Economies of Diversification in the US Credit Union Sector*

Emir Malikov[1] Shunan Zhao[2] Subal C. Kumbhakar[2]

[1]Department of Agricultural Economics & Rur. Soc., Auburn University, Auburn, AL
[2]Department of Economics, State University of New York at Binghamton, Binghamton, NY

First Draft: May 25, 2014
This Draft: November 13, 2016

Abstract

Significant scale economies have been recently cited to rationalize a dramatic growth in the US retail credit union sector over the past few decades. In this paper, we explore another plausible supply-side explanation for the growth of the industry, namely, economies of diversification. We focus on the fact that credit unions differ among themselves in the range of financial services they offer to their members. Since larger credit unions tend to offer a more diversified financial service menu than credit unions of a smaller size, the incentive to grow in size may be fueled not only by present scale economies but also by economies of diversification. This paper provides the first robust estimates of such economies of diversification for the credit union sector. We estimate a flexible semiparametric smooth coefficient quantile panel data model with correlated effects that is capable of accommodating a four-way heterogeneity among credit unions. Our results indicate the presence of non-negligible economies of diversification in the industry. We find that as many as 27–91% (depending on the type and the cost quantile) of diversified credit unions enjoy substantial economies of diversification; the cost of most remaining credit unions is invariant to the scope of services. We also find the overwhelming evidence of increasing returns to scale in the industry.

Keywords: correlated effects, credit union, diversification, economies of scope, returns to scale, semiparametric, varying coefficient, quantile regression

JEL Classification: C13, C14, D24, G21

* *Email*: emalikov@auburn.edu (Malikov), szhao15@binghamton.edu (Zhao), kkar@binghamton.edu (Kumbhakar).

We would like to thank the editor and three anonymous referees for many insightful comments and suggestions that helped improve this article. We are also thankful to Robin Sickles and other conference participants at the 2015 NY Camp Econometrics X in Bolton Landing for helpful feedback.

1 Introduction

In 2014, the Credit Union National Association[1] (CUNA) launched an online campaign of soliciting "selfies" from members of retail credit unions from all over the US in the celebration of the industry's closing in on the milestone of 100 million memberships. This movement was an exemplary testimony to a steady growth that the retail credit union sector had been enjoying over the past few decades (e.g., see Goddard et al., 2002; Goddard and Wilson, 2005). Credit unions currently cater to over 40% of the adult population in the US[2] and account for 9.3% of total national consumer savings totaling 1.1 trillion dollars in assets (CUNA, 2013). In the past ten years alone, the average size of federally-insured retail credit unions has more than doubled from 71.8 million to 162.0 million in total assets (NCUA, 2013). The industry has also been undergoing a wave of consolidation (Fried et al., 1999; Bauer et al., 2009).

This transformation of the industry has taken place despite the decline in the relative advantages that credit unions — locally-oriented small-scale mutual consumer cooperatives serving primarily retail consumers — have conventionally enjoyed over commercial banks. The increasing availability of credit information via nation-wide credit-reporting bureaus, establishment of the federal deposit insurance fund for credit unions and the growth in credit card lending by larger commercial banks have dramatically eroded benefits of the so-called "soft" information that lies at the core of credit unions' business formula (Walter, 2006).

Scale economies have recently been invoked to rationalize the observed industry growth (Emmons and Schmid, 1999; Wilcox, 2005, 2006). Indeed, Wheelock and Wilson (2011) and Malikov et al. (2015b) document pervasive increasing returns to scale among retail credit unions in the US in the 1989–2006 and 1994–2011 periods, respectively. These findings provide a plausible explanation to the growth of the industry, suggesting that credit unions grow in size to capitalize on scale economies in order to increase their competitive margins. They also suggest that scale economies are yet to be fully exploited by credit unions.[3]

In this paper, we seek to complement the above studies of scale economies among credit unions by exploring another plausible cost-side trigger[4] to the growth of the industry, namely, economies of diversification (scope). More specifically, we focus our analysis around the fact that credit unions differ among themselves in the scope/range of financial services they offer to their members. For instance, only 16% of retail credit unions in our sample offer business loans. Further, larger credit unions tend to provide a more diversified financial service menu to their membership pool than credit unions of a smaller size. Thus, the incentive to grow in size may be fueled not only by present scale economies but also by economies of diversification, which are said to exist if the average cost that credit unions incur declines with the increase in the scope of services provided. If, by growing their membership, credit unions are able to secure the demand for a more diversified scope of services, they would naturally have the incentive to grow in order to also capitalize on economies of diversification. Hence, the explanation of the recent growth in the industry by the presence of scale economies alone is likely to be missing another important factor such as economies of diversification. To our knowledge, this paper is the first attempt to measure economies of diversification (if any) for the US credit union sector.

[1] A national not-for-profit trade group serving credit unions in the US.

[2] Here, an "adult" is defined as a person of 18 years of age and older. The calculations are based on 2013 population estimates from the US Census Bureau.

[3] Some recent studies of scale economies in the *banking* sector include, among many others, Feng and Serletis (2010), Restrepo-Tobón and Kumbhakar (2015) and Malikov et al. (2015a).

[4] We acknowledge that the industry growth can also, at least partly, be attributed to demand-side factors. In this paper, we however exclusively focus on the supply/cost side of the phenomenon.

The "diversification" as a business strategy of financial intermediaries has been thoroughly studied, although from a quite different perspective, in the commercial banking literature [see Goddard et al. (2008) for an excellent review]. The above studies primarily focus on the reduced-form relationships between (revenue) diversification and the shareholder value, profitability, volatility of revenue and other financial metrics. These studies traditionally conceptualize the diversification as the split of the revenue sources either *between* interest and non-interest activities or *within* interest activities of the bank. The diversification itself is often measured as the revenue share of the either activity or as the (revenue) Herfindahl index. In a similar fashion, Goddard et al. (2008) examine how the diversification into non-interest income-generating activities affects the financial performance of US credit unions.

Our approach to studying benefits/costs of diversification differs from that described above in several aspects. First, the economies of diversification that we investigate has a sound foundation in the microeconomic theory and is measured in terms of cost savings due an increase in the scope of products via the integration of the multiple production processes into one (e.g., Berger et al., 1987; Rime and Stiroh, 2003). Second, we focus on the product diversification within a group of interest-earning activities only (namely, different kinds of loans). Lastly, we examine both the diversification and scale economies in the context of a (structural) model of credit union costs (Smith, 1984; Fried et al., 1999; Frame and Coelli, 2001; Frame et al., 2003; Malikov et al., 2015b).

In a pursuit of robust estimates of economies of diversification and scale for credit unions, we estimate a flexible semiparametric smooth coefficient quantile panel data model with correlated effects (Cai et al., 2013) that is capable of accommodating four kinds of heterogeneity among credit unions: *(i)* heterogeneity across credit unions offering different financial service menus, *(ii)* heterogeneity across entities of different size, *(iii)* heterogeneity across the (conditional) quantiles of the credit union cost distribution, and *(iv)* unobserved heterogeneity correlated with covariates in the estimated equation (e.g., management quality). The model we employ is essentially a synthesis of a traditional parametric quantile regression model (Koenker and Bassett, 1978; Koenker, 2005; Abrevaya and Dahl, 2008) with a semiparametric smooth coefficient mean regression model (Cai et al., 2000; Li et al., 2002). We estimate our model via local-linear kernel fitting.

Our results provide evidence in favor of economically and statistically significant economies of diversification present among retail credit unions continuously operating in the US during the 2001–2006 period. We find that, for a credit union in the middle of the cost distribution, the consolidation of the issuance of real estate loans with consumer loans and investments reduces credit union costs, on average, by astounding 45%. A similar integration of business loans and the mix of real estate and consumer loans and investments is estimated to reduce credit union costs by the median of 37–47% depending on the quantile of the cost. Overall, we find no notable evidence in support of diseconomies of diversification in the industry. On the contrary, as many as 27–91% (depending on the type and the cost quantile) of diversified credit unions enjoy statistically significant economies of diversification. The cost of most remaining credit unions is invariant to the scope of financial services. Consistent with findings of Wheelock and Wilson (2011) and Malikov et al. (2015b) [who however investigate credit union costs at the conditional mean only], we also find the overwhelming evidence of statistically significant increasing returns to scale across credit unions of all types and quantiles of the cost distribution.

Our findings of statistically significant, non-negligible economies of diversification and scale suggest that the growth of the credit union sector is far from reaching its peak. The industry-wide trends, such as the diversification of financial services offered to members as well as the consolidation of the sector, are therefore likely to persist over coming years.

The rest of the paper proceeds as follows. Section 2 describes the conceptual framework to modeling credit union costs as well as explains our methodology for computing economies of diversification. We present our econometric model of four-way heterogeneity and describe the associated nonparametric estimation methodology in Section 3. Section 4 provides the description of the data. Section 5 presents the empirical results. We conclude in Section 6.

2 Modeling Credit Union Costs

In order to empirically assess whether credit unions enjoy economies of diversification and/or of scale, we first need to formalize the model of their costs. Unlike commercial banks that seek to maximize profits via building up the wedge between the interest rates paid on deposits and those charged on loans, credit unions are not profit-maximizers. Rather, credit unions are (consumer) cooperatives that seek to maximize service provision to their members in terms of quantity, price and variety of services (Smith, 1984; Fried et al., 1999). This includes not only the provision of favorable[5] rates on both the deposits and loans to members but also offering the types of financial services most desired by members. Consistent with the above, we follow a wide practice in the credit union literature (e.g., Frame and Coelli, 2001; Frame et al., 2003) and model credit union costs via the "service provision approach". Specifically, we model credit unions as service providers that, given their prevailing production technologies, minimize variable non-interest cost subject to the levels and types of outputs (i.e., financial services) and the variable inputs prices.

We consider the following four outputs (financial services) that credit unions produce: real estate loans (y_1), business and agricultural loans (y_2), consumer loans (y_3) and other investments (y_4). Following Frame et al. (2003) and Wheelock and Wilson (2011, 2013), we further specify two quasi-fixed netputs (services) to capture the price dimension of the service provision by credit unions: the average interest rate on saving deposits ($\widetilde{y}_5$) and the (inverse of) average interest rate on loans ($\widetilde{y}_6$). Conditioning credit union costs on $\widetilde{y}_5$ and $\widetilde{y}_6$ is meant to account for the provision of benefits to members in the form of favorable rates. Like Malikov et al. (2016), one thus may prefer thinking of these price variables as quasi-fixed outputs.[6] Further, we define two variable inputs: financial capital (x_1) and labor (x_2) with the vector of corresponding input prices $\boldsymbol{w} = (w_1, w_2)$. All these variables are taken as arguments of the dual (short-run) variable non-interest cost function of credit unions.

The dual cost framework that we employ in this paper is more advantageous over an alternative primal specification, which can also be used to model credit union operations, because it avoids using input quantities during the estimation which can lead to simultaneity problems since the variable input allocation is endogenous to credit unions. Treating output quantities and input prices as exogenous is however justified theoretically by the cost minimization premise and widely accepted in the financial services literature [e.g., see Hughes and Mester (2015) for an excellent review].

As briefly discussed in the Introduction, credit unions widely differ from one another in the range of financial services they offer to their members. Malikov et al. (2015b) document that the overwhelming majority of retail, or so-called natural-person, credit unions (more than 99%) offer one of the following three financial service menus to their members: *(i)* consumer loans and investments $\boldsymbol{y}_1 = (y_3, y_4)$, *(ii)* real estate and consumer loans as well as investments $\boldsymbol{y}_2 = (y_1, y_3, y_4)$, and *(iii)* all types of financial services: real estate, business and consumer loans, and investments $\boldsymbol{y}_3 = (y_1, y_2, y_3, y_4)$. We label these financial service menus (i.e., output mixes) as "1", "2" and "3", respectively, and refer to corresponding retail credit unions as "Type 1", "Type 2" and "Type 3".

[5]Favorable over those offered by commercial banks.

[6]We therefore consider the *inverse* of the average interest rate on loans to enforce positive monotonicity (in outputs) of the cost function.

Table 1 in Section 4 reports, among other things, the distribution of credit unions in our sample by their types. Ignoring this *observed* heterogeneity in the range of financial services among credit unions would amount to making a strong and rather unrealistic assumption that all credit unions share the same cost structure that is invariant to the menu of services they provide. Given the marked heterogeneity between credit union types along all dimensions, as can be easily seen from Table 1, the assumption of invariance of credit union costs to the service menu is likely to result in the loss of information and the misspecification of the econometric model, the risk of which we do not wish to take on. We therefore model the cost structure of each of the three identified types of credit unions separately.

We explicitly recognize that, under the "service provision approach", credit unions minimize non-interest variable cost subject to *different* types of outputs among other relevant constraints. Consequently, the associated production technologies (and the cost relationships) are allowed to be heterogeneous over credit union types. We consider the following credit-union-type-specific dual (short-run) variable cost function[7]

$$C_r\left(\boldsymbol{y}_r, \widetilde{\boldsymbol{y}}, \boldsymbol{w}\right) = \inf_{\boldsymbol{x}} \left\{\boldsymbol{x}'\boldsymbol{w} \mid \mathbb{T}_r\left(\boldsymbol{y}_r, \widetilde{\boldsymbol{y}}, \boldsymbol{x}\right) \leq 1\right\}\ , \qquad r \in \{1, 2, 3\} \tag{2.1}$$

where $\widetilde{\boldsymbol{y}} = (\widetilde{y}_5, \widetilde{y}_6)$ is a vector of quasi-fixed netputs, $\boldsymbol{x} = (x_1, x_2)$ is a vector of variable inputs, and $\boldsymbol{w} = (w_1, w_2)$ is a vector of the variable input prices. The output vector $\boldsymbol{y}_r$ and the associated transformation $\mathbb{T}_r(\cdot)$ and cost $C_r(\cdot)$ functions are allowed to vary across credit union types and are thus indexed by the type $r \in \{1, 2, 3\}$. By modeling the cost structure for each of the credit union types separately, we are able to explicitly accommodate potential discontinuities in the credit union's cost function across different financial service menus, which are likely to arise due to the change in dimensionality of the cost function's support across the three credit union types (since the number of outputs produced by credit unions is not the same across the three types). Further, note that, unlike an alternative model where a common cost relationship is assumed to be shared by all credit unions regardless of their service menus, the type-specific cost function (2.1) does not suffer from the problem of having to deal with many zeros reported for outputs that credit unions of a particular type do not produce. In our case, the cost function contains only those outputs that are produced by credit unions (of a given type) in practice and thus are non-zero. That is, if the credit union does not offer a given kind of output, the latter is treated not as a simple matter of the variable taking a zero value but rather as an indication that the variable does not enter the credit union's cost optimization problem whatsoever.[8]

2.1 Economies of Diversification

The credit-union-type-specific dual cost functions given in (2.1) can readily be used to compute economies of diversification exhibited in the US retail credit union sector. Here, we define diversification in terms of the scope/range of financial services that credit unions offer to their members. Such economies of diversification are said to exist if the average non-interest cost, which credit unions incur, tends to decline with the increase in the number of outputs produced. The latter may occur for a number of reasons including the potential to spread fixed costs, to reuse credit information about customers as well as to diversify risks across different assets (Berger et al., 1987). As it changes its product mix, the credit union might be able to lower its cost not only due to supply/cost-side factors, which we focus on in this paper, but also due to the change in the demand-side factors

[7] We call this a variable cost function since it only includes costs of the two variable inputs and may differ from the total operating cost, which is likely to include some other costs.

[8] For more on the zero-value observations problem, see Malikov et al. (2015b).

it faces (e.g., a different customer base). We acknowledge that our approach is unable to disentangle these two forces, and the measure of economies of diversification that we employ in this paper (discussed in detail below) captures their cumulative effect.

While also closely linked to the concept of subadditivity of the cost function, our definition of economies of diversification somewhat differs from Baumol et al.'s (1982) traditional notion of economies of scope that are said to exist if the cost of producing outputs individually is greater than the cost of their joint production. More concretely, consider the case of a credit union of Type 2 producing three outputs (y_1, y_3, y_4) via technology $\mathbb{T}_2$ with the corresponding cost function $C_2(\cdot)$. Using Baumol et al.'s (1982) definition, a Type 2 credit union enjoys economies of scope if

$$C_2(y_1, 0, 0) + C_2(0, y_3, 0) + C_2(0, 0, y_4) > C_2(y_1, y_3, y_4) \ , \tag{2.2}$$

where we have suppressed all other arguments of the cost function besides outputs. The assessment of (2.2) requires the computation of the counterfactual cost of producing each of the outputs separately. The counterfactual here is defined as the case of fully specialized production processes. The above approach naturally suffers from excessive extrapolation (Evans and Heckman, 1984; Hughes and Mester, 1993) since the counterfactual requires a projection of the estimated Type 2 cost function $C_2(\cdot)$ to the often non-existent case of complete specialization of the firms. Furthermore, not only does Baumol et al.'s (1982) measure extrapolate the estimated cost function to its boundary, but it also assumes that the specialized credit unions share the same technology and incur the same fixed costs as the integrated credit unions, which may be quite unrealistic in practice (see Fuss and Waverman, 2002).

To somewhat alleviate the problem of excess extrapolation, Berger et al. (1987) have proposed an expansion-path measure of subadditivity, whereby a Type 2 credit union is said to be "competitively viable" if

$$C_2\Big(\varpi_1^A y_1, \varpi_3^A y_3, \varpi_4^A y_4\Big) + C_2\Big(\varpi_1^B y_1, \varpi_3^B y_3, \varpi_4^B y_4\Big) + C_2\Big(\varpi_1^C y_1, \varpi_3^C y_3, \varpi_4^C y_4\Big) > C_2\big(y_1, y_3, y_4\big) \ , \tag{2.3}$$

where weights $\varpi_1^\kappa \geq 0$, $\varpi_3^\kappa \geq 0$ and $\varpi_4^\kappa \geq 0$ are such that $\varpi_m^A + \varpi_m^B + \varpi_m^C = 1$ for all $\kappa \in \{A, B, C\}$ and $m \in \{1, 3, 4\}$. This measure nests the conventional definition of scope economies in (2.2) as a special case. However, it does *not* allow the measurement of economies of scope or diversification *per se* in the instance of non-zero weights, since it would compare the cost of a larger multi-product firm and that of several smaller multi-product firms as opposed to comparing the cost of joint versus disjoint production processes (Grosskopf et al., 1992). Further, Berger et al. (1987) provide little systematic guidance on how to select the weights in practice.

In this paper, we generalize the approach of Grosskopf et al. (1992) who combine the conventional measure of economies of scope with that of expansion-path subadditivity, whereby we can measure the reduction in the credit union's cost, if any, as a result of diversifying its output scope while requiring a more realistic *partial* (as opposed to full) specialization of credit unions. Using this measure, we are able to significantly mute the noise in the estimates of counterfactuals.

More specifically, the credit union types, which we have defined on the basis of the output mix (financial service menu), imply a particular direction in which credit unions diversify in practice. The three credit union types are evidently "nested" with a distinct ordering: a shift from Type 1 to Type 2 implies the diversification by offering an additional product (y_1), and a shift from Type 2 to Type 3 requires the offering of yet another product (y_2). Thus, rather than computing economies of diversification defined as the benefit/loss of consolidating *non-existent* completely specialized credit unions, we seek to compute the benefit/loss of a *plausible* diversification process where credit unions evolve from Type 1 to Type 2 and, further, from Type 2 to Type 3.

For instance, for Type 2 credit unions, we compare the fitted cost of producing their actual bundle of outputs $\boldsymbol{y}_2 = (y_1, y_3, y_4)$ with the cost of splitting the production into two separate processes: the specialized Type 1 production of $\boldsymbol{y}_1 = (y_3, y_4)$ and the new relatively more specialized Type 2 production exhibiting a *higher* degree of specialization in y_1. Specifically, the Type 2 credit union is said to enjoy economies of diversification if

$$C_1\big(\varpi_3 y_3, \varpi_4 y_4\big) + C_2\Big(y_1, (1-\varpi_3)y_3, (1-\varpi_4)y_4\Big) > C_2\big(y_1, y_3, y_4\big) \tag{2.4}$$

for some $0 \le \varpi_3 \le 1$ and $0 \le \varpi_4 \le 1$. Note that the more appropriate Type 1 (rather than Type 2) cost function is used when evaluating the cost of producing the Type 1 bundle $\boldsymbol{y}_1$. Our approach differs from a conventional one because we do not compare the cost of producing $\boldsymbol{y}_2$ with the cost of producing y_1, y_3 and y_4 individually, *all* predicted using the Type 2 cost function. Further, our measure in (2.4) does not require the computation of the cost for a non-existent *fully* specialized credit union producing exclusively y_1 (unless $\varpi_3 = \varpi_4 = 1$). Our counterfactual is therefore "less counter" to reality. Combined with the fact that we employ the more appropriate Type 1 cost function when evaluating the cost of producing $\boldsymbol{y}_1$, the latter renders our methodology more realistic and improved than that routinely employed in the vast number of papers that seek to assess economies of scope/diversification in industries with no fully specialized firms by means of Baumol et al.'s (1982) traditional formula à la (2.2), which necessitates the extrapolation of an integrated firm's multi-product cost function to the case of single-product full specialization.

To measure economies of diversification in practice, we first need to specify non-negative weights ϖ_3 and ϖ_4 that control the split of common outputs (y_3, y_4). To further avoid excessive extrapolation, we follow Evans and Heckman (1984) and restrict the choice of weights ϖ_3 and ϖ_4 to the "admissible region" defined by the two data-determined constraints. First, we ensure that each counterfactual credit union does not produce less of each output than credit unions do in our sample. That is, when measuring economies of diversification for Type 2 credit unions in (2.4), we require that $\varpi_3 y_3 \ge \underline{y}_3$, $(1-\varpi_3)y_3 \ge \underline{y}_3$, $\varpi_4 y_4 \ge \underline{y}_4$ and $(1-\varpi_4)y_4 \ge \underline{y}_4$, where the underscored variables denote sample minima, i.e., $\underline{y}_m = \min\{y_m\}$ for all $m = 1, \ldots, 4$.[9] To make this constraint more operable, we redefine the (normalized) measure of economies of diversification (ED) as a credit union expands its service menu from Type 1 to Type 2 as follows

$$ED_{1\to 2} = \frac{\Big[C_1\big(\varpi_3 y_3^* + \underline{y}_3, \varpi_4 y_4^* + \underline{y}_4\big) + C_2\Big(y_1, (1-\varpi_3)y_3^* + \underline{y}_3, (1-\varpi_4)y_4^* + \underline{y}_4\Big)\Big] - C_2\big(y_1, y_3, y_4\big)}{C_1\big(\varpi_3 y_3^* + \underline{y}_3, \varpi_4 y_4^* + \underline{y}_4\big) + C_2\Big(y_1, (1-\varpi_3)y_3^* + \underline{y}_3, (1-\varpi_4)y_4^* + \underline{y}_4\Big)} ,$$

where $y_m^* = y_m - 2\underline{y}_m$ for all $m = 1, \ldots, 4$. The ED measures the within-sample economies of diversification, positive (negative) values of which correspond to economies (*dis*economies) of diversification. The case of $ED = 0$ suggests invariance to the scope of outputs.

The second constraint defining the admissible region ensures that each counterfactual credit union does not specialize in either output to a greater extent than credit unions do in our sample. In other words, we ensure that ratios of outputs for each counterfactual credit union fall in the range of such ratios observed in the sample (Evans and Heckman, 1984):

$$\begin{aligned} \min\left\{\frac{y_3}{y_4}\right\} &\le \frac{\varpi_3 y_3^* + \underline{y}_3}{\varpi_4 y_4^* + \underline{y}_4} \le \max\left\{\frac{y_3}{y_4}\right\} , \\ \min\left\{\frac{y_3}{y_4}\right\} &\le \frac{(1-\varpi_3) y_3^* + \underline{y}_3}{(1-\varpi_4) y_4^* + \underline{y}_4} \le \max\left\{\frac{y_3}{y_4}\right\} . \end{aligned} \tag{2.5}$$

[9]Thus, we measure the *within-sample* economies of diversification (also see Hughes and Mester, 1993).

Similar to the $ED_{1\to2}$ measure above, we can also define the (normalized) measure of economies of diversification as a credit union expands its service menu from Type 2 to Type 3, i.e.,

$$ED_{2\to3} = \frac{\Big[C_2\big(\varpi_1 y_1^* + \underline{y}_1, \varpi_3 y_3^* + \underline{y}_3, \varpi_4 y_4^* + \underline{y}_4\big) + C_3\Big((1-\varpi_1)y_1^* + \underline{y}_1, y_2, (1-\varpi_3)y_3^* + \underline{y}_3, (1-\varpi_4)y_4^* + \underline{y}_4\Big)\Big] - C_3\big(y_1, y_2, y_3, y_4\big)}{C_2\big(\varpi_1 y_1^* + \underline{y}_1, \varpi_3 y_3^* + \underline{y}_3, \varpi_4 y_4^* + \underline{y}_4\big) + C_3\Big((1-\varpi_1)y_1^* + \underline{y}_1, y_2, (1-\varpi_3)y_3^* + \underline{y}_3, (1-\varpi_4)y_4^* + \underline{y}_4\Big)}$$

such that, in addition to the constraints in (2.5), the following is satisfied:

$$\begin{aligned}
&\min\left\{\frac{y_1}{y_3}\right\} \le \frac{\varpi_1 y_1^* + \underline{y}_1}{\varpi_3 y_3^* + \underline{y}_3} \le \max\left\{\frac{y_1}{y_3}\right\}\ ,\\
&\min\left\{\frac{y_1}{y_3}\right\} \le \frac{(1-\varpi_1) y_1^* + \underline{y}_1}{(1-\varpi_3) y_3^* + \underline{y}_3} \le \max\left\{\frac{y_1}{y_3}\right\}\ ,\\
&\min\left\{\frac{y_1}{y_4}\right\} \le \frac{\varpi_1 y_1^* + \underline{y}_1}{\varpi_4 y_4^* + \underline{y}_4} \le \max\left\{\frac{y_1}{y_4}\right\}\ ,\\
&\min\left\{\frac{y_1}{y_4}\right\} \le \frac{(1-\varpi_1) y_1^* + \underline{y}_1}{(1-\varpi_4) y_4^* + \underline{y}_4} \le \max\left\{\frac{y_1}{y_4}\right\}
\end{aligned} \tag{2.6}$$

for some $0 \le \varpi_1 \le 1$. Also note that our measure of ED has an upper bound of one by construction, which facilitates a more meaningful comparison of $ED_{1\to2}$ and $ED_{2\to3}$.

Clearly, our ED measures are likely to be sensitive to the choice of weights $(\varpi_1, \varpi_3, \varpi_4)$. To circumvent this problem, we extend Grosskopf et al.'s (1992) approach in the spirit of Evans and Heckman (1984). Specifically, we propose a conservative approach to measuring economies of diversification, whereby we choose $(\varpi_1, \varpi_3, \varpi_4)$ that provide the smallest ED estimate in the admissible region defined by the credit union's observed output vector. Specifically, we measure the "global" economies of diversification (for each credit-union-year) as follows:

$$\begin{aligned}
GED_{1\to2} &\overset{\text{def}}{=} \inf_{\varpi_3,\varpi_4} ED_{1\to2}(\varpi_3, \varpi_4)\ ,\\
GED_{2\to3} &\overset{\text{def}}{=} \inf_{\varpi_1,\varpi_3,\varpi_4} ED_{2\to3}(\varpi_1, \varpi_3, \varpi_4)\ .
\end{aligned} \tag{2.7}$$

If economies of diversification of the *smallest* value are still statistically significantly above zero, we can conclude that economies of diversification are "globally" significant over the credit union's entire (admissible) output space in a given year.

3 Econometric Model

3.1 Four-way Heterogeneity

In a pursuit of robust estimates of economies of diversification and scale for credit unions, our econometric strategy is to devise a flexible model that is capable of accommodating as many sources of heterogeneity among credit unions as possible while still maintaining (partially parametric) structure for the sake of a better rate of convergence, computational ease and tractability.

As discussed in Section 2, we estimate the credit union cost function that explicitly accounts for observed heterogeneity across credit unions providing different financial services to their members. This renders the estimated cost relationship *credit-union-type*-specific. However, the cost structure is also unlikely to be common to all credit unions *within* a given type, which, if overlooked, can distort our results as well. For instance, Wheelock and Wilson (2011) argue at length that the size of credit unions (commonly measured by total assets) matters considerably in shaping their cost structure and that any parametric specification of the cost function that overlooks this relationship is thus likely to suffer from parameter instability. We concur with this sentiment and agree that it may be inappropriate to assume that the cost structure of a small credit union is the same as that of a large credit union, even if both belong to the same credit union type. To accommodate this heterogeneity among credit unions of different sizes, we follow Malikov et al. (2016) and allow parameters of the credit-union-type-specific cost function (2.1) to also vary (in an unspecified, nonparametric way) with the size of credit unions. Such a specification yields *credit-union*-specific estimates of the cost function parameters.[10]

The cost structure of credit unions is also likely to vary across the cost distribution itself. It is quite plausible for two credit unions of the same service menu type and the same asset size to still incur different costs due to differences in the input prices and/or in the distribution of outputs across output types. Thus, the two can exhibit quite distinct cost structures. Estimation of the cost function at the mean is however likely to mask these potential heterogeneity across higher and lower quantiles of the (conditional) credit union cost distribution. Further, mean estimation methods are sensitive to the presence of outliers in the data. Therefore, our econometric approach is to model the conditional quantiles of credit union costs as opposed to a mere conditional mean.

The cost performance of credit unions naturally depends on many other *unobserved* factors that are unique to an institution. For instance, one may expect a credit union run by a skilled experienced manager to operate at a lower cost than a credit union managed by a novice, *ceteris paribus*. Ignoring this unobserved heterogeneity when estimating the credit union's cost function (as customarily done in the existing literature) may produce inconsistent estimates since unobserved heterogeneity is likely to be correlated with covariates present in the estimated equation. While such credit-union-specific unobserved effects cannot be accounted for in a cross-sectional setting due to the incidental parameters problem, we address this issue in our case by taking advantage of the panel structure of the data.

In what follows, we describe a flexible semiparametric smooth coefficient quantile panel data model with correlated effects which we employ in order to robustly estimate the cost structure of credit unions. Our approach is capable of accommodating all four above-mentioned kinds of heterogeneity among credit unions, namely: *(i)* heterogeneity across different service menu types, *(ii)* heterogeneity across credit unions of different sizes, *(iii)* heterogeneity across the (conditional) quantiles of the cost distribution, and *(iv)* unobserved heterogeneity correlated with covariates in the estimated equation.[11]

3.2 Semiparametric Smooth Coefficient Quantile Panel Data Model

As opposed to a traditional approach to modeling the credit union cost by specifying its conditional mean (e.g., Frame and Coelli, 2001; Frame et al., 2003; Wheelock and Wilson, 2011), we

[10] Another way to allow the cost function to be credit-union-specific is to specify the latter as an unknown nonparametric function. For instance, Wheelock and Wilson (2011, 2013) favor such a specification (although they fit a common cost function for all credit unions regardless of their different financial service menus). While being a valid alternative, such a *fully* nonparametric local-approximation approach is however not only more computationally intensive but it also exhibits a slower rate of convergence given that it smooths over a larger number of covariates.

[11] As rightly pointed out by a referee, the first two types of heterogeneity — namely, heterogeneity across different credit union types and that across credit unions of different sizes within a given type — may be codependent implying that credit unions might be endogenously selecting their financial service menus (i.e., types) based on their size, among other factors. Given the already high complexity of the model, we abstract away from such a possibility in this paper. However, the issue of such a selectivity is explored in Malikov et al. (2015b).

instead model the conditional quantiles of the credit union cost distribution. Specifically, letting $\mathbb{F}_r(\log(C_{r,it})|\boldsymbol{y}_{r,it}, \widetilde{\boldsymbol{y}}_{it}, \boldsymbol{w}_{it})$ denote the conditional probability distribution of the log of variable non-interest cost of a credit union i of type r in time period t, the conditional quantile function of the log of cost given the conditioning set $(\boldsymbol{y}_{r,it}, \widetilde{\boldsymbol{y}}_{it}, \boldsymbol{w}_{it})$ for any quantile $0 < \tau < 1$ is defined as

$$\mathbb{Q}_\tau\left[\log(C_{r,it})|\boldsymbol{y}_r, \widetilde{\boldsymbol{y}}, \boldsymbol{w}\right] \stackrel{\text{def}}{=} \inf\left\{\log(C_{r,it}) :\ \mathbb{F}_r\left(\log(C_{r,it})|\boldsymbol{y}_{r,it}, \widetilde{\boldsymbol{y}}_{it}, \boldsymbol{w}_{it}\right) \geq \tau\right\}\ . \tag{3.1}$$

The quantile function $\mathbb{Q}_\tau[\cdot]$ can equivalently be expressed as (Koenker and Bassett, 1978)

$$\mathbb{Q}_\tau[\log(C_{r,it})|\boldsymbol{y}_r, \widetilde{\boldsymbol{y}}, \boldsymbol{w}] \stackrel{\text{def}}{=} \underset{q}{\operatorname{argmin}}\ \mathbb{E}\left[\rho_\tau\{\log(C_{r,it}) - q\}\right]\ , \tag{3.2}$$

where $\rho_\tau\{\cdot\}$ is the so-called "check function" defined as

$$\rho_\tau\{\zeta\} \stackrel{\text{def}}{=} \zeta(\tau - \mathbb{1}\{\zeta < 0\}) \tag{3.3}$$

for some $\zeta \in \mathbb{R}$, where $\mathbb{1}\{\cdot\}$ is the indicator function (also see Koenker and Hallock, 2001; Koenker, 2005).

We specify the unknown conditional quantile function $\mathbb{Q}_\tau[\cdot]$ as a *semiparametric* second-order expansion in logs where we let the coefficients be unknown functions of a "contextual" variable z_{it} as well as allow for unobserved credit-union-specific effects μ_i:

$$\begin{aligned}\mathbb{Q}_\tau[\log(C_{r,it})|\cdot] = \beta_{r,0,\tau}(z_{it}) + \log(\mathbf{v}_{r,it})'\boldsymbol{\beta}_{r,\tau}(z_{it}) + \tfrac{1}{2}\log(\mathbf{v}_{r,it})'\mathbf{B}_{r,\tau}(z_{it})\log(\mathbf{v}_{r,it}) + \mu_i\ ,\\ i = 1, \ldots, N;\ t = 1, \ldots, T;\ r = 1, 2, 3\end{aligned} \tag{3.4}$$

where $\mathbf{v}_{r,it} = (\boldsymbol{y}'_{r,it}, \widetilde{\boldsymbol{y}}'_{it}, \boldsymbol{w}'_{it})'$ is a $J_r \times 1$ vector of the (strictly exogenous) arguments to the cost function for credit unions of type r;[12] $\beta_{r,0,\tau}(\cdot)$ is an unknown smooth scalar intercept function; and $\boldsymbol{\beta}_{r,\tau}(\cdot) = (\beta_{r,1,\tau}(\cdot), \ldots, \beta_{r,J_r,\tau}(\cdot))'$ and $\mathbf{B}_{r,\tau}(\cdot) = [\beta_{r,jj,\tau}(\cdot) :\ j = 1, \ldots, J_r]$ are a $J_r \times 1$ vector and a $J_r \times J_r$ symmetric matrix of unknown smooth parameter functions, respectively. For convenience, we define a collection of these parameter functions $\mathcal{B}_{r,\tau}(\cdot)$, i.e.,

$$\mathcal{B}_{r,\tau}(\cdot) \equiv \{\beta_{r,0,\tau}(\cdot),\ \boldsymbol{\beta}_{r,\tau}(\cdot),\ \mathbf{B}_{r,\tau}(\cdot)\}\ . \tag{3.5}$$

A few remarks are warranted about our model in (3.4). First, all unknown parameter functions are indexed by r which allows them to vary with the credit union type defined by the service menu offered. Hence, the dimension of $\boldsymbol{\beta}_{r,\tau}(\cdot)$, $\mathbf{B}_{r,\tau}(\cdot)$ as well as of the covariate vector $\mathbf{v}_{r,it}$ is varying with the credit union type too. Second, the unknown parameter functions $\mathcal{B}_{r,\tau}(\cdot)$ are quantile-specific which permit us to accommodate heterogeneity across credit unions belonging to different (conditional) quantiles of the cost distribution. Third, we model unknown coefficients $\mathcal{B}_{r,\tau}(\cdot)$ as smooth nonparametric functions of a (scalar) "contextual" variable z_{it}, which in our case is the (log of) total asset size of credit unions, as discussed in Section 2. This formulation permits us to obtain credit-union-specific estimates of unknown coefficients. Here, we do not index variable z_{it} by the credit union type r since it pertains to credit unions of *all* types.[13] Lastly, we allow the unobserved credit-union-specific effects μ_i to be correlated with the right-hand-side covariates.

[12] During the estimation, we also include the time trend as an additional covariate in $\mathbf{v}_{r,it}$, which we obviously do not log. See Section 5 for more details.

[13] Note that, in general, the contextual variable z_{it} need not be a scalar, and the coefficient functions $\mathcal{B}_{r,\tau}(\cdot)$ can be specified as a function of a vector of variables.

Model (3.4) is a semiparametric smooth coefficient quantile generalization of the standard translog cost function in the panel data setting. It closely relates to a *partially* varying coefficient quantile panel data model recently proposed by Cai et al. (2013), from whom we differ by excluding constant coefficient terms.[14] While model (3.4) is not as flexible as a fully nonparametric quantile regression,[15] our model is however less prone to the "curse of dimensionality", which all nonparametric methods suffer from (Li and Racine, 2007; Ichimura and Todd, 2007). Specifically, when employing optimal bandwidths, the asymptotic integrated mean square error (IMSE) of a fully nonparametric local-linear estimator converges with rate $N^{-4/(J_r+4)}$ which is significantly slower than $N^{-4/5}$ exhibited by our semiparametric estimator that smooths over a single (scalar) "contextual" covariate.[16] Further, model (3.4) is still flexible enough to nest several popular parametric and semiparametric specifications. For instance, the case of $\mu_i = 0$ and $z_{it} = 1$ for all i and t, which renders constant parameters in (3.4), corresponds to a traditional pooled (parametric) linear quantile regression model of Koenker and Bassett (1978). Allowing z_{it} to vary across observations (but maintaining $\mu_i = 0$) transforms our model into a standard smooth coefficient quantile regression model (e.g., Kim, 2007; Cai and Xu, 2008; Cai and Xiao, 2012). If we further restrict all parameters other than the intercept $\beta_{r,0,\tau}(\cdot)$ to be constant, our model would collapse to a (pooled) partially linear quantile regression model of Lee (2003). On the other hand, maintaining the restriction $z_{it} = 1$ for all i and t but allowing for unobserved heterogeneity μ_i, model (3.4) collapses to a parametric counterpart considered by Koenker (2004) who treats μ_i as fixed effects.

Before we proceed to the discussion of the estimation methodology, we first need to formulate the treatment of unobserved effects μ_i. Following Abrevaya and Dahl (2008) and Cai et al. (2013), who all consider quantile panel data models in the presence of unobserved heterogeneity, we specify μ_i as correlated (random) effects which permits it to correlate with $\mathbf{v}_{r,it}$ and z_{it}. More specifically, we can approximate unobserved effects μ_i conditional on covariates in the spirit of Chamberlain (1980) by specifying the latter in the form of a semiparametric linear projection on lags and leads of observables:

$$\begin{aligned}\mu_i =&\ \log(\mathbf{v}_{r,i1})'\boldsymbol{\delta}_{r,1}(z_{it}) + \cdots + \log(\mathbf{v}_{r,iT})'\boldsymbol{\delta}_{r,T}(z_{it}) \\ \stackrel{\text{def}}{=}&\ \log(\mathbf{v}_{r,i})'\boldsymbol{\delta}_{r}(z_{it}) \ ,\end{aligned} \tag{3.6}$$

where, for convenience, we have defined $\log(\mathbf{v}_{r,i}) \stackrel{\text{def}}{=} \log(\mathbf{v}'_{r,i1}, \ldots, \mathbf{v}'_{r,it}, \ldots, \mathbf{v}'_{r,iT})'$ and a corresponding vector of auxiliary parameter functions of a conformable dimension $\boldsymbol{\delta}_r(\cdot) \stackrel{\text{def}}{=} (\boldsymbol{\delta}'_{r,1}(\cdot), \ldots, \boldsymbol{\delta}'_{r,t}(\cdot), \ldots, \boldsymbol{\delta}'_{r,T}(\cdot))'$. Note that, in the case of small T, one may alternatively opt for a more parsimonious specification of correlated effects à la Mundlak (1978) by restricting $\boldsymbol{\delta}_{r,1}(\cdot) = \boldsymbol{\delta}_{r,t}(\cdot) = \boldsymbol{\delta}_{r,T}(\cdot)$ which implies the following formulation

$$\mu_i = \overline{\log(\mathbf{v}_{r,i})}'\boldsymbol{\varrho}_r(z_{it}) \ , \tag{3.7}$$

where $\overline{\log(\mathbf{v}_{r,i})} \stackrel{\text{def}}{=} T^{-1}\sum_t \log(\mathbf{v}_{r,it})$ is a $J_r \times 1$ vector of time averages and $\boldsymbol{\varrho}_r(\cdot)$ is a conformable

[14] That is, based on Cai et al.'s (2013) specification, some of the parameter functions in $\mathcal{B}_{r,\tau}(\cdot)$ would have been assumed to be constant and not dependent on a "contextual" variable z_{it}. Given our pursuit of a full parameter heterogeneity in the cost function across credit unions of different sizes, restricting some coefficients in (3.4) to be size-invariant seems to be rather unwarranted. Furthermore, since there is no a priori knowledge about which parameters might be constant, any designation of them would be ad hoc.

[15] To our knowledge, a fully nonparametric quantile panel data regression model, which allows for unobserved effects correlated with the right-hand-side covariates, is yet to be considered in the literature. For an excellent review of the existing *pooled* nonparametric (kernel and spline) quantile regression models, see Cai et al. (2009).

[16] The number of continuous arguments of the cost function for the Type r credit unions, which a fully nonparametric estimator is to smooth over, is $J_r \geq 7$. To keep things in perspective, recall that the parametric rate for the IMSE is N^{-1}.

vector of auxiliary parameter functions.[17] Given that, in our empirical application, T is particularly small relative to N, in this paper we prefer this parsimonious Mundlak-type parameterization over that in (3.6). In fact, since (3.6) nests (3.7), we can formally discriminate between the two specifications by means of a nonparametric likelihood-ratio test, which we discuss in more details in Section 5.

Thus, making use of the parameterization of unobserved effects μ_i in (3.7), from (3.4) we get the following estimatable expression for the conditional quantile function:

$$\begin{aligned}\mathbb{Q}_\tau[\log(C_{r,it})|\cdot] = \beta_{r,0,\tau}(z_{it}) + \log(\mathbf{v}_{r,it})'\boldsymbol{\beta}_{r,\tau}(z_{it}) +& \\ \tfrac{1}{2}\log(\mathbf{v}_{r,it})'\mathbf{B}_{r,\tau}(z_{it})\log(\mathbf{v}_{r,it}) + \overline{\log(\mathbf{v}_{r,i})}'\boldsymbol{\varrho}_{r,\tau}(z_{it})& \ . \end{aligned} \tag{3.8}$$

Before we proceed, for the ease of notation, we define an augmented collection of all parameter functions in (3.8), i.e.,

$$\mathscr{B}_{r,\tau}(\cdot) \equiv \{\mathcal{B}_{r,\tau}(\cdot),\ \boldsymbol{\varrho}_{r,\tau}(\cdot)\} = \{\beta_{r,0,\tau}(\cdot),\ \boldsymbol{\beta}_{r,\tau}(\cdot),\ \mathbf{B}_{r,\tau}(\cdot),\ \boldsymbol{\varrho}_{r,\tau}(\cdot)\} \ . \tag{3.9}$$

Estimation. We estimate the unknown coefficient functions $\mathscr{B}_{r,\tau}(\cdot)$ in (3.8) using local-linear kernel fitting, which has numerous well-documented advantages over a more commonly used local-constant estimator that suffers from non-adaptation and boundary effects. Specifically, under the assumption that smooth coefficient functions are twice differentiable in the neighborhood of z, the jth element in $\mathscr{B}_{r,\tau}(\cdot)$ can then be approximated using the first-order Taylor expansion around $z_{it} = z$, i.e.,

$$\mathscr{B}_{r,j,\tau}(z_{it}) \approx \mathscr{B}_{r,j,\tau}(z) + \nabla\mathscr{B}_{r,j,\tau}(z)(z_{it} - z) \ , \tag{3.10}$$

where $\nabla\mathscr{B}_{r,j,\tau}(z)$ is the first derivatives of $\mathscr{B}_{r,j,\tau}(z_{it})$ with respect to z_{it}, evaluated at z.

We start by vectorizing our model in (3.8):

$$\begin{aligned}\mathbb{Q}_\tau[\log(C_{r,it})|\cdot] = \beta_{r,0,\tau}(z_{it}) + \log(\mathbf{v}_{r,it})'\boldsymbol{\beta}_{r,\tau}(z_{it}) +& \\ \tfrac{1}{2}\mathrm{vec}\left\{\log(\mathbf{v}_{r,it})\log(\mathbf{v}_{r,it})'\right\}'\mathrm{vec}\left\{\mathbf{B}_{r,\tau}(z_{it})\right\} + \overline{\log(\mathbf{v}_{r,i})}'\boldsymbol{\varrho}_{r,\tau}(z_{it})& \ , \end{aligned} \tag{3.11}$$

where vec$\{\cdot\}$ denotes the column-wise vectorization operator. Next, for each subset of $\mathscr{B}_{r,\tau}(\cdot)$ from (3.9), respectively, we define

$$\boldsymbol{\alpha}_{r,\tau}(z) \overset{\text{def}}{=} \begin{bmatrix}\beta_{r,0,\tau}(z) & \nabla\beta_{r,0,\tau}(z)\end{bmatrix} \tag{3.12a}$$

$$\boldsymbol{\gamma}_{r,\tau}(z) \overset{\text{def}}{=} \begin{bmatrix}\boldsymbol{\beta}_{r,\tau}(z) & \nabla\boldsymbol{\beta}_{r,\tau}(z)\end{bmatrix} = \begin{bmatrix}\beta_{r,1,\tau}(z) & \nabla\beta_{r,1,\tau}(z) \\ \vdots & \vdots \\ \beta_{r,J_r,\tau}(z) & \nabla\beta_{r,J_r,\tau}(z)\end{bmatrix} \tag{3.12b}$$

$$\boldsymbol{\Gamma}_{r,\tau}(z) \overset{\text{def}}{=} \begin{bmatrix}\mathrm{vec}\left\{\mathbf{B}_{r,\tau}(z)\right\} & \nabla\mathrm{vec}\left\{\mathbf{B}_{r,\tau}(z)\right\}\end{bmatrix} = \begin{bmatrix}\beta_{r,11,\tau}(z) & \nabla\beta_{r,11,\tau}(z) \\ \vdots & \vdots \\ \beta_{r,J_rJ_r,\tau}(z) & \nabla\beta_{r,J_rJ_r,\tau}(z)\end{bmatrix} \tag{3.12c}$$

$$\boldsymbol{\vartheta}_{r,\tau}(z) \overset{\text{def}}{=} \begin{bmatrix}\boldsymbol{\varrho}_{r,\tau}(z) & \nabla\boldsymbol{\varrho}_{r,\tau}(z)\end{bmatrix} = \begin{bmatrix}\varrho_{r,1,\tau}(z) & \nabla\varrho_{r,1,\tau}(z) \\ \vdots & \vdots \\ \varrho_{r,J_r,\tau}(z) & \nabla\varrho_{r,J_r,\tau}(z)\end{bmatrix} \ . \tag{3.12d}$$

[17]For a similar Mundlak-type specification of correlated effects in parametric panel data models, e.g., see Malikov and Kumbhakar (2014).

Defining a 2×1 "vector of deviations" $\mathcal{Z}_{it}(z) \stackrel{\text{def}}{=} (1, z_{it} - z)'$, we can approximate (3.11) by

$$\mathbb{Q}_\tau[\log(C_{r,it})|\cdot] \approx \boldsymbol{\alpha}_{r,\tau}(z)\mathcal{Z}_{it}(z) + \log(\mathbf{v}_{r,it})'\boldsymbol{\gamma}_{r,\tau}(z)\mathcal{Z}_{it}(z) + \\ \tfrac{1}{2}\text{vec}\left\{\log(\mathbf{v}_{r,it})\log(\mathbf{v}_{r,it})'\right\}'\boldsymbol{\Gamma}_{r,\tau}(z)\mathcal{Z}_{it}(z) + \overline{\log(\mathbf{v}_{r,i})}'\boldsymbol{\vartheta}_{r,\tau}(z)\mathcal{Z}_{it}(z) \ , \qquad (3.13)$$

where we have replaced $(\beta_{r,0,\tau}(z_{it}),\ \boldsymbol{\beta}_{r,\tau}(z_{it}),\ \text{vec}\{\mathbf{B}_{r,\tau}(z_{it})\},\ \boldsymbol{\varrho}_{r,\tau}(z_{it}))$ appearing in (3.11) with $(\boldsymbol{\alpha}_{r,\tau}(z)\mathcal{Z}_{it}(z),\ \boldsymbol{\gamma}_{r,\tau}(z)\mathcal{Z}_{it}(z),\ \boldsymbol{\Gamma}_{r,\tau}(z)\mathcal{Z}_{it}(z),\ \boldsymbol{\vartheta}_{r,\tau}(z)\mathcal{Z}_{it}(z))$, respectively.

We estimate (a superset of) parameter functions in (3.13) (including the gradient functions), i.e.,

$$\mathscr{G}_{r,\tau}(\cdot) \equiv \{\mathcal{B}_{r,\tau}(\cdot),\ \nabla\mathcal{B}_{r,\tau}(\cdot)\} \equiv \{\boldsymbol{\alpha}_{r,\tau}(\cdot),\ \boldsymbol{\gamma}_{r,\tau}(\cdot),\ \boldsymbol{\Gamma}_{r,\tau}(\cdot),\ \boldsymbol{\vartheta}_{r,\tau}(\cdot)\} \ , \qquad (3.14)$$

from the following (local) kernel-weighted semiparametric quantile regression problem for a given prespecified quantile of interest τ:

$$\min_{\mathscr{G}_{r,\tau}(z)} \sum_i \sum_t \rho_\tau\Big\{\log(C_{r,it}) - \boldsymbol{\alpha}_{r,\tau}(z)\mathcal{Z}_{it}(z) - \log(\mathbf{v}_{r,it})'\boldsymbol{\gamma}_{r,\tau}(z)\mathcal{Z}_{it}(z) - \\ \tfrac{1}{2}\text{vec}\left\{\log(\mathbf{v}_{r,it})\log(\mathbf{v}_{r,it})'\right\}'\boldsymbol{\Gamma}_{r,\tau}(z)\mathcal{Z}_{it}(z) - \overline{\log(\mathbf{v}_{r,i})}'\boldsymbol{\vartheta}_{r,\tau}(z)\mathcal{Z}_{it}(z)\Big\}\ \mathcal{K}\left(\tfrac{z_{it}-z}{h_{r,\tau}}\right) \ , \qquad (3.15)$$

where $\mathcal{K}(\cdot)$ is a nonnegative symmetric kernel function such that $\int \mathcal{K}(v)dv = 1$ which provides weights in the estimation of $\mathscr{G}_{r,\tau}(z)$, and $h_{r,\tau}$ is the bandwidth controlling the degree of smoothing/weighting. The L_1-estimator defined in (3.15) is consistent and asymptotically normal (see Cai et al., 2013).[18] We solve the weighted objective function in (3.15) via linear programming using Barrodale and Roberts's (1974) modified simplex algorithm for L_1-regressions. The estimated parameters of the conditional τ-quantile are local to the credit-union-specific neighborhood $z_{it} = z$.

Bandwidth. We select the optimal smoothing parameter (bandwidth) for each (r, τ) pair via the data-driven cross-validation procedure. Unlike the three-stage *partially* smooth coefficient estimator of Cai et al. (2013), ours conveniently requires one stage only.[19] Hence, we do not need to preoccupy ourselves with the selection of the bandwidth more than once (for each r and τ) and to ensure that the first-stage estimates are under-smoothed.

To avoid over-smoothing in dense ranges of the support of the data while under-smoothing in sparse tails, which the fixed bandwidth parameter is well-known to produce, we opt to employ an adaptive bandwidth capable of adapting to the local distribution of the data.[20] Specifically, we use a $k_{r,\tau}$-nearest-neighbor bandwidth $h_{r,\tau}(z)$ defined as the Euclidean distance between some point z and the kth nearest neighbor of z among z_{it}.

The value of $k_{r,\tau}$ such that $k_{r,\tau} \to \infty$ and $k_{r,\tau}/N \to 0$ as $N \to \infty$ is selected via cross-validation for each (r, τ) pair. Due to the presence of unobserved effects, the standard leave-one-(observation)-out cross-validation is however unlikely to perform well in our case. We therefore instead consider a leave-one-*cross-section*-out cross-validation in order to select optimal bandwidth. Specifically, for each conditional τ-quantile estimation of the costs for r-type credit unions, we select the optimal

[18] Unlike Koenker (2004) who requires both N and T to go to infinity for consistency of their linear fixed effects quantile panel data estimators, the semiparametric estimator in (3.15) requires only $N \to \infty$ for consistency (for details, see Cai et al., 2013).

[19] Cai et al.'s (2013) partially linear model, owing to the fact that some coefficients in it are assumed to be constant and not varying with a contextual variable, requires a three-stage estimation procedure necessary to recover both the constant and varying coefficients.

[20] We are thankful to a referee for this suggestion.

$k_{r,\tau}$ by minimizing the following L_1-norm cross-validation objective function:[21]

$$\min_{k_{r,\tau}} \ \mathrm{CV}(k_{r,\tau}) \stackrel{\text{def}}{=} \sum_i \sum_t \rho_\tau \Big\{ \log(C_{r,it}) - \widehat{\mathbb{Q}}_{\tau,-it} \left[\log(C_{r,it}) | \log(\mathbf{v}_{r,it}), z_{it} \right] \Big\} \ , \tag{3.16}$$

where $\widehat{\mathbb{Q}}_{\tau,-it} [\log(C_{r,it})|\cdot]$ is the leave-one-cross-section-out estimate of the τth conditional quantile of $\log(C_{r,it})$, obtained by solving the following problem:

$$\begin{aligned} \min_{\mathscr{G}_{r,\tau}(z_{it})} \ & \sum_{j \neq i} \sum_t \rho_\tau \Big\{ \log(C_{r,jt}) - \boldsymbol{\alpha}_{r,\tau}(z_{it}) \mathcal{Z}_{jt}(z_{it}) - \log(\mathbf{v}_{r,jt})' \boldsymbol{\gamma}_{r,\tau}(z_{it}) \mathcal{Z}_{jt}(z_{it}) - \\ & \tfrac{1}{2} \mathrm{vec} \left\{ \log(\mathbf{v}_{r,jt}) \log(\mathbf{v}_{r,jt})' \right\}' \boldsymbol{\Gamma}_{r,\tau}(z_{it}) \mathcal{Z}_{jt}(z_{it}) - \overline{\log(\mathbf{v}_{r,j})}' \boldsymbol{\vartheta}_{r,\tau}(z_{it}) \mathcal{Z}_{jt}(z_{it}) \Big\} \ \mathcal{K} \left(\tfrac{z_{jt} - z_{it}}{h_{r,\tau}(z_{it})} \right) \ . \end{aligned} \tag{3.17}$$

Bootstrap Inference. Despite that the estimator given in (3.15) is asymptotically normal under fairly standard regularity conditions (Cai et al., 2013), it is well-documented that asymptotic inference is often unreliable in finite samples due to finite-sample biases as well as the first-order asymptotic theory's poor ability to approximate the distribution of estimators in finite samples (Horowitz, 2001). Bootstrap however provides means not only to reduce the estimator's finite-sample bias (and hence finite-sample MSE) but also to account for higher-order moments in the sampling distribution. Furthermore, the evaluation of our estimator's asymptotic covariance matrix, the need for which is obviated if one were to employ bootstrap, is quite cumbersome too. In order to conduct statistical inference, we therefore resort to wild (residual) bootstrap. However, Feng et al. (2011) show that a traditional wild bootstrap procedure is invalid for quantile estimators due to nonlinear score functions associated with the objective function (3.15). They propose a modified wild bootstrap algorithm to suit the asymmetric loss function $\rho_\tau\{\cdot\}$ used in the quantile estimation, which we adapt to a panel data setting to allow for the error correlation within cross-sectional clusters. The algorithm for the panel data consists of the following steps:

For each credit union type r and the prespecified quantile $0 < \tau < 1$:

(1) Estimate the model in (3.15). Denote the estimated parameter functions and residuals as $\widehat{\mathscr{G}}_{r,\tau}(z_{it})$ and $\widehat{\epsilon}_{r,\tau,it}$ for $i = 1, \ldots, N;\ t = 1, \ldots, T$, respectively.

(2) Generate weights $\{\omega_i;\ i = 1, \ldots, N\}$ from the following two-point mass distribution:

$$\omega_i = \begin{cases} 2(1-\tau) & \text{with prob.} \quad (1-\tau) \\ -2\tau & \text{with prob.} \quad \tau \end{cases} \ . \tag{3.18}$$

Next, generate the weighted bootstrap residuals $\epsilon^*_{r,\tau,it} = \omega_i \times |\widehat{\epsilon}_{r,\tau,it}|$, where $|\cdot|$ denotes the absolute value.

(3) Construct the bootstrap sample $\{\log(C_{r,it})^*, \log(\mathbf{v}_{r,it}), z_{it};\ i = 1, \ldots, N;\ t = 1, \ldots, T\}$, where

$$\begin{aligned} \log(C_{r,it})^* = \ & \widehat{\beta}_{r,0,\tau}(z_{it}) + \log(\mathbf{v}_{r,it})' \widehat{\boldsymbol{\beta}}_{r,\tau}(z_{it}) + \\ & \tfrac{1}{2} \log(\mathbf{v}_{r,it})' \widehat{\mathbf{B}}_{r,\tau}(z_{it}) \log(\mathbf{v}_{r,it}) + \overline{\log(\mathbf{v}_{r,i})}' \widehat{\boldsymbol{\varrho}}_{r,\tau}(z_{it}) + \epsilon^*_{r,\tau,it} \ . \end{aligned} \tag{3.19}$$

[21] For instance, Cai and Xu (2008) employ an alternative data-driven method of selecting optimal bandwidths for smooth coefficient quantile regression models via minimizing the bias-corrected AIC criterion. However, their bandwidths are not adaptive but fixed.

(4) Reestimate the model in (3.15) using the bootstrap sample from step (3).[22] Denote the bootstrap estimates of the parameter functions as $\{\widehat{\mathscr{G}}^{*}_{r,\tau}(z_{it});\ i = 1, \ldots, N, t = 1, \ldots, T\}$.

(5) Repeat steps (2)–(4) B times to obtain the bootstrap distribution of the estimator. For each observation $i = 1, \ldots, N;\ t = 1, \ldots, T$, the $(1 - a) \times 100\%$ bootstrap confidence interval for $\widehat{\mathscr{G}}_{r,\tau}(z_{it})$ can be estimated as the interval between the $[a/2 \times 100]$th and $[(1 - a/2) \times 100]$th percentiles of the corresponding bootstrap distribution consisting of B copies of $\widehat{\mathscr{G}}^{*}_{r,\tau}(z_{it})$.

4 Data

The data we use in this paper come from year-end Call Reports available from the National Credit Union Administration (NCUA), a federal regulatory body that supervises credit unions in the US While the data are available for both retail, or so-called natural-person, and corporate credit unions, we focus on retail credit unions only. We exclude corporate credit unions (whose customers are the retail credit unions) from the sample in order to minimize noise in the data due to apparent non-homogeneity between these two types of financial depositories.[23]

The sample period we consider in this paper runs from 2001 to 2006. We intentionally limit our analysis to a pre-crisis period in an attempt to minimize the influence of potential structural changes in the industry during the crisis and in its aftermath on the estimation results. Further, given the large wave of mergers and acquisitions taking place in the US credit union sector over our sample period (e.g., see Bauer et al., 2009), we limit ourselves to continuously operating retail credit unions only. The latter allows us to mitigate a potential impact of entries and exits on the estimation results. Lastly, since more than 99% of all retail credit unions operating in the US fall into either of the three types described in Section 2, we omit the remaining credit unions from our analysis in order to minimize heterogeneity in the sample.[24]

Prior to estimation, we also discard observations with negative values of outputs and total cost. Likewise, we exclude observations with non-positive values of variable input prices, quasi-fixed netputs, and total assets. Since $\widetilde{\mathbf{y}}$ and w_1 are interest rates, we follow Wheelock and Wilson (2011) and Malikov et al. (2016) and also eliminate those observations for which values of these variables lie outside the unit interval. These excluded observations are likely to be the result of erroneous data reporting. Our data sample thus consists of $N = 4,543$ retail credit unions continuously operating during the 2001–2006 period ($T = 6$). See Appendix of Malikov et al. (2015b), whom we follow closely, for the details on construction of the variables.

Table 1 reports summary statistics of the variables used in our analysis. We deflate all nominal stock variables to 2011 US dollars using the GDP Implicit Price Deflator. A comparison of sample mean and median estimates of variables shows clear differences between the credit union types. Expectedly, the size of the credit unions, proxied by total assets (z), increases as the range of financial services grows from Type 1 to Type 3. The dramatic differences between the three types favor our view that the assumption of a common cost relationship across credit union types is unlikely to be supported by the data and may result in the misspecification of the econometric model.

[22] As means of asymptotic refinement, it is generally advised to use a larger bandwidth in the bootstrap procedure (or, put differently, to undersmooth the "original data" estimator) in order to reduce the estimator's finite-sample bias and thereby finite-sample MSE, which helps improve coverage probabilities of confidence intervals (Härdle, 1990; Horowitz, 2001).

[23] This results in a loss of less than 0.7% of observations in the sample.

[24] The omitted credit unions are likely to be either outliers or reporting errors.

Table 1. Data Summary Statistics

Var.	1st Qu.	Median	Mean	3rd Qu.	1st Qu.	Median	Mean	3rd Qu.
	Type 1 Credit Unions				**Type 2 Credit Unions**			
C	73.68	144.36	216.42	263.29	437.22	1,025.48	2,901.19	2,706.89
y_1	—	—	—	—	1,197.56	4,608.46	22,562.79	15,866.66
y_2	—	—	—	—	—	—	—	—
y_3	1,221.78	2,473.37	3,604.14	4,534.62	5,009.36	10,650.50	32,030.02	27,116.22
y_4	231.51	867.51	2,027.92	2,536.62	2,238.88	6,386.39	24,132.22	17,242.05
$\widetilde{y}_5$	0.014	0.020	0.021	0.028	0.013	0.019	0.020	0.026
$1/\widetilde{y}_6$	0.075	0.088	0.091	0.102	0.073	0.084	0.086	0.096
w_1	0.013	0.018	0.020	0.024	0.016	0.021	0.022	0.026
w_2	26.73	37.79	38.10	47.65	40.53	47.73	49.72	56.46
z	218.13	4,590.68	6,708.35	8,570.89	12,667.46	28,635.80	89,728.72	72,546.33
% Obs.		19.18				64.78		
	Type 3 Credit Unions							
C	1,739.56	4,636.01	11,079.75	12,567.57				
y_1	11,908.82	38,566.46	135,761.60	112,633.60				
y_2	200.66	856.25	6,071.57	4,026.35				
y_3	15,462.30	38,959.35	117,129.10	112,538.60				
y_4	5,650.69	17,296.62	76,559.45	57,877.53				
$\widetilde{y}_5$	0.015	0.020	0.021	0.026				
$1/\widetilde{y}_6$	0.070	0.078	0.080	0.088				
w_1	0.016	0.020	0.021	0.024				
w_2	43.95	50.80	52.63	58.91				
z	48,975.53	121,641.40	374,025.90	363,062.00				
% Obs.		16.04						

NOTES: Credit union types are as defined in Section 2. C – total variable, non-interest cost; y_1 – real estate loans, y_2 – business and agricultural loans; y_3 – consumer loans; y_4 – investments; $\widetilde{y}_5$ – average interest rate on saving deposits; $\widetilde{y}_6$ – the *inverse* of average interest rate on loans; w_1 – price of capital; w_2 – price of labor; z – total assets. $(C, y_1, y_2, y_3, y_4, w_2, z)$ are in thousands of real 2011 US dollars; $(\widetilde{y}_5, 1/\widetilde{y}_6, w_1)$ are unit-free interest rates.

5 Estimation and Empirical Results

We estimate our (local-linear) semiparametric smooth coefficient quantile panel data model with correlated effects given in (3.15) for the three credit union types $r = \{1, 2, 3\}$ separately. For each of these types, we estimate the model for three pre-specified quantiles, namely, $\tau = \{0.25, 0.50, 0.75\}$. The latter quantiles correspond to the first, median and the third quartiles of the conditional distribution of credit union costs.

As briefly mentioned in Section 3.2, given that in our sample T is rather small relative to N, we formulate the unobserved correlated effects in our model using the parsimonious Mundlak-type device in (3.7) instead of the parameterization à la Chamberlain in (3.6). This choice of ours is strongly supported by the data across all credit union types. Specifically, note that the Chamberlain-type specification in (3.6) nests that of Mundlak given in (3.7) as a special case, implying that we can formally discriminate between the two by means of a specification test. In this paper, we make use of Koenker and Machado's (1999) likelihood-ratio test for quantile regressions based on the difference between the restricted (under the null) and unrestricted (under the alternative) sums of estimated check functions. Since our model is semiparametric, we use the *kernel-weighted* sums of estimated check functions in (3.15) when constructing the test statistic in the spirit of a nonparametric likelihood-ratio test by Ullah (1985). Given that kernel-based nonparametric

Table 2. Economies of Diversification Estimates

Model	*Point Estimates* 1st Qu.	Median	Mean	3rd Qu.	*Categories, %* ED<0	ED=0	ED>0
			Type 1 → Type 2				
$\mathbb{Q}_{0.25}$	0.444	0.481	0.452	0.499	0.26	8.96	90.78
$\mathbb{Q}_{0.50}$	0.455	0.483	0.452	0.496	0.07	8.57	91.36
$\mathbb{Q}_{0.75}$	0.392	0.450	0.350	0.471	4.00	8.63	87.37
			Type 2 → Type 3				
$\mathbb{Q}_{0.25}$	0.419	0.473	0.449	0.505	1.19	51.31	47.50
$\mathbb{Q}_{0.50}$	0.264	0.388	0.357	0.494	1.46	71.70	26.84
$\mathbb{Q}_{0.75}$	0.267	0.368	0.353	0.436	2.33	70.92	26.75

NOTE: Percentage points may not sum up to a hundred due to rounding.

tests usually suffer from substantial finite-sample size distortions, we bootstrap our test statistic to obtain its null distribution [see Lee and Ullah (2003) for the description of the appropriate bootstrap procedure]. For the semiparametric conditional median regression model ($\tau = 0.50$), we consistently fail to reject the null of the Mundlak-type specification across all three credit union types with the p-value being at least as large as 0.998.

The model (3.15) is estimated subject to symmetry restrictions on $\mathbf{\Gamma}_{r,\tau}(\cdot)$ [owing to symmetry of $\mathbf{B}_{r,\tau}(\cdot)$ in (3.8)] as well as the linear homogeneity of the cost function in input prices. We impose the latter by dividing the cost and input prices by the price of labor w_2. Throughout, we use the second-order Gaussian kernel, and the optimal adaptive k-nearest-neighbor bandwidth for each (r, τ) pair is selected via the data-driven cross-validation as discussed in Section 3.

In addition to standard arguments to the cost function under the "service provision approach" discussed in Section 2 — output quantities, quasi-fixed netputs and input prices $(\boldsymbol{y}_{r,it}, \widetilde{\boldsymbol{y}}_{it}, \boldsymbol{w}_{it})$ — we also include the time trend t (along with its square and cross-product terms) in the specification of the conditional quantile in (3.4) in order to capture temporal shifts in the credit union cost relationship. Lastly, since our model already allows for heterogeneity in many dimensions, we follow Cai et al. (2013) and restrict the contextual variable z_{it} (the log of total assets) to only vary across credit unions but not time by taking its time average.[25] This effectively reduces the number of iterative kernel estimations for each (r, τ) pair from NT to N, thus significantly speeding up the estimation as well as cross-validation.[26]

Economies of Diversification. We use the fitted model (3.15) to compute economies of diversification (ED) exhibited in the US retail credit union sector. Such economies of diversification are said to exist if the average non-interest cost, which credit unions incur, tends to decline with the increase in the number of outputs produced. As discussed earlier, the three credit union types observed in the data can be described as being "nested", implying a distinct diversification process whereby credit unions evolve from Type 1 to Type 2 and, further, from Type 2 to Type 3.

[25] One can potentially be concerned whether time-averaging is appropriate in our application because credit unions have been documented to exhibit a rather rapid growth in their average size during recent decades. However, the latter phenomenon is primarily driven by mergers and acquisitions widely taking place in the industry. Since our sample includes only continuously operating credit unions and *excludes* such mergers and acquisitions, time-averaging of the contextual variable does not seem to pose problems.

[26] The computational burden of estimating our model is not trivial because, unlike standard semi- or nonparametric conditional mean models, ours does not have a closed-form solution requiring a numerical search for the optimum at each kernel iteration.

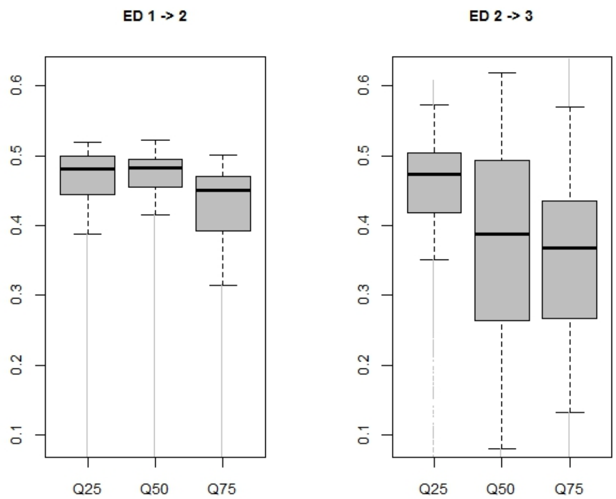

Figure 1. Economies of Diversification Estimates across Cost Quantiles

Consequently, we center our analysis around the estimates of $ED_{1\to2}$ and $ED_{2\to3}$ (see Section 2.1).

Since both $ED_{1\to2}$ and $ED_{2\to3}$ measures depend on the allocation [controlled by the choice of weights $(\varpi_1, \varpi_3, \varpi_4)$] of a diversified credit union's output quantities across more specialized production units, we employ a conservative approach to measuring economies of diversification by focusing on the *smallest* ED estimates in the data-defined observation-specific admissible regions defined by the diversified credit unions' output vectors. Essentially, for each credit-union-year, we perform a grid search of the lowest ED value across different combinations of (relevant) weights $(\varpi_1, \varpi_3, \varpi_4)$ yielding estimates that lie within the admissible region. The grid search is performed over the weights' permissible range between 0 and 1 at the 0.10 increments. The obtained smallest ED estimate is referred to as the measure of the credit union's "global" economies of diversification, i.e., GED. The reported $GED_{1\to2,\tau}$ and $GED_{2\to3,\tau}$ estimates are obtained using actual data for the already diversified Type 2 and Type 3 credit unions, respectively. This gives us the distribution of the fitted GED estimates for each quantile of the cost. We therefore caution the reader to not confuse quantiles of the cost distribution (τ), for which our model is estimated, with quantiles of the fitted GED distribution for each τ. For more details on the construction of the GED measure, please see Supplementary Appendix A.

Table 2 reports the point estimates of "global" economies of diversification across the 0.25, 0.50 and 0.75 quantiles of the (conditional) credit union cost distribution. The upper and lower panels report summaries of $GED_{1\to2,\tau}$ and $GED_{2\to3,\tau}$ estimates, respectively. We find that, on average, economies of diversification in the "Type 1 $\to$ Type 2" direction somewhat decrease in magnitude as one moves from the 0.25 to 0.75 quantile of the credit union cost distribution. The median estimates of $GED_{1\to2,\tau}$ are 0.48, 0.48 and 0.45 for the 0.25, 0.50 and 0.75 τ-quantiles of costs, respectively, pointing to a somewhat timid variation in the magnitude of economies of diversification across quantiles of the conditional cost distribution. However, in the "Type 2 $\to$ Type 3" direction, the decrease in economies of diversification with the quantile of costs is far more pronounced: the corresponding median estimates of $GED_{2\to3,\tau}$ are 0.47, 0.39 and 0.37 (see Table 2). These regularities are easier to recognize by looking at Figure 1 which depicts box-and-whiskers plots of the distributions of GED estimates for each τ. For a more conventional graphical representation

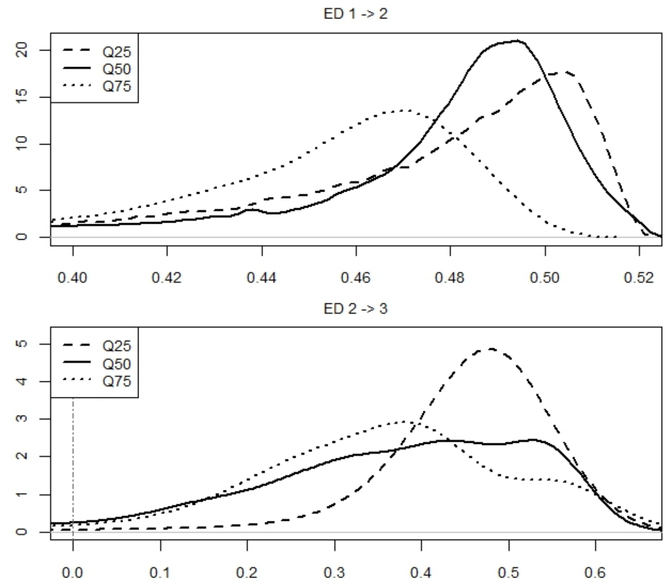

Figure 2. Kernel Densities of Economies of Diversification across Cost Quantiles

of the empirical distribution of the GED estimates, see Figure 2 which plots kernel densities of these point estimates. The distributions are negatively skewed regardless of the quantile of the cost for which the cost function is estimated, indicating the overwhelming presence of substantial economies of diversification across all τ-quantiles.

Overall, our estimates indicate economically and statistically significant economies of diversification present in the US retail credit union sector. For a credit union in the middle of the cost distribution ($\tau = 0.5$), the $ED_{1\to 2,0.5}$ estimates imply that consolidating the issuance of real estate loans (y_1) with consumer loans and investments (y_3, y_4) reduces credit union costs, on average, by astounding 45%. A similar integration of business loans (y_2) and the mix of real estate and consumer loans and investments (y_1, y_3, y_4) is estimated to reduce credit union costs by the median of 37–47% depending on the quantile of the cost. However, our discussion has so far been based solely on point estimates, where no sampling error associated with the estimation of the model has been taken into account. The right panel of Table 2 reports the breakdown of Type 2 and Type 3 credit unions into three categories: diseconomies of, invariance to and economies of diversification. This classification is based on the GED point estimate being statistically less/equal/greater than zero at the 95% significance level. For inference, we use bootstrap percentile intervals as described in Section 3.[27]

Having accounted for four sources of heterogeneity among retail credit unions in a pursuit of robust estimates of the credit union cost structure, we find *no* notable empirical evidence in support of diseconomies of diversification in the industry. On the contrary, as many as 87–91% and 27–48% of diversified Type 2 and Type 3 credit unions, respectively, enjoy both economically and statistically significant economies of diversification. The cost of most remaining credit unions is invariant to the scope of financial services (see Table 2). In general, the data indicate that Type 2 credit unions exhibit economies of diversification more often than Type 3 credit unions, which tend to be substantially larger in size. The latter suggests that credit unions may naturally exhaust their diversification-driven cost reduction potential (via reusing credit information about customers or diversifying risks across different assets) as the volume of their operations expands. For instance,

[27]We perform 399 bootstrap replications.

Table 3. Returns to Scale Estimates

Model	*Point Estimates* 1st Qu.	Median	Mean	3rd Qu.	*Categories, %* DRS	CRS	IRS
			Type 1 Credit Unions				
$\mathbb{Q}_{0.25}$	1.897	2.208	2.338	2.567	0.00	12.50	87.50
$\mathbb{Q}_{0.50}$	1.951	2.242	2.375	2.616	0.00	0.00	100.00
$\mathbb{Q}_{0.75}$	2.158	2.456	2.739	2.976	0.00	0.00	100.00
			Type 2 Credit Unions				
$\mathbb{Q}_{0.25}$	1.710	1.896	1.965	2.137	0.00	0.00	100.00
$\mathbb{Q}_{0.50}$	1.709	1.904	2.018	2.200	0.00	0.00	100.00
$\mathbb{Q}_{0.75}$	1.621	1.777	1.872	2.020	0.00	0.00	100.00
			Type 3 Credit Unions				
$\mathbb{Q}_{0.25}$	1.637	1.768	1.765	1.877	0.00	0.00	100.00
$\mathbb{Q}_{0.50}$	1.538	1.627	1.676	1.734	0.00	0.00	100.00
$\mathbb{Q}_{0.75}$	1.471	1.540	1.551	1.611	0.00	0.05	99.95

NOTE: Percentage points may not sum up to a hundred due to rounding.

larger credit unions may see their diversification-driven cost savings be, at least partly, offset by extra costs associated with having to adopt more complex management systems necessary to coordinate larger operations.

Economies of Scale. We next proceed to measuring the degree of scale economies (if any) exhibited by credit unions. Economies of scale are said to exist if the average non-interest cost, which credit unions incur, declines with the expansion of the scale of operations. The latter has recently become a subject of particular scholarly interest (e.g., Emmons and Schmid, 1999; Wilcox, 2005, 2006; Wheelock and Wilson, 2011; Malikov et al., 2015b, 2016).

The returns to scale are defined as the inverse of the sum of output elasticities of the cost and are computed as (for each credit union type r and quantile of the cost distribution τ):

$$RTS_{r,\tau} = \left[\sum_{m}^{M_r} \frac{\partial \log(C_{r,\tau})}{\partial \log(y_m)}\right]^{-1} ,$$

where the number of outputs M_r varies with the type r. Note that, in this paper, we estimate the (short-run) variable cost function which includes the cost of labor and financial capital only and is likely to be smaller than the (long-run) total operating cost. Since the short-run elasticities of the cost are at least as large as their long-run counterparts, the returns to scale estimates based on the (short-run) variable cost function are likely to be higher than those we would have obtained if estimating the long-run cost function.

Table 3 reports the summary of point estimates of the returns to scale for the 0.25th, median and 0.75th quantiles of the (conditional) cost distribution for each of the three credit union types. The right panel of the table also reports the groupings of credit unions based on the returns to scale categories: decreasing returns to scale (DRS), constant return to scale (CRS) and increasing returns to scale (IRS). We classify a credit union as exhibiting DRS/CRS/IRS if the point estimate of its returns to scale is found to be statistically less than/equal to/greater than one at the 95% significance level.

The empirical evidence overwhelmingly suggests the presence of IRS across credit unions of *all* types and quantiles of the cost distribution. We find that almost every single credit union of all

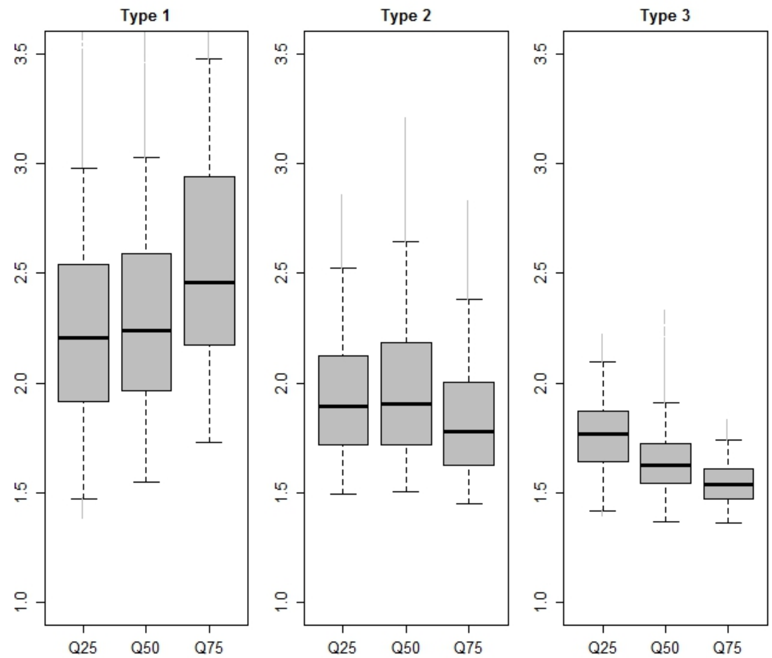

Figure 3. Returns to Scale Estimates across Cost Quantiles and Credit Union Types

types in our sample enjoys statistically significant scale economies (IRS) except for about 12.5% of small (the 0.25th quantile) Type 1 credit unions whose costs are found to be invariant to scale. For instance, a 1% scale expansion of the most diversified (Type 3) credit union is associated with an increase of its non-interest cost by a median of 0.57–0.64% only, depending on the quantile of the conditional cost distribution. Our finding of prevalent IRS is consistent with the results reported in Wheelock and Wilson (2011) and Malikov et al. (2015b), who however investigate credit union costs at the *conditional mean* only. One thus can interpret our results as providing the robust assurance of the findings previously reported in the literature.

Figure 3 graphs box-and-whiskers plots of the distributions of the returns to scale estimates across credit union types and the quantiles of the cost distribution. This figure conveniently permits us to analyze the relationship between returns to scale of a credit union and its scale, roughly proxied by its service menu type and the quantile of the cost distribution. First, we find that the estimates of returns to scale largely fall as credit unions evolve from Type 1 to Type 2 and, further, to Type 3. The latter is an expected finding given that a progressive movement across service menu types is associated with an increase in the average scale of operations (see Table 1 in Section 4). The same inverse relation holds across the quantiles of the cost distribution (which provides another rough measure of the credit union's scale within a given service menu type) for Type 2 and Type 3 credit unions. In the instance of Type 1 credit unions, however, the returns to scale increase with the quantiles of the credit union cost distribution. The latter may be because, given a relative small size of Type 1 credit unions to begin with, an increase in available resources as a credit union grows (captured by higher costs) enables it to adopt new information processing technologies (e.g., internet banking, automated teller machines, use of electronic money) that are initially expensive to install but, once installed, are substantial cost-savers. Given that the median credit union of Type 1 is as small as an entity with only 1.5 full-time equivalent employees, many of these credit unions are not financially capable of adopting the above-mentioned technologies until they grow in size. The impact of these financial constraints however naturally wears out as credit unions continue to grow which we observe in the instance of Type 2 and Type 3 credit unions.

Lastly, we also investigate the consequences of overlooking the heterogeneity across credit unions of different sizes as well as the heterogeneity across different (conditional) quantiles of the cost distribution by estimating two alternative, more restrictive, specifications of the credit union cost relationship. To conserve space, the discussion of these models is relegated to Supplementary Appendix B.

6 Conclusion

This paper explores economies of diversification as a plausible cost-side trigger to the recent growth of the retail credit union industry in the U.S. We focus on the fact that credit unions differ among themselves in the scope/range of financial services they offer to their members. Larger credit unions tend to provide a more diversified financial service menu to their membership pool than credit unions of a smaller size. Thus, the incentive to grow in size may be fueled not only by present scale economies (as has been argued in the literature) but also by economies of diversification, which are said to exist if the average cost that credit unions incur declines with the increase in the scope of services provided. If, by growing its membership, credit unions are able to secure the demand for a more diversified scope of services, they would naturally have the incentive to grow in order to also capitalize on economies of diversification.

We focus on the product diversification within a group of interest-earning activities (namely, different kinds of loans), which we examine in the context of a model of credit union costs. In a pursuit of robust estimates of economies of diversification for credit unions, we estimate a flexible semiparametric smooth coefficient quantile panel data model with correlated effects that is capable of accommodating four kinds of heterogeneity among credit unions: *(i)* heterogeneity across credit unions offering different financial service menus, *(ii)* heterogeneity across entities of different size, *(iii)* heterogeneity across the (conditional) quantiles of the credit union cost distribution, and *(iv)* unobserved heterogeneity correlated with covariates in the estimated equation.

Our results provide evidence in favor of economically and statistically significant economies of diversification present among retail credit unions continuously operating in the US during the 2001–2006 period. We find that, for a credit union in the middle of the cost distribution, the consolidation of the issuance of real estate loans with consumer loans and investments reduces credit union costs, on average, by astounding 45%. A similar integration of business loans and the mix of real estate and consumer loans and investments is estimated to reduce credit union costs by the median of 37–47% depending on the quantile of the cost. Overall, we find no notable evidence of diseconomies of diversification in the industry. On the contrary, as many as 27–91% (depending on the type and the cost quantile) of diversified credit unions enjoy substantial economies of diversification. The cost of most remaining credit unions is invariant to the scope of financial services.

On a final note, when measuring economies of diversification for credit unions, we make use of observable diversification paths in the industry which uniquely enables us to rely on fewer counterfactuals necessary to estimate potential cost savings attributed to the expanded scope. The latter makes our estimates of economies of diversification "less counter" to *factum* thereby significantly reducing the excessive extrapolation problem associated with the measurement of scope economies in general. In fact, we seek to further mitigate excessive extrapolation by limiting our analysis to the data-determined "admissible region" within which we only focus on the most conservative estimates of economies of diversification. While our methodology provides a significant *improvement* over the more traditional and widely practiced approaches to measuring scope economies in the literature, we nonetheless acknowledge that it is not completely perfect because it still relies on counterfactuals. However, the reader may take solace in the fact that an admittedly more desirable counterfactual-free measurement of economies of diversification is generally infeasible for the U.S. credit union sector due to the non-existence of fully specialized firms in the industry which makes the direct estimation of product-specific cost functions impossible. Therefore, our improved methodology provides a good compromise.

References

Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight: Evidence from quantile estimation on panel data. *Journal of Business and Economic Statistics*, 26:379–397.

Barrodale, I. and Roberts, F. (1974). Solution of an overdetermined system of equations in the L1 norm. *Communications of the ACM*, 17:319–320.

Bauer, K. J., Miles, L. L., and Nishikawa, T. (2009). The effect of mergers on credit union performance. *Journal of Banking & Finance*, 33(12):2267–2274.

Baumol, W., Panzar, J., and Willig, R. (1982). *Contestable Markets and the Theory of Industry Structure*. Harcourt, Brace & Jovanovich, San Diego.

Berger, A. N., Hanweck, G. A., and Humphrey, D. B. (1987). Competitive viability in banking: Scale, scope, and product mix economies. *Journal of Monetary Economics*, 20:501–520.

Cai, Z., Chen, L., and Fang, Y. (2013). Quantile panel data models with partially varying coefficients: Estimating the growth effect of FDI. Working Paper, University of Kansas.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902.

Cai, Z., Gu, J., and Li, Q. (2009). Some recent developments on nonparametric econometrics. *Advances in Econometrics*, 25:495–549.

Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167:413–425.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimation for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103:1595–1608.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47:225–238.

Credit Union National Association (CUNA) (2013). *Credit Union Report Year-End 2013*.

Emmons, W. R. and Schmid, F. A. (1999). Credit unions and the common bond. *Federal Reserve Bank of St. Louis Review*, 81(September/October).

Evans, D. S. and Heckman, J. J. (1984). A test for subadditivity of the cost function with an application to the Bell System. *American Economic Review*, 74(4):615–623.

Feng, G. and Serletis, A. (2010). Efficiency, technical change, and returns to scale in large US banks: Panel data evidence from an output distance function satisfying theoretical regularity. *Journal of Banking & Finance*, 34:127–138.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, 98:995–999.

Frame, S. W. and Coelli, T. J. (2001). US financial services consolidation: The case of corporate credit unions. *Review of Industrial Organization*, 18(2):229–241.

Frame, S. W., Karels, G. 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(1):35–47.

Fried, H. O., Lovell, C., and Yaisawarng, S. (1999). The impact of mergers on credit union service provision. *Journal of Banking & Finance*, 23(2):367–386.

Fuss, M. A. and Waverman, L. (2002). Econometric cost functions. In Cave, M. E., Majumdar, S. K., and Vogelsang, I., editors, *Handbook of Telecommunications Economics*, volume 1. Elsevier.

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

Goddard, J. A., McKillop, D. G., and Wilson, J. O. (2008). The diversification and financial performance of US credit unions. *Journal of Banking & Finance*, 32(9):1836–1849.

Goddard, J. A. and Wilson, J. O. (2005). US credit unions: An empirical investigation of size, age and growth. *Annals of Public & Cooperative Economics*, 76:375–406.

Grosskopf, S., Hayes, K., and Yaisawarng, S. (1992). Measuring economies of diversification: A frontier approach. *Journal of Business and Economic Statistics*, 10(4):453–459.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

Horowitz, J. L. (2001). The bootstrap. In Heckman, J. J. and Leamer, E., editors, *Handbook of Econometrics*, chapter 52, pages 3159–3228. Elsevier Science B.V., 5 edition.

Hughes, J. P. and Mester, L. J. (1993). A quality and risk-adjusted cost function for banks: Evidence on the "too-big-to-fail" doctrine. *Journal of Productivity Analysis*, 4:293–315.

Hughes, J. P. and Mester, L. J. (2015). Measuring the performance of banks: Theory, practice, evidence, and some policy implications. In Berger, A., Molyneux, P., and Wilson, J., editors, *Oxford Handbook of Banking*, pages 247–270. Oxford University Press, Oxford, 2 edition.

Ichimura, H. and Todd, P. E. (2007). Implementing nonparametric and semiparametric estimators. In Heckman, J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6, pages 5369–5468. Elsevier.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *Annals of Statistics*, 35:92–108.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91:74–89.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15:143–156.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94:1296–1310.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, 19:1–31.

Lee, T.-H. and Ullah, A. (2003). Nonparametric bootstrap specification testing in econometric models. In Giles, D. E., editor, *Computer-Aided Econometrics*, pages 451–477. CRC Press, NY.

Li, Q., Huang, C. J., Li, D., and Fu, T.-T. (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, 20(3):412–422.

Li, Q. and Racine, J. C. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton.

Malikov, E. and Kumbhakar, S. C. (2014). A generalized panel data switching regression model. *Economics Letters*, 124(3):353–357.

Malikov, E., Kumbhakar, S. C., and Sun, Y. (2016). Varying coefficient panel data model in the presence of endogenous selectivity and fixed effects. *Journal of Econometrics*, 190(2):233–251.

Malikov, E., Restrepo-Tobón, D. A., and Kumbhakar, S. C. (2015a). Estimation of banking technology under credit uncertainty. *Empirical Economics*, 49(1):185–211.

Malikov, E., Restrepo-Tobón, D. A., and Kumbhakar, S. C. (2015b). Heterogeneous credit union production technologies with endogenous switching and correlated effects. *Econometric Reviews*. forthcoming.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.

National Credit Union Administration (NCUA) (2013). *2013 Annual Report*.

Restrepo-Tobón, D. and Kumbhakar, S. C. (2015). Nonparametric estimation of returns to scale using input distance functions: An application to large U.S. banks. *Empirical Economics*, 48(1):143–168.

Rime, B. and Stiroh, K. J. (2003). The performance of universal banks: Evidence from Switzerland. *Journal of Banking & Finance*, 27:2121–2150.

Smith, D. J. (1984). A theoretic framework for the analysis of credit union decision making. *Journal of Finance*, 39(4):1155–1168.

Ullah, A. (1985). Specification analysis of econometric models. *Journal of Quantitative Economics*, 1:187–209.

Walter, J. R. (2006). Not your father's credit union. *Federal Reserve Bank of Richmond Economic Quarterly*, 92(4):353–377.

Wheelock, D. C. and Wilson, P. W. (2011). Are credit unions too small? *Review of Economics and Statistics*, 93(4):1343–1359.

Wheelock, D. C. and Wilson, P. W. (2013). The evolution of cost-productivity and efficiency among US credit unions. *Journal of Banking & Finance*, 37:75–88.

Wilcox, J. A. (2005). Economies of scale and continuing consolidation of credit unions. *FRBSF Economic Letter*, (Nov 4).

Wilcox, J. A. (2006). Performance divergence of large and small credit unions. *FRBSF Economic Letter*, (Aug 4).