q. (2000), “Nonparametric estimation of regression functions with both categorical and continuous data,” Unpublished working paper, Department of Economics, University of South Florida, Tampa, FL.
- (2004), “Nonparametric estimation of regression functions with both categorical and continuous data,” *Journal of Econometrics*, 119, 99–130.
- Smith, D. (1984), “A theoretic framework for the analysis of credit union decision making,” *Journal of Finance*, 69, 1155–1168.
- Spong, K. (2000), *Banking Regulation: Its Purposes, Implementation, and Effects*, fifth ed., Kansas City, Missouri: Federal Reserve Bank of Kansas City.
- Szegő, G. (1959), *Orthogonal Polynomials*, New York: American Mathematical Society.
- Walter, J. R. (2006), “Not your father’s credit union,” *Federal Reserve Bank of Richmond Economic Quarterly*, 92, 353–377.
- Watson, G. (1964), “Smooth regression analysis,” *Sankhya Series A*, 26, 359–372.
- Wheelock, D. C., and Wilson, P. W. (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 US commercial banks,” *Journal of Monetary Economics*, 47, 115–132.
- (2008), “Robust nonparametric quantile estimation of efficiency and productivity change in U. S. commercial banking, 1985–2004,” *Journal of Business and Economic Statistics*. Forthcoming.
- White, H. (1980), “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica*, 48, 817–838.
- Wilcox, James A. (2005), “Economies of scale and continuing consolidation of credit unions,” *Federal Reserve Bank of San Francisco Economic Letter*.
- (2006), “Performance divergence of large and small credit unions,” *Federal Reserve Bank of San Francisco Economic Letter*.
- Wilson, P. W., and Carey, K. (2004), “Nonparametric analysis of returns to scale and product mix among US hospitals,” *Journal of Applied Econometrics*, 19, 505–524.

Table 1: Variable Definitions

Y_1 — Real estate loans: amount of first mortgage real estate loans (CUSA0243) + amount of other real estate loans (CUSA0244).

Y_2 — Commercial loans: for years 1989–2003, amount of commercial loans (CUSA0257) + amount of agricultural loans to members (CUSA1235); for years 2004–2006, member business loans, total amount outstanding (CUSA4899).

Y_3 — Consumer loans: total loans and leases, amount (CUSA1263) $-(Y_1 + Y_2)$.

Y_4 — Investments: for years 1989–2005, total investments (less derivatives contracts) (CUSA4577); for year 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).

Y_5 — Savings pricing: [dividends on shares (CUSA4278) + interest on deposits (CUSA4279)] / total shares and deposits (CUSA2197).

Y_6 — Loan pricing: interest and fee income on loans, total (CUSA4010 / amount of total loans and leases (CUSA1263)).

W_1 — Price of 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 by total shares and deposits (CUSA2197).

W_2 — Price of labor: labor expenses, i.e. officers and employee compensation (CUSA4137), divided by number of full-time credit union employees (CUSA6047) + (1/2 times) number of part-time credit union employees (CUSA6048).

T — Time: equals 1 for 1989, 2 for 1990, ..., 18 for 2006.

D_1 — Dummy variable: equals 1 if $Y_1 > 0$; 0 otherwise.

D_2 — Dummy variable: equals 1 if $Y_2 > 0$; 0 otherwise.

C — Variable cost: capital expenses + labor expenses.

Table 2: Summary Statistics, 1989–2006

	Minimum	Q1	Median	Mean	Q3	Maximum	#0s
<i>COST</i>	1.68	66.34	223.20	1219.00	796.70	598100.00	0
Y_1	0.00	0.00	354.50	10090.00	3756.00	8731000.00	67172
Y_2	0.00	0.00	0.00	417.10	0.00	511900.00	157261
Y_3	0.00	1198.00	3335.00	14790.00	9657.00	8581000.00	12
Y_4	0.00	564.40	2117.00	12230.00	7267.00	6260000.00	94
Y_5	0.0000	0.0254	0.0351	0.0360	0.0449	0.3333	0
Y_6	0.0047	0.0805	0.0924	0.0958	0.1106	0.7778	0
W_1	0.0001	0.0124	0.0178	0.0189	0.0237	0.4583	0
W_2	0.0026	22.3100	29.8500	29.7700	36.9800	4768.0000	0
T	1.00	4.00	9.00	8.81	13.00	18.00	0
D_1	0.00	0.00	1.00	0.63	1.00	1.00	67172
D_2	0.00	0.00	0.00	0.14	0.00	1.00	157261
<i>ASSETS</i>	11.95	2201.00	6955.00	39550.00	23540.00	22840000.00	0

NOTE: *COST*, Y_1 , Y_2 , Y_3 , Y_4 , and *ASSETS* are measured in thousands of (year 2000) dollars; Y_5 , Y_6 , and W_1 are dimensionless ratios, and W_2 is measured in thousands of (year 2000) dollars per full-time equivalent employees. Columns labelled “Q1” and “Q3” give first and third quartiles for each variable, while the last column (labelled “#0s”) gives the number of observations where each variable equals zero.

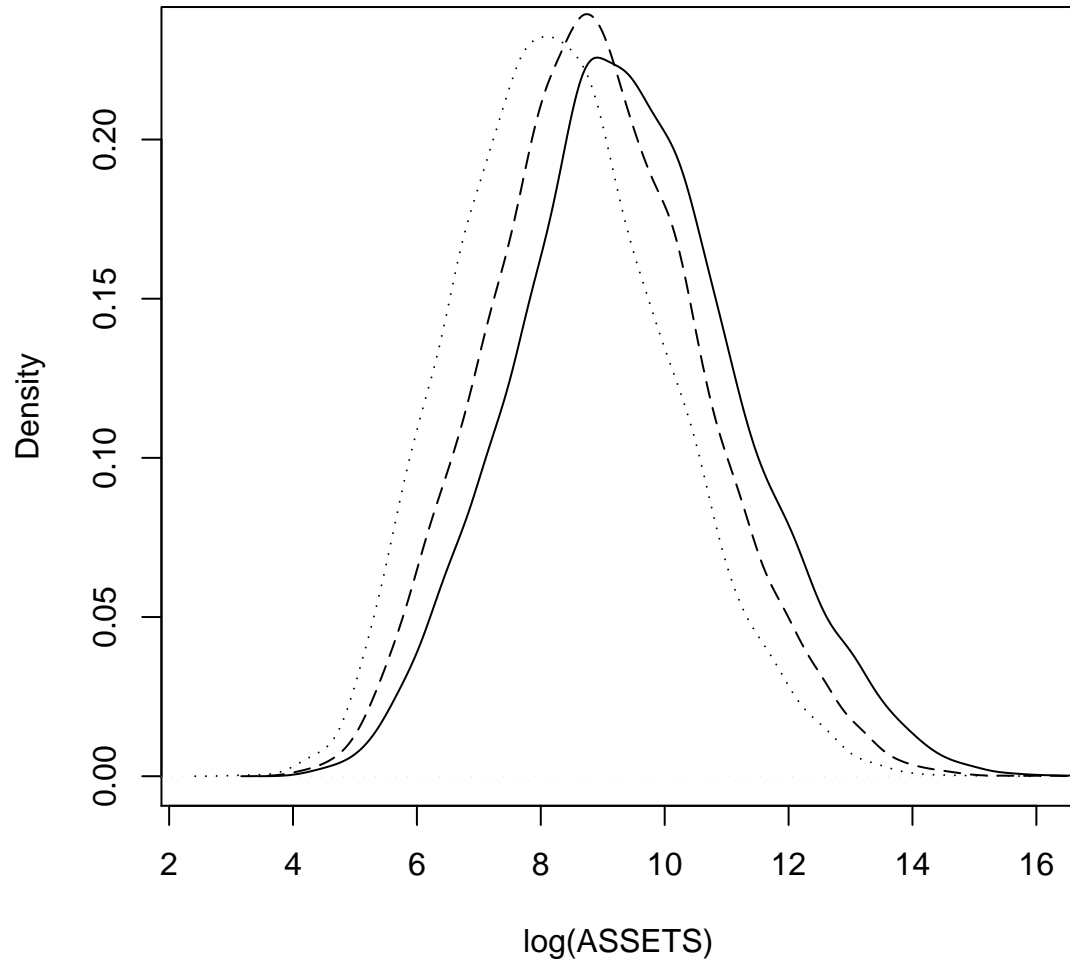
Table 3: Number of Observations per Year

Year	#obs	Year	#obs	Year	#obs
1989	12438	1995	10766	2001	9322
1990	12213	1996	10539	2002	8970
1991	12022	1997	10389	2003	8895
1992	11727	1998	10169	2004	8574
1993	11340	1999	9935	2005	8278
1994	11065	2000	9598	2006	8039

Table 4: Expansion-Path Scale Economies

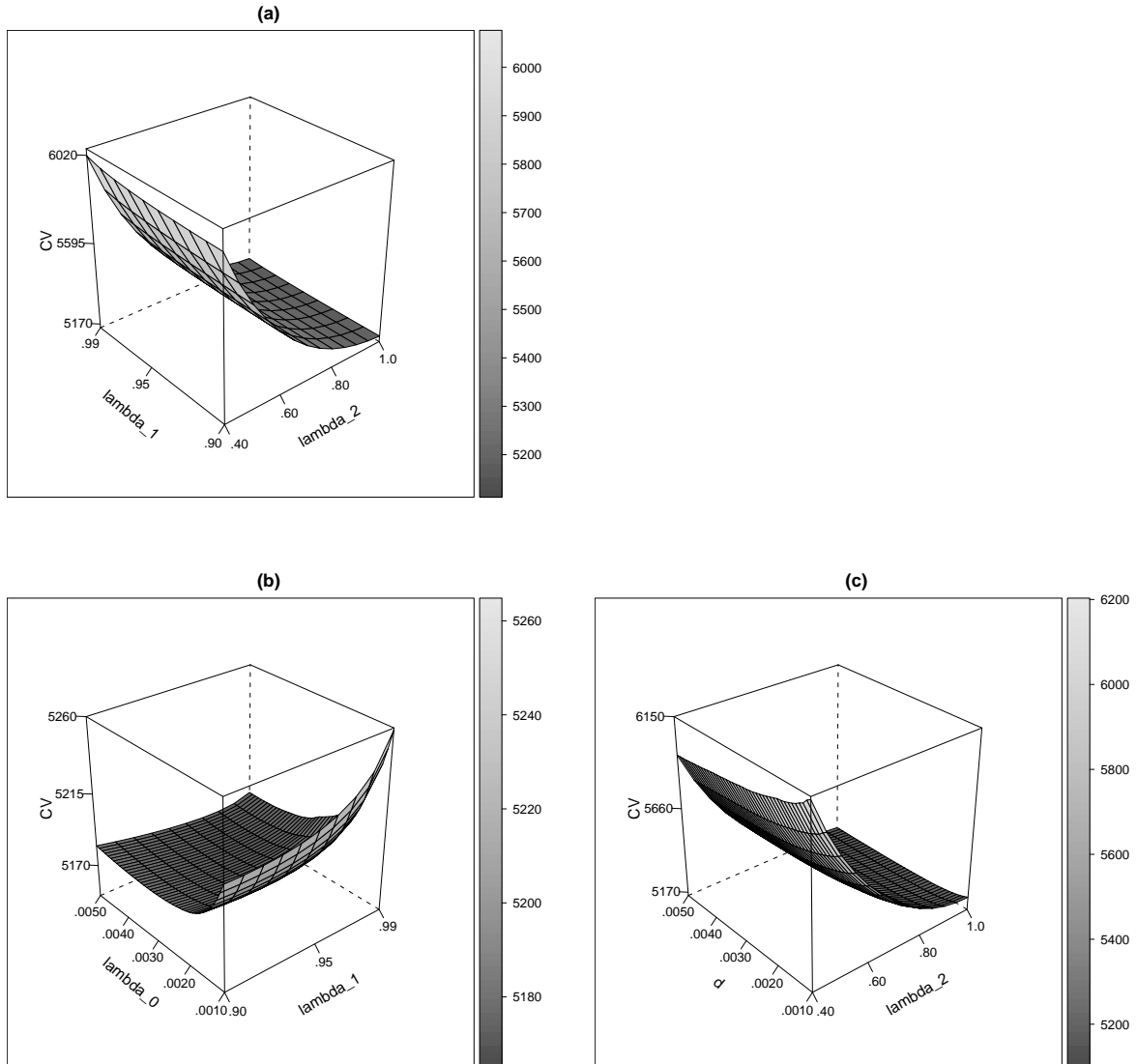
Quintile	Median	IRS	CRS	DRS	# Obs.
—1989—					
1	1.0107	1578	162	748	2488
2	0.8450	2170	96	220	2486
3	0.8208	2311	53	123	2487
4	0.8180	2323	66	100	2489
5	0.8202	1964	284	240	2488
—1997—					
1	0.8644	1791	79	207	2077
2	0.8136	2014	22	43	2079
3	0.8013	2051	9	17	2077
4	0.8017	2039	17	22	2078
5	0.8014	1805	159	114	2078
—2006—					
1	0.8792	1209	165	233	1607
2	0.8137	1454	109	46	1609
3	0.8019	1491	81	35	1607
4	0.7974	1514	70	24	1608
5	0.7743	1379	119	110	1608

Figure 1: Density of Total (Log) Assets



Note: Kernel estimates of the density of (log) total assets for 1989, 1997, and 2006 are shown by the dotted, dashed, and solid curves, respectively. Total assets are measured in thousands of constant year 2000 dollars.

Figure 2: Least Squares Cross-Validation



Note: Panel (a) shows the cross-validation function as a function of λ_1 and λ_2 , with λ_0 fixed at its optimal value. Similarly, panel (b) shows the cross-validation function as a function of λ_0 and λ_1 , with λ_2 fixed at its optimal value; panel (c) shows the cross-validation function as a function of λ_0 and λ_2 , with λ_1 fixed at its optimal value.

Figure 3: Ray Scale Economies ($Y_1 = Y_2 = 0$)

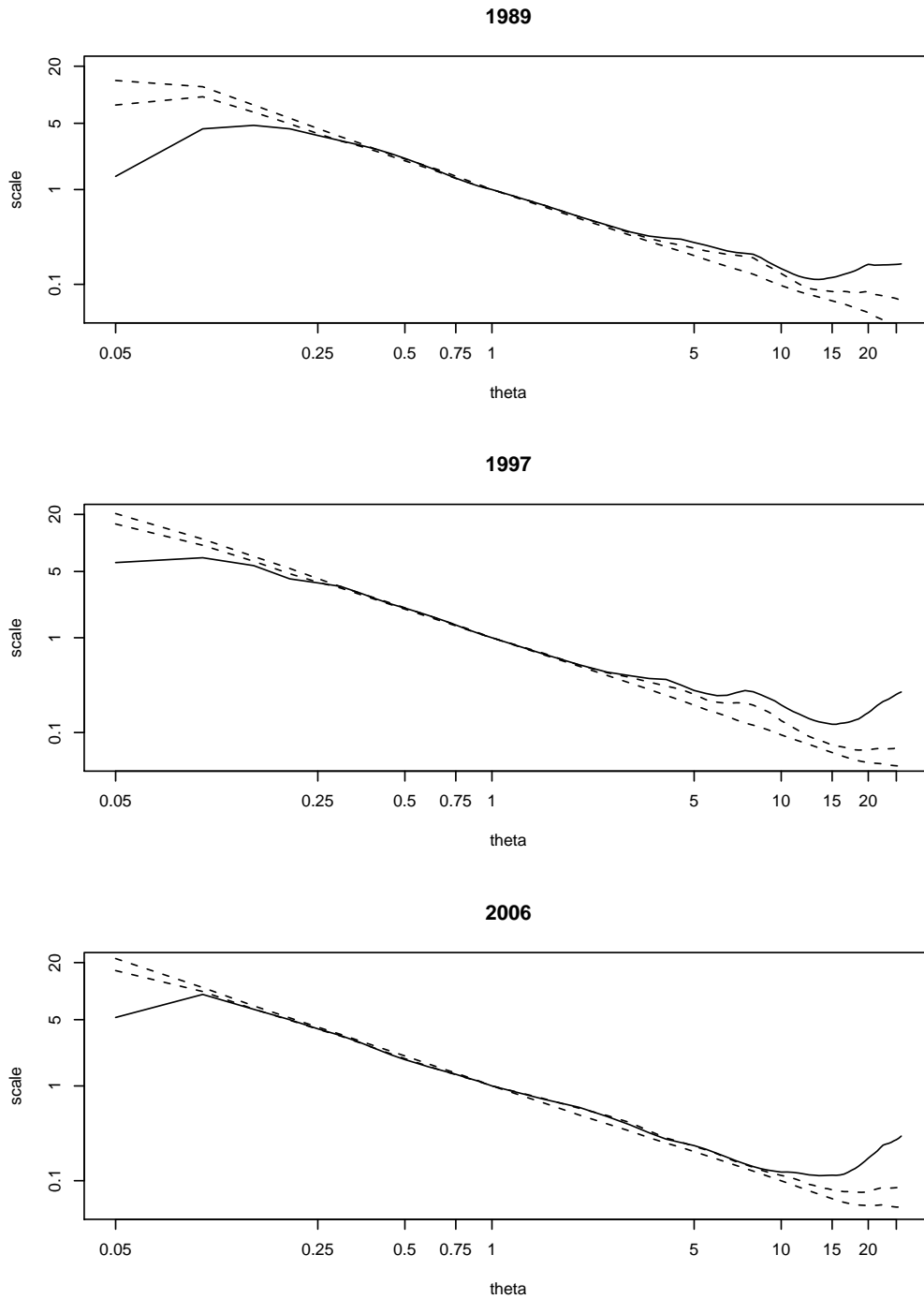


Figure 4: Ray Scale Economies ($Y_1 = 0, Y_2 > 0$)

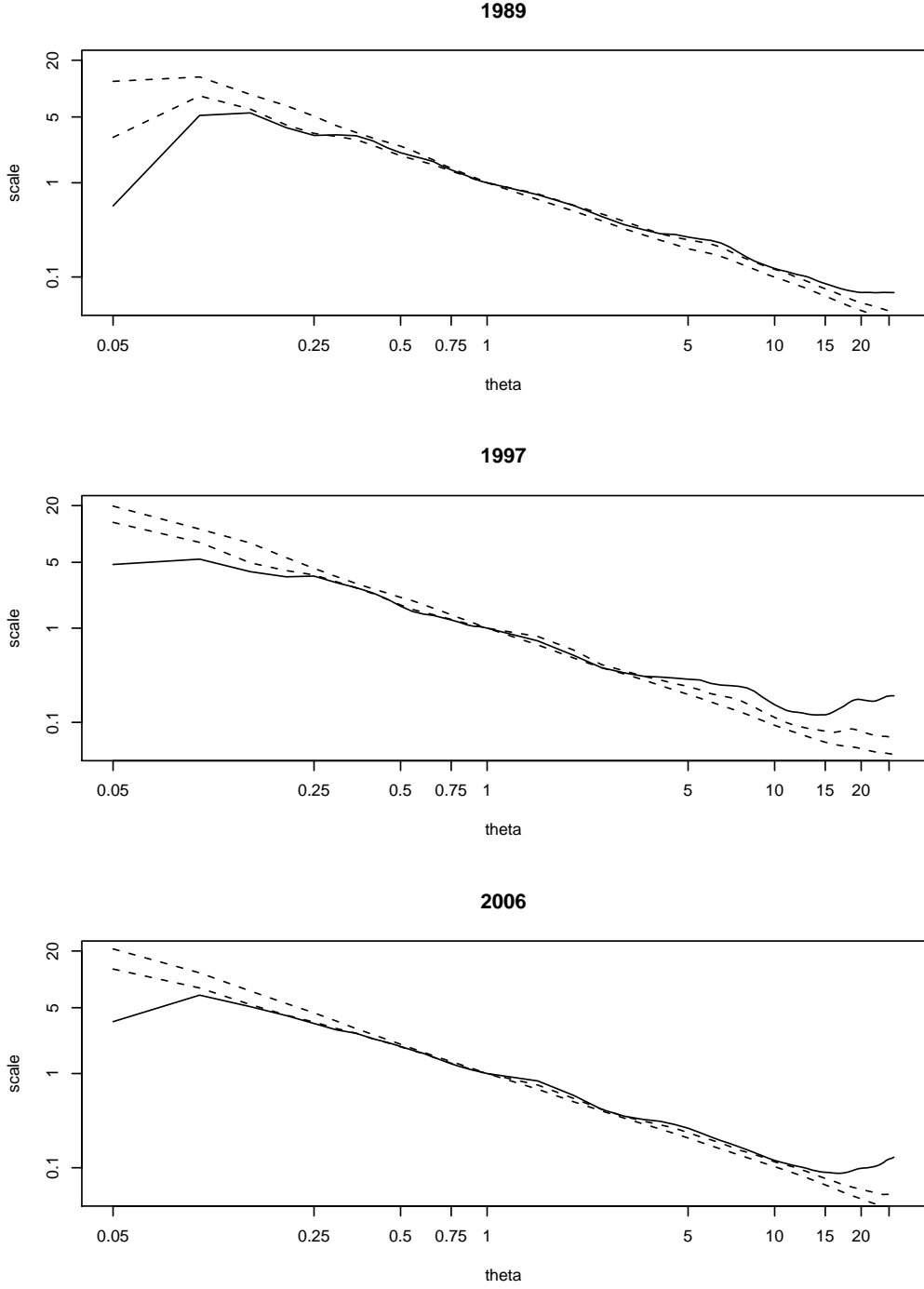


Figure 5: Ray Scale Economies ($Y_1 > 0, Y_2 = 0$)

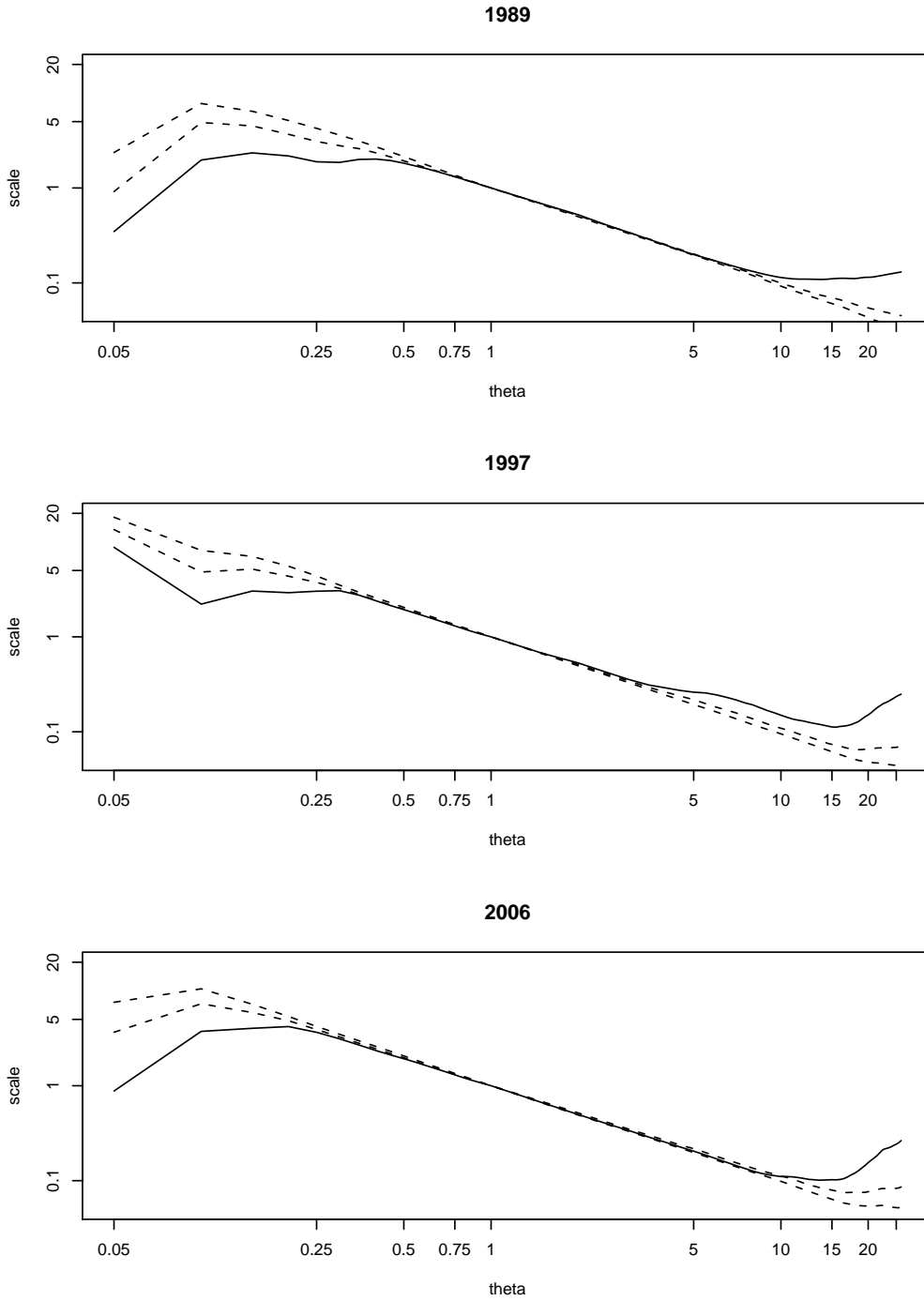


Figure 6: Ray Scale Economies ($Y_1 > 0, Y_2 > 0$)

