NONPARAMETRIC ESTIMATION OF RETURNS TO SCALE USING INPUT DISTANCE FUNCTIONS: AN APPLICATION TO LARGE U.S. BANKS

Diego Restrepo-Tobón*

Subal C. Kumbhakar†

This version: June 14, 2013

Abstract

In this paper we derive new measures of returns to scale based on input distance functions (IDF) and estimate them using nonparametric regression methods. In contrast to cost functions, IDF dispenses with input prices which are usually unavailable or measured imprecisely. In addition, we can appropriately account for equity and physical capital in the IDF. These items are traditionally excluded from the analysis (especially in a cost function approach) or treated as quasi fixed inputs because their prices are not readily available. Using data for bank holding companies and large commercial banks in the U.S. from 2000 to 2010, we find that although some of these institutions enjoy increasing returns to scale, scale economies are economically small. Thus, concerns about potential cost increases from breaking up large banking organizations seem exaggerated, especially from the scale economies viewpoint.

JEL Classification: D24, G21, L13

Keywords: Nonparametric regressions, returns to scale, distance functions.

*Corresponding author, Department of Economics, State University of New York, Binghamton, USA and EAFIT University, Medellín, Colombia. *e-mail*: drestre1@binghamton.edu. Restrepo acknowledges financial support from the Colombian Fulbright Commission, the Colombian Administrative Department of Science, Technology and Innovation (Colciencias), and EAFIT University.

†Department of Economics, State University of New York, Binghamton, USA.

1 Introduction

More than two years after The Dodd-Frank Wall Street Reform and Consumer Protection Act was signed into law, regulators, policymakers, and academics in the U.S. are still pondering the possibility and desirability of limiting the size of large banking organizations. Recent speeches by top Federal Reserve Bank officials in the U.S. and of the Bank of England in the U.K. revived the unsettled debate on "too big to fail" (TBTF) and the feasibility of limiting the scale and scope of bank activities, calling for further research on banking industry structure in general and on economies of scale and scope in particular.[1]

If large banking organizations enjoy economies of scale, limiting or shrinking their size may pose substantial losses to the economy. However, recent research focusing on the existence of economies of scale and on the cost of shrinking or capping the size of banks is still inconclusive. Wheelock and Wilson (2011, 2012) present evidence indicating that all U.S. commercial banks, bank holding companies (BHC), and credit unions operate under increasing returns to scale (RTS). Likewise, Hughes and Mester (2011) show evidence of increasing RTS for BHC operating in 2007. For large U.S. commercial banks, those with assets in excess of $1 billion, Feng and Serletis (2010) present evidence of slightly increasing RTS, Feng and Zhang (2012) find decreasing RTS, and Restrepo-Tobón, Kumbhakar and Sun (2012) find increasing RTS for about 73% of them and constant or slightly decreasing RTS for the rest.

Among these studies, only Wheelock and Wilson (2012) and Hughes and Mester (2011) estimate the potential cost of breaking up large BHC. Using a back-of-the-envelope calculation, Wheelock and Wilson (2012) estimate that the additional cost of breaking up the four largest U.S. BHC operating in 2010 into firms with no more than $1 trillion of assets would be around $79.1 billion per year.[2] This value exceeds their combined net income in each year from 2003 to 2006. Hughes and Mester (2011) conduct a similar calculation to estimate the additional cost of breaking up the

[1] See Tarullo (2012a,c,b), Haldane (2012), Rosenblum (2011), and the ensuing media coverage and market participants' analyses (e.g. Johnson 2012, Wack 2012, Wallison 2012, and Harrison 2012).

[2] The four BHC are: Bank of America Corporation with $2.268 trillion; J.P. Morgan Chase with $2.118 trillion; Citigroup with $1.914 trillion; and Wells Fargo with $1.258 trillion.

17 BHC with assets exceeding \$100 billion into firms with assets of no more than \$100 billion, while keeping their output mix unchanged. They find that the additional cost would be around \$990 billion per year. Allowing BHC to change their output mix to a value in line with their new hypothetical smaller size, however, yields cost savings between \$21 and \$147 billion per year, depending on the method used.

Boyd and Heitz (2012) estimates that the potential benefits to the society from economies of scale of big financial institutions are unlikely to ever exceed the potential costs due to their contribution to systemic risk. They find that the cost to the economy as a whole due to increased systemic risk is of an order of magnitude larger than the potential benefits due to any economies of scale.

To contribute to this literature and enhance the policy debate, we derive new measures of returns to scale (RTS) based on input distance functions (IDF) and estimate them using nonparametric regression methods for the largest U.S. banking organizations. Compared with a cost function-based approach, an IDF gives a primal representation of the underlying technology and its estimation requires no information on total costs or input prices. On the other hand, nonparametric methods circumvent the potential misspecification problem inherent in parametric models and dispense with assumptions about the true unknown functional form of the underlying production technology.

Conventional economies of scale studies are potentially misleading since they rely on estimated cost functions that may identify TBTF lower funding costs as evidence of scale economies. RTS estimates from an IDF however, are partially shielded from this problem since in their estimation one uses information that is less likely to be driven by TBTF funding advantages embodied in firms' costs and the input prices they face.

Using an IDF allows us to appropriately account for equity and physical capital in the estimation. Many studies based on cost functions exclude them from the analysis or treat them as quasi-fixed inputs given that their associated prices are not readily available (e.g., Wheelock and Wilson 2012 and Berger and Mester 2003). Output distance functions (ODF) also shares these advantages with the IDF approach (see Feng and Serletis, 2010 and Feng and Zhang, 2012). However, estimating an ODF leads to biased and inconsistent estimates when inputs are endogenous; which is the case when the assumption of cost minimization is appropriate. In contrast, the estimation of the IDF does not suffer from this problem, Das and Kumbhakar 2012.

The closest papers to our work are Wheelock and Wilson (2011, 2012) (W&W). However, our papers differs from theirs along several dimensions. First, while W&W use a cost function, we use an IDF. As mentioned before, using an IDF instead of a cost function avoids making the assumptions that physical and financial capital are quasi-fixed factors of production; which seems not to be the case for the long periods traditionally studied in the literature.

Second, appealing to economic intuition, W&W define two different measures of scale economies based on the cost function—ray-scale economies (RSE) and expansion-path scale economies (EPSE). They show that these measures can be used to investigate the nature of scale economies. We generalize the RSE measure by deriving it from the basic properties of the underlying bank technology for cost, output distance, and input distance functions. Further, we prove that the EPSE measure, as defined by W&W, is just the ratio between two RSE at two different points and that, contrary to W&W's claims, EPSE is not a better measure of RTS; in some cases it may not accurately indicate the nature and magnitude of RTS. In addition, we show that it is unnecessary to use the slope of the RSE measure to investigate the nature of RTS, as W&W advocate, since the appropriate measure is the RSE itself.

Third, we use a more recent dataset which is more relevant for current policy debates and focus only on banks with assets above $500 million—the existence of scale economies at smaller banking organizations is not an issue in the literature. In particular, we use data for the period after deregulation of the banking industry to avoid that differences in regulations across banks contaminate the estimation of RTS. This strategy allows us to have a smaller sample and avoid the use of principal components analysis to reduce the dimensionality of the data as is done by W&W.

Finally, W&W avoid using gradient-based measure of RTS based on nonparametric functions arguing that gradient estimates may be noisy. We compute such measures and find that the conclusion derived using the RSE are similar to those derived from gradient-based measures of RTS. Thus, W&W's argument in this respect seems to have little empirical support.

We use annual data for U.S. BHC and the biggest U.S. commercial banks from 2000 to 2010. Contrary to Wheelock and Wilson (2012), but in line with conventional wisdom, we find that not all BHC and commercial banks enjoy increasing returns to scale (IRTS). In addition, economies of scale for those banking organizations operating under IRTS are small. Our RTS estimates are generally close to unity but are estimated very precisely so that we can distinguish between increasing, decreasing, and constant returns to scale with enough accuracy. Thus, despite the existence of IRTS at some of the biggest banking organizations, the cost of breaking up some of these institutions into smaller and more manageable organizations may pose few costs to the economy.

In the next section we derive our measure of RTS and compared them with those presented in related studies. Section 3 presents our model of bank production and describes the data. Section 4 discusses the econometric estimation of the model and Section 5 presents the main empirical results. We conclude by highlighting the policy implications in Section 6.

2 Methodology

In this section we derive our RTS measures based on basic properties of the underlying production technology for cost, input distance, and output distance functions; and highlight the main differences with related measures used in the literature.

2.1 Modelling Banks' Technology

We assume that banks have a production technology $\mathbb{T} : \mathbb{R}^N_+ \times \mathbb{R}^M_+ \to T = \{(x,y) : x \text{ can produce } y\}$ that transforms input vectors $x = (x_1,\ldots,x_N) \in \mathbb{R}^N_+$ into output vectors $y = (y_1,\ldots,y_M) \in \mathbb{R}^M_+$. The output correspondence $\mathbb{P} : \mathbb{R}^N_+ \to P(x) = \{y : (x,y) \in T\}$ maps the input vectors (x) into output sets $P(x)$ which contains all producible output vectors from $x \in \mathbb{R}^N_+$. The input correspondence $\mathbb{L} : \mathbb{R}^M_+ \to L(y) = \{x : (x,y) \in T\}$ maps output vectors y into input sets $L(y)$ which contains all input vectors that can produce $y \in \mathbb{R}^M_+$. By definition, $y \in P(x) \iff (x,y) \in T \iff x \in L(y)$. Thus, $\mathbb{T}$, $\mathbb{P}$, and $\mathbb{L}$ are equivalent representations of the production technology.

The cost, output, and input distance functions are defined, respectively, as follows:[3]

$$C(y,w) = \min_x \{px : x \in L(y)\}, y \in \text{Dom } L(y), w > 0. \tag{1}$$

$$D_o(x,y) = \inf_\theta \{\theta > 0 : (y/\theta) \in P(x)\} \text{ for all } x \in \mathbb{R}^N_+. \tag{2}$$

$$D_I(y,x) = \sup_\lambda \{\lambda > 0 : (x/\lambda) \in L(y)\} \text{ for all } y \in \mathbb{R}^M_+. \tag{3}$$

Under minimal standard assumptions, the cost function is a dual representation of the underlying technology. Moreover, the output and input distance functions are themselves representations of the underlying production technology. The properties of the cost function are well known and those of the ODF and IDF are detailed in Färe (1988). For our purposes, we highlight some important properties of the IDF. The IDF is nondecreasing and linear homogeneous in x; and nonincreasing and, in general, not homogeneous in y. We exploit these properties to derive our measures of RTS based on the IDF.

The literature shows that the production technology exhibits nonincreasing returns to scale (NIRTS) if and only if for all $\gamma \geq 1$, the following (subhomogeneity) conditions hold:

I. $P(\gamma x) \subseteq \gamma P(x)$
II. $L(\gamma y) \subseteq \gamma L(y)$

For $\lambda \in (0,1]$ the signs in I and II have to be reversed. If the opposite (superhomogeneity) conditions hold, the technology exhibits nondecreasing returns to scale (NDRTS). In addition, the technology exhibits constant returns to scale (CRTS) if the signs in I and II are changed for equality signs (homogeneity conditions).

Under NIRTS and for $\gamma \geq 1$, conditions I and II imply the following relations for the cost,

[3] See Färe and Primont (1995).

output, and input distance functions:

$$\gamma C(y,w) \leq C(\gamma y,w) \tag{4}$$

$$(1/\gamma)D_o(x,y) \leq D_o(\gamma x,y) \tag{5}$$

$$(1/\gamma)D_I(y,x) \geq D_I(\gamma y,x) \tag{6}$$

For $\gamma \in (0,1)$ the signs in (4) - (6) have to be reversed. The conditions for NDRTS are obtained by reversing the inequality signs. Likewise, the conditions for CRTS are obtained by substituting for equality signs.

We prove (6) for $\gamma \geq 1$. From (3)

$$\begin{aligned} D_I(\gamma y,x) =& \sup_{\lambda}\{\lambda > 0 : \frac{x}{\lambda} \in L(\gamma y)\} \leq \sup_{\lambda}\{\lambda > 0 : \frac{x}{\lambda} \in \gamma L(y)\} = \sup_{\lambda}\{\lambda > 0 : \frac{x}{\lambda\gamma} \in L(y)\} \\ =& \frac{1}{\gamma}\sup_{\lambda}\{\lambda\gamma > 0 : \frac{x}{\lambda\gamma} \in L(y)\} = \frac{1}{\gamma}\sup_{\phi}\{\phi > 0 : \frac{x}{\phi} \in L(y)\} = \frac{1}{\gamma}D_I(y,x) \end{aligned} \tag{7}$$

where the inequality follows from condition II. Conditions (4) and (5) are proved analogously.

The economic intuition underlying the condition (4) is easy to understand. Under NIRTS, if outputs were increased by a factor $\gamma > 1$, total costs would increase by a factor greater than γ. Otherwise, with fixed output prices, it would be profitable to expand production since revenues would increase by a factor γ but total costs would increase by a factor less than γ. The economic intuition of conditions (5) and (6), on the other hand, is less evident and requires familiarity with the concepts of output and input distance functions—we return to this issue in the next subsection. For this reason, only some versions of (4) has been used in the literature to measure the nature of scale economies using discrete changes in outputs and inputs.

2.2 Measuring Returns to Scale (RTS)

Eliminating the inequalities in conditions (4)–(6), we can define the following measures of RTS based on the cost and the output and input distance functions as follows:

RTS based on the cost function:

$$S^{cost}(\gamma|y,w) = \frac{C(\gamma y,w)}{\gamma C(y,w)} \tag{8}$$

For $\gamma > 1$, the technology exhibits NIRTS, CRTS, or NDRTS as $S^{cost}(\gamma|y,w) \gtreqless 1$, respectively. For $\gamma \in (0,1]$, the signs need to be reversed.

RTS based on the ODF:

$$S^{ODF}(\gamma|y,x) = \frac{D_o(\gamma x,y)}{(1/\gamma)D_o(x,y)} \tag{9}$$

For $\gamma > 1$, the technology exhibits NIRTS, CRTS, or NDRTS as $S^{ODF}(\gamma|y,w) \gtreqless 1$, respectively. For $\gamma \in (0,1]$, the signs need to be reversed.

RTS based on the IDF:

$$S^{IDF}(\gamma|y,x) = \frac{D_I(\gamma y,x)}{(1/\gamma)D_I(y,x)} \tag{10}$$

For $\gamma > 1$, the technology exhibits NIRTS, CRTS, or NDRTS, as $S^{IDF}(\gamma|y,x) \lesseqgtr 1$, respectively. For $\gamma \in (0,1]$, the signs need to be reversed.

Since the IDF is linear homogeneous in x, (10) can be rewritten as:

$$S^{IDF}(\gamma|y,x) = \frac{D_I(\gamma y,\gamma x)}{D_I(y,x)} \tag{11}$$

which makes it clear that $S^{IDF}(\gamma|y,x)$ measures how the distance function changes when all outputs and inputs are scaled up or down by a factor γ. If the IDF does not change, it means that the technology exhibits CRTS: an equiproportional increase of all inputs leads to an equal equiproportional increase in all outputs. If the technology exhibits NDRTS, then scaling up (down) all outputs and inputs leads to an increase (decrease) of the IDF. Likewise, if the technology exhibits NIRTS, then scaling up (down) all outputs and inputs leads to a decrease (increase) of the IDF.

We can use (8)–(10) to quantify the magnitude of RTS. Using the cost based RTS measure in (8), an increase in output quantities by a factor $\gamma > 1$ leads to an increase in cost by a factor $S^{cost}(\gamma|y,w) \times \gamma$. Conversely, a decrease in output quantities by a factor γ^{-1} leads to a decrease in cost by a factor $(S^{cost}(\gamma|y,w)\times\gamma)^{-1}$. Likewise, using the IDF based measure of RTS in (10), an increase in output quantities by a factor $\gamma > 1$ requires an increase in input quantities by a factor $\gamma/S^{IDF}(\gamma|y,x)$. Conversely, a decrease in output quantities by a factor γ^{-1} requires a decrease in input quantities by a factor $S^{IDF}(\gamma|y,x)/\gamma$. An analogous interpretation applies for (9).

As we later show, the relation in (12) allows us to interpret IDF-based measures of RTS in terms of cost and *vice versa*. However, the IDF based measure is more general since it does not require the cost minimization assumption upon which (8) is defined.

Using the especial two-inputs one-output case, figures 1 and 2 illustrate how our proposed RTS measure based on the IDF, (10), can identify the nature and quantify the magnitude of scale economies. Figure 1 illustrates the concept of the IDF. The IDF, $D_I(y,x)$, is given by the ratio between β/α and represents the maximum factor by which one needs to divide an arbitrary vector of inputs $x=(x_1,x_2)\in L(y)$ along the ray through x, so that the resulting vector x^* still belongs to $L(y)$. By definition of the IDF, $x/D_I(y,x)$ is contained in $L(y)$, but no point southwest of it is in $L(y)$.

In Figure 2, we draw several isoquants for different output levels. For simplicity, we consider only RTS along the ray $\mathscr{R}_0$ for which $x_1=x_2$, but any other ray can be considered. It is clear that up to $y=3$, the technology exhibits increasing RTS; from $y=3$ to $y=4.5$ it exhibits constant RTS; and after $y=4.5$ it exhibits decreasing RTS. Now, for any arbitrary $x\in L(y)$, we can compute our scale economy measure in (10) along the ray $\mathscr{R}_0$. Note that along $\mathscr{R}_0$, β and α are easily computed and so is $D_I(y,x)$.

Consider $x=(8,8)$. Along $\mathscr{R}_0$, the IDF associated with $y=1$ is $D_I(y=1,x)=8$. Likewise, $D_I(y=2,x)=16/3$. Thus, in this case, $S^{IDF}(\gamma=2|y=1,x)=(16/3)/(1/2\times 8)=4/3$, indicating IRTS as expected. Thus, doubling y from $y=1$ to $y=2$ requires only an increase of 50% in all inputs. Similarly, $D_I(y=3,x)=4$. Thus, considering increasing y from $y=2$ to $y=3$ yields $S^{IDF}(\gamma=3/2|y=2,x)=(4)/(3/2\times 4/3)=2$, indicating IRTS since this only requires an increase of 1/3 in inputs. For $y=3.75$, $D_I(y=3.75,x)=16/5$. Thus, increasing y from $y=3$ to $y=3.75$ yields $S^{IDF}(\gamma=1.25|y=3,x)=(16/5)/(4/5\times 4)=1$, indicating CRTS. The same results follows

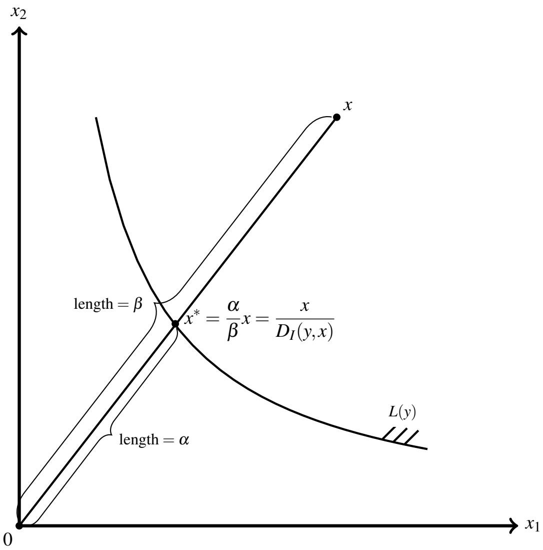

Figure 1: Input Distance Function

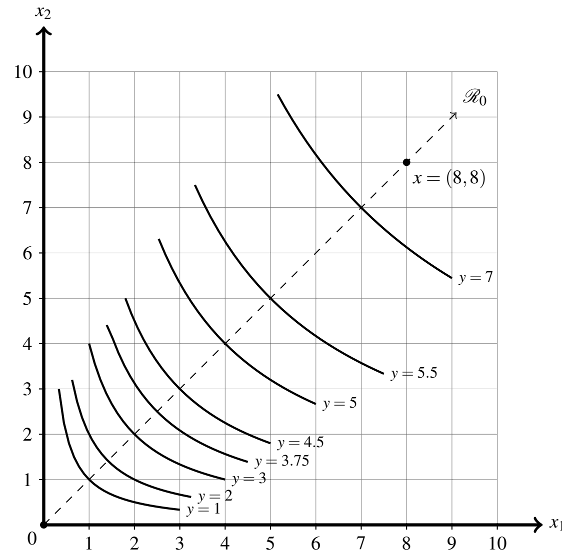

Figure 2: Returns to Scale

for $S^{IDF}(\gamma = 1.2|y = 3.75, x) = 1$. For $y = 5$, $D_I(y = 5, x) = 2$. Thus, $S^{IDF}(\gamma = 10/9|y = 4.5, x) = (2)/(9/10 \times 8/3) = 5/6 < 1$, indicating DRTS. Then, increasing y from $y = 4.5$ to $y = 5$ requires an increase in inputs of $\gamma/S^{IDF}(\gamma = 10/9|y = 4.5, x) = (10/9)/(5/6) = 4/3$ which is greater than the increase in output of 10/9. An analogous reasoning yields DRTS around $y = 5.5$ and $y = 7$.

Figure 2 also illustrates how our measure differs from the cost-based measured in (8). The cost-based measure assumes that isocost lines are tangent to the isoquant curves at each level of output y. Thus, it assumes that the firm is cost minimizing at the observed output level y and at all the hypothesized scaled output levels γy. In addition, one needs input price and total cost information to determine the nature and magnitude of RTS using the cost-based measure. In contrast, using the IDF-based measure one can dispense with the assumption of cost minimization at each level of output. In particular, the IDF-based measure is robust to the presence of technical inefficiency since it is well define even when $D_I(y,x) > 1$.

To our knowledge, neither the measure of RTS we propose based on the IDF nor the preceding analysis have appeared previously in the literature. In general, our measure of RTS based on the IDF can be used to compute RTS along any ray from the origin through any input vector. In particular, note that no cost minimizing assumption is involved in the analysis. Thus, measuring scale economies using the IDF is more general than using a cost function. Moreover, the technology may exhibit varying degrees of RTS along different rays and our measure based on the IDF will still be able to identify them. Another advantage of our approach is that we can measure RTS when outputs are scaled up or down by discrete factors and not at infinitesimal changes as traditional elasticity-based measures of RTS do.

Under the assumption of cost minimization the cost and IDF are related through $wx = C(y,w)\,D_I(y,x)$ (see Färe, 1988, p. 152). Thus, the following equality holds[4]

$$S^{cost}(\gamma|y,w) = \frac{1}{S^{IDF}(\gamma|y,x)} \tag{12}$$

Wheelock and Wilson (2012) refer to $S^{cost}(\gamma|y,w)$ in (8) as *ray-scale economies*. However, instead of using (8) directly, W&W based their analysis on how (8) changes when γ changes. This procedure is unnecessary since (8) itself is the appropriate measure of RTS in this context. In addition, W&W define an alternative measure of scale economies, viz., *expansion-path scale economies* (EPSE). We show that this latter measure is nothing new; in fact it is a ratio of two *ray-scale economies* defined in (8). The EPSE in Wheelock and Wilson (2011, 2012) is defined as

$$S^{W\&W}(\theta|y,w) = \frac{C(\theta(1-\gamma)y,w)}{\theta C((1-\gamma)y,w)} \tag{13}$$

where $\theta = (1+\gamma)/(1-\gamma)$ is 1.105 because they set $\gamma = 0.05$. Thus,

$$\begin{aligned} S^{W\&W}(\theta|y,w) &= \frac{C(\theta(1-\gamma)y,w)}{\theta C((1-\gamma)y,w)} = \frac{C((1+\gamma)y,w)}{\frac{1+\gamma}{1-\gamma}C((1-\gamma)y,w)} = \frac{C((1+\gamma)y,w)}{(1+\gamma)C(y,w)} \times \frac{(1-\gamma)C(y,w)}{C((1-\gamma)y,w)} \\ &= S^{cost}((1+\gamma)|y,w) \times \frac{1}{S^{cost}((1-\gamma)|y,w)} \end{aligned} \tag{14}$$

Thus, the EPSE measure of W&W is nothing but the ratio of two *ray-scale economies*. According

[4]If RTS is measured based on elasticity formula, the relationship in (12) can be obtained directly from the duality between cost and transformation functions (see Caves, Christensen and Swanson, 1981). We come to this in Section 5.2.

to W&W, it gives an indication of RTS for a particular bank moving along the path from the origin through its observed output vector y, starting at $(1-\gamma)$ of y and continuing to $(1+\gamma)$ of y. However, the correct interpretation of (13) is that it measures RTS when outputs are scaled up by $\theta > 1$, starting at $(1-\gamma)$ of the observed output level. So, it measures RTS when outputs are scaled up by θ from $(1-\gamma)\times y$ to $(1+\gamma)\times y$ which is different than measuring RTS at or around y. That is, it does not measure RTS at or around the observed output level. The reason is that if from $(1-\gamma)\times y$ to $(1+\gamma)\times y$ the technology uniformly exhibits NIRTS, NDRTS, or CRTS this measure is well defined. However, it is possible that when y is scaled up by $(1+\gamma)$ the technology might exhibit NDRTS (NIRTS or CRTS) but when y is scaled down by $(1-\gamma)$ it might exhibit NIRTS (NDRTS). In this case, the EPSE measure is not well defined to determine the nature of RTS since it does not measure RTS around the observed output level y. For instance, suppose we want to measure RTS around an hypothetical observed output level $y = 3$ with $\gamma = 1/3$. In this case, $\theta = 2$. Measuring RTS using the EPSE implies going from $y = 2$ to $y = \theta \times 2 = 4$. As Figure 2 shows, in this case we will conclude that the technology exhibits NDRTS. However, the main interest is about what happens around $y = 3$, the observed output level. From Figure 2, it is clear that the technology exhibits NDRTS above $y = 3$ and CRTS below $y = 3$. So, the EPSE measure gives an indication of the nature and magnitude of RTS going from $y = 2$ to $y = 4$ but not around $y = 3$. That is, it gives an indication of RTS between two hypothetical points above and below the observed output level $y = 3$ but not at or around it.

Our RTS measures are based on global properties of the underlying technology. Thus, they can be readily extended to other available methods for estimating distance functions. For instance, directional distance function along a particular direction vector can also be used. Our estimated measures are local in the sense that they measure RTS at the chosen value of γ. Choosing different values of γ allows us to make inferences about the robustness of our procedure. If for a given observation, $\hat{S}^{IDF}(\gamma|x_0)$ changes wildly as γ changes, it may indicate that the nonparametric estimate of the IDF is unstable and unable to estimate RTS precisely. We check the stability of our results in the empirical section.

Under the assumption of cost minimization, the *ray-scale economies* measure of RTS based on IDF is equivalent to measures based on cost functions. This allows us to infer returns to scale and investigate how changes in the level of outputs will change total costs for a given banking organization even if input prices are unavailable. Further, our measure is also valid if the cost minimization assumption fails to hold in the data. Therefore, our results are more general than those presented in the recent literature.

In the next section, we present our model of bank production, describe our strategy to estimate (10) nonparametrically, and show how to make inferences about scale economies.

3 A Model of Bank Production

Consistent with the widely used intermediation approach of Sealey and Lindley (1977), we assume that banks' balance-sheets capture the essential structure of banks' core business: (i) liabilities, together with physical and financial (equity) capital, and labor, are inputs into the bank production process and (ii) assets, other than physical assets, are outputs. In addition, we define off-balance sheet activities as an additional output. Liabilities include core deposits and purchased funds while assets include loans and trading securities. Off-balance sheet activities include all revenue sources other than lending and securities trading. Therefore, banks use labor, physical capital, financial capital, and liabilities to produce loans, invest in financial assets, and facilitate other financial services.

To keep our results comparable with the literature and, in particular, to those of Wheelock and Wilson (2012), we define the following output variables: consumer loans (y_1), real estate loans (y_2), business loans (y_3), securities (y_4), and off-balance sheet activities (y_5). The input variables are: purchased funds (x_1), core deposits (x_2), and number of full- time equivalent employees (x_3). In addition, we include physical (x_4) and financial (equity) capital (x_5) as inputs for the IDF or quasi-fixed inputs for the cost function. We also include time as an extra variable to account for shifts in the technology and possible variations in RTS estimates over time.

We use three different samples for the estimation covering the period between 2000 and 2010. We obtain data for BHC from the Federal Reserve end-of-year FR Y-9C reports and for commercial banks from end-of-year Call reports. Except for employees' information, all our variables are nominal stock variables. We use the U.S. GDP implicit price deflator to transform all nominal variables to 2010 prices. Table 1 reports summary statistics for the three samples. Sample I includes 8,265 observations for 1,418 top-tier BHC with assets above \$500 million. Sample II includes 10,229 observations for 1,640 commercial banks with assets above \$500 million. Sample III includes sample I plus 1,287 observations for 262 independent commercial banks not owned by a reporting BHC with assets above \$500 million.

4 Returns to Scale Estimation

To compute our measure of RTS given by (10), we need estimates for the input distance functions $D_I(y,x)$ and $D_I(\gamma y,x)$. Following Wheelock and Wilson (2012, 2011), we assume away inefficiency and use nonparametric kernel estimation methods to avoid making arbitrary assumptions about the functional form of the underlying technology. After imposing the linear homogeneity property (in inputs) on the IDF and including time into $D_I(y,x)=1$, the estimating equation becomes:

$$-\ln x_1 = \ln D_I(y,\tilde{x},t) + \varepsilon_{it} \tag{15}$$

where $\tilde{x} = \{\ln(x_2/x_1), \ln(x_3/x_1), ..., \ln(x_5/x_1)\}$, $y = \{\ln y_1, \ln y_2, ..., \ln y_5\}$, $t = \{1,2,...,11\}$ indexes time periods from 2001 to 2010, and the random error term, ε_{it}, captures the stochastic nature of the IDF.

Advances in nonparametric econometrics allows us to estimate (15) by smoothing both the continuous (y and $\tilde{x}$), and the ordered (t) variables. To do this, we use a generalized kernel estimation along the lines of Li and Racine (2004) and Racine and Li (2004). Specifically, we use a Local Linear Least Squares (LLLS) estimator which performs weighted least-squares regressions around a point x_0 with weights determined by a kernel function and a bandwidth vector. Observations closer to x_0 receive more weight as a function of their associated bandwidth. We use a second order Gaussian kernel and Least Squares Cross-Validation (LSCV) to choose the bandwidths. We use the Wang and Van Ryzin (1981) bandwidth for the ordered variable. Hall, Li and Racine (2007) show that LSCV has desirable properties like the ability to smooth away irrelevant variables or to detect if some variables enter linearly in the estimation.

After estimating (15), we obtain $\hat{D}_I(x_0)$, an estimate of $D_I(y,\tilde{x},t)$ evaluated at an arbitrary vector of regressors x_0. To compute our measure of RTS, $S^{IDF}(\gamma|y,x)$, given in (10), we evaluate $\hat{D}_I(\gamma y,\tilde{x},t)$ for $\gamma = \{0.95, 0.97, ..., 1.05\}$, where $(y,\tilde{x},t)$ corresponds to the vector of original regressors. For $\gamma = 1$, we obtain the estimated values of $D_I(y,\tilde{x},t)$ which enter in the denominator of (15). Use of different values of $\gamma \in [0.95, 1.05]$ allows us to estimate RTS between the 95% and the 105% of the observed output vector for each banking organization. The special case considered by W&W corresponds to computing the ratio between $S^{IDF}(1.05|x_0)$ and $S^{IDF}(0.95|x_0)$, as we showed before.

To determine if a given observation show evidence consistent with increasing, decreasing, or constant RTS, we need to compute confidence intervals for each point estimate of our RTS measures. A given value of $\hat{S}^{IDF}(\gamma|x_0)$ indicates increasing RTS (IRTS) if it exhibits NDRTS and not CRTS. It indicates decreasing RTS (DRTS) if it exhibits NIRTS and not CRTS. It indicates constant RTS (CRTS) if it exhibits neither NIRTS nor NDRTS.

To construct bias-corrected confidence interval (CI), we use the wild bootstrap procedure. We bootstrap $S^{IDF}(\gamma|x_0)$ for each observation 399 times. This involves estimating $\hat{D}_I(\gamma y,\tilde{x},t)$ for $\gamma = \{0.95, 0.97, ..., 1.05\}$ 399 times, taking their associated residuals, drawing new residuals from a two-value distribution, as detailed in Hardle and Mammen (1993), and estimating $S^{IDF}(\gamma|x_0)$ again. After this, we estimate the standard errors and the bias of $\hat{S}^{IDF}(\gamma|x_0)$ and construct 95% asymptotic normal CI.

For $\gamma > 1$, a given observation exhibits IRTS if the lower bound of its associated CI lies above 1. It exhibits CRTS if the lower bound of its associated CI lies below 1 and its upper bound above 1. It exhibits DRTS if the upper bound of its associated CI lies below 1. For $\gamma < 1$, the reverse conditions holds. That is, a given observation exhibits IRTS if the upper bound of its associated CI lies below 1. It exhibits CRTS if the lower bound of its associated CI lies below 1 and its upper bound above 1. It exhibits DRTS if the lower bound of its associated CI lies above 1.

5 Empirical Results

5.1 Ray-Scale Economies

We estimate the model in (15) as described in Section 4 for samples I, II, and III. We obtain similar results across the three samples. Table 2 presents the bandwidths estimates, their associated scale factors, residual standard errors, cross-validation objective functions (CVOF), and R^2 values for the three samples. All variables are treated as continuous variables except for t, which is treated as an ordered variable.[5] The R^2 values indicate a good fit of the model in (15) for the three different samples. The estimated bandwidths differ across the different samples but their variation seems small. In addition, the estimated bandwidth are small indicating that all variables are relevant in the estimation.

Using the estimated models for the three different samples, we compute our *ray-scale economies* (RSE) measure of RTS based on the IDF using (10) and their corresponding confidence intervals as described in the previous section. Table 3 reports[6] the median values of RSE estimates by size quartiles based on total assets for $\gamma = \{0.95, 0.97, ..., 1.05\}$. For $\gamma < 1$ a median value of RSE less (greater) than 1 indicates IRTS (DRTS). For $\gamma > 1$, a median value of RSE greater (less) than 1 indicates IRTS (DRTS). If the median value equals 1, it indicates CRTS. Thus, according with the median values of RSE for each quartile and each value of γ in Table 3, the median banking organization for each sample exhibits IRTS.

To determine if a given observation shows evidence consistent with IRTS, CRTS, or DRTS, we

[5]We use the parallel implementation of the np-package in R , Hayfield and Racine 2008, for our estimations. The R^2 values are computed as the squared of the correlation coefficient between the left-hand-side variable in (15) and its nonparametric estimate.

[6]We dropped 9 observations from sample I, 14 observations from sample II, and 20 observations from sample III for which the estimated RSE were highly implausible.

construct bias corrected confidence intervals for RSE as described in the previous section. The last three columns for each sample show the percentage of observations in each size quartile (based on total assets) for which the bias corrected wild bootstrap confidence intervals indicate IRTS, CRTS, or DRTS. For instance, for Sample I and $\gamma = 0.95$, 47% of the observations falling within the first quartile show evidence consistent with IRTS, 40.84% with CRTS, and 12.16% with DRTS. For the fourth quartile, the corresponding values are 43.51%, 26.89%, and 29.60%. Overall, for $\gamma = 0.95$, 41.40%, 37.49%, and 29.60% of the observations show evidence consistent with IRTS, CRTS, and DRTS, respectively. An important result shown in Table 3 is that the commercial banks, which are included in Sample II, show stronger evidence in favor of IRTS compared to the BHC, which are included in Sample I.

The most salient result in Table 3 is that, contrary to Wheelock and Wilson (2012), not all banking organizations exhibit IRTS. This results holds across the three different samples. Another important result is that the number of observations indicating DRTS increase monotonically as total assets increase. For samples I and III, the number of observations indicating IRTS decreases as total assets increase up to the third quartile. However, the number of observations indicating IRTS tends to be higher for the largest banking organizations, those in the 4^{th} quartile. In contrast, for sample II, which includes commercial banks only, the number of observations indicating IRTS decreases monotonically as total assets increase.

The results in Table 3 offer a more detailed description of the nature of RTS around the observed output and input levels. For each banking organization, starting at 95% and continuing until 105% of its observed output levels, the Table shows that RSE measures change depending on the scaling factor γ by which outputs are scaled up or down. However, for different values of γ, the number of observations indicating IRTS, CRTS, or DRTS varies only slightly across all size quartiles. This indicates that around $\pm 5\%$ of the observed output levels, our measure of RTS is robust and stable. In Sample I, for $\gamma = 0.95$, 41.40% of the observations indicate IRTS while for $\gamma = 1.05$ this number is only 39.51%. Thus, some observations (156) show IRTS when outputs are scaled down to 95% of their observed level but CRTS or DRTS when outputs are scaled up to 105% of their observed level. The same holds true for samples II and III. In these cases, the *expansion-path scale economy* measure of W&W shown in (13) will be ill defined and will fail to give a correct indication of the nature of RTS for some observations.

Table 4 presents summary statistics of the proportion by which input quantities will change when all outputs are multiplied by a factor $\gamma > 0$. Panel A, B, and C, show results for samples I, II, and III, respectively. For sample I, multiplying all outputs by $\gamma = 1.05$ requires an average increase in all input quantities by a factor of 1.0488 which is less than $\gamma = 1.05$, indicating IRTS. Alternatively, multiplying all output quantities by $\gamma = 0.95$ requires an average decrease in all input quantities by a factor of 0.9511 which is greater than $\gamma = 0.95$, indicating IRTS as well. Under cost minimization, the factors presented in Table 4 represent the factors by which total costs will increase or decrease when all outputs are scaled up or down by γ.

Table 4 shows that despite the statistical evidence presented in Table 3 indicating the existence of economies of scale for some of the biggest BHC and commercial banks in the U.S., economies of scale seems to be economically small for the range of output changes we consider. Regardless of the value of γ, the median RSE estimate for all three samples is around 1, indicating nearly CRTS for all observations. Less than 1% of the observations show RSE measures consistent with economically significant scale economies. For instance, only around 34 observations indicate economies of scale greater than 1.03, in which case a 10% increase in all outputs will require an increase in all inputs by approximately 9.7% ($1/1.03 \times 10\%$). For all other observations, an increase of 10% in all outputs will require an increase in all inputs of approximately the same magnitude.

Our results regarding the economic significance of RTS for the largest U.S. banking organizations sharply contrast with the findings of Wheelock and Wilson (2012). They find that even the largest BHC exhibit sizable economies of scale. However, our empirical evidence gives a different picture. Despite the existence of scale economies for some large BHC and commercial banks, these economies of scale seem to be economically small. This result is surprising since our methodology closely follow Wheelock and Wilson (2012) and includes their model of bank costs as a special case. We use exactly the same definitions for outputs and inputs used in Wheelock and Wilson

(2012) and our econometric approaches are similar. The main two differences between our work and theirs is that we use more recent sample periods and avoid using dimension reduction techniques based on principal components analysis (PCA). However, we think our sample period is more relevant for policy analysis and is sufficiently large to give very precise estimates. Further, we see data reductions techniques unnecessary in our framework and instead use all the available information in the raw data. Using PCA transformations may produce noisier estimates. For example, PCA reduces the data to the same linear combinations of right-side variables regardless of their relation with the left-hand side variable. In addition, PCA is only able to find linear relations among the subject variables and cannot properly account for nonlinearities. Thus, our method offers several advantages over Wheelock and Wilson (2012) that make our results more compelling.

For completeness, we estimate the cost model of Wheelock and Wilson (2012) using Sample I for BHC. They consider the same output and input variables we use for the IDF. However, since no accurate input prices exist for physical and financial capital, they consider them as quasi-fixed inputs. Table 5 reports the nonparametric regression bandwidth estimates for this model. Table 6 presents the median RSE estimates using (8) and the number of observations exhibiting IRTS, CRTS, or DRTS. For comparability with our IDF results and unlike Wheelock and Wilson (2012), we do not use any dimension reduction technique to estimate their model. Thus, the results are readily comparable with those presented in Table 3 for our IDF based RSE estimates. Using the cost function, we find more evidence in favor of IRTS. However, we still find observations indicating constant or decreasing RTS. Overall, about 73% of the observations show evidence of IRTS and 20% indicate CRTS, and the remaining 7% indicate DRTS.

In terms of the economic significance of the RTS estimates from the cost function, our results indicate that despite the existence of IRTS, economies of scale seems to be small and not as sizable as those presented in Wheelock and Wilson (2012). Table 7 presents summary statistics of the proportion by which total costs would change if outputs were scaled up or down by γ. For instance, multiplying all outputs by $\gamma = 1.05$ leads to an average increase in total cost by approximately 4.32% which is less than the 5% increase in all outputs, indicating IRTS. Likewise, multiplying all output quantities by $\gamma = 0.95$ leads to an average decrease in total cost by about 5.66% which is more than the decrease in all outputs of 5%, again, indicating IRTS. Overall, the results from both the IDF and the cost function show that economies of scale in the banking industry seems to be small and that there are some banking organizations that operate under constant or decreasing returns to scale.

The differences between our RTS estimates from the IDF and the cost function are not surprising since these two approaches use different data. Moreover, RTS estimates from the cost function should be interpreted as short run RTS since the cost model includes physical and financial capital as quasi-fixed inputs. In contrast, RTS estimates from the IDF assumes that physical and financial capital are variable inputs; a more appropriate assumption for the long run. In any case, our evidence does not support the view that all U.S. banking organizations exhibit IRTS and given that the benefits from scale economies seem to be small, breaking up some of the largest banking organization may not lead to substantial welfare losses.

5.2 Elasticity-Based Returns to Scale

Traditional measures of RTS involve estimating elasticities of the cost function with respect to each output. In this case, a measure of RTS is given by:

$$\eta(w,y) = \left(\sum_{l} \frac{\partial \log C(w,y)}{\partial \log y_l} \right)^{-1} \tag{16}$$

Alternatively, using an input distance function, this measure is equivalent to[7]

$$\eta(x,y) = -\left(\sum_{l} \frac{\partial \log D(x,y)}{\partial \log y_l} \right)^{-1} \tag{17}$$

Values of $\eta(\cdot) \gtreqless 1$ indicate increasing, constant, or decreasing RTS, respectively. To compute (16) ((17)), one needs to estimate the derivatives of the cost function (IDF). Wheelock and Wilson

[7]This elasticity based formula is a special case of the formula based on the transformation function $F(x,y) = 1$ for which RTS is defined as $RTS(x,y) = -\sum_j \frac{\partial \log F(\cdot)}{\partial \log x_j} \div \sum_l \frac{\partial \log F(\cdot)}{\partial \log y_l}$ (Caves et al., 1981). For IDF, $\sum_j \frac{\partial \log F(\cdot)}{\partial \log x_j} = 1$.

(2012, 2011) argue that derivative estimates of the nonparametric cost function are noisier than estimates of the cost function itself and for this reason they estimate RTS using (13) instead of (16). However, they did not offer any empirical evidence to support their claims. In the following, we estimate (16) and (17) and compare the results with our previous RSE estimates.

We refer to (16) and (17) as *elasticity-based* RTS (EB-RTS). To estimate them, we need the estimated derivatives of the cost and input distance functions with respect to each output. Fortunately, since these values are obtained as a by-product of the local linear least squares nonparametric regression approach; the only additional task is to estimate confidence intervals.

We compute (16) for Sample I and (17) for Samples I, II, and III. Tables 8 and 9 report the results.[8] Panel A of Table 8 shows the results for the EB-RTS estimates using the IDF and Panel B those using the cost function. Again, we report separate results for each size quartile based on total assets. Overall, the results from the IDF for the three different samples closely mirror those reported in Table 3 in terms of the proportion of observations for each sample consistent with IRTS, CRTS, and DRTS. For instance, the results presented for Sample I in Table 3 show that, on average, 40.39%, 37.74%, and 21.88% of the observations show evidence of IRTS, CRTS, and DRTS, respectively. Table 8 shows that for Sample I, the corresponding numbers using the elasticity-based RTS estimates are 39.36%, 40.80%, and 19.84%. The similarities also hold within each size quartile. In addition, the results are also similar for samples II and III. Thus, using the EB-RTS estimates, our previous results regarding the nature of economies of scale of the biggest banking organizations are essentially unchanged.

In terms of the EB-RTS estimates using the cost function, the results are also comparable with our previous results based on RSE and presented in Table 6. It can be seen from that table that 73%, 20%, and 7% of the observations show evidence consistent with IRTS, CRTS, and DRTS, respectively. Table 8 shows that we obtain similar results using the EB-RTS estimates—75%, 20%, and 5%, respectively. Again, our previous results remain unchanged.

[8] The last column of Table 9 presents the number of observations included in the computations. We exclude values of RTS estimates below the 0.5% and above the 99.5% percentiles which are outliers or economically implausible. The presence of outliers partially support the argument in Wheelock and Wilson (2012, 2011) regarding the estimation of derivatives of nonparametric functions.

Table 8 shows that the median EB-RTS estimates using the IDF is close to 1 for Sample I and III, and slightly above 1 for Sample II. These results are consistent with our previous finding that economies of scale seem to be economically small. Table 9 presents additional summary statistics for EB-RTS estimates. This table shows that, in general, EB-RTS estimates are higher than their counterparts derived using (10).

An important difference is that EB-RTS estimated using the cost function seem economically significant and higher than the corresponding values estimated using (8). Further, they seem to be higher than those estimated using the IDF. The differences may stem from the fact that the IDF treats physical and financial capital as variable inputs while the cost function treats them as quasi-fixed inputs. To make them comparable, we adjust the cost function EB-RTS estimates by multiplying them by one minus the sum of the elasticities of the cost function with respect to both physical and financial capital (see Caves et al. 1981). The results of doing this adjustment are presented in the Panel B of Table 9. This adjustment lowers the median estimate of RTS from 1.17 to 1.05, suggesting that the unadjusted EB-RTS are upwardly biased. Further, adjusted EB-RTS estimates are closer to those estimated from the IDF.[9] Overall, we find that the EB-RTS estimates favor our earlier results that not all the largest banking organization in the U.S. exhibit increasing returns to scale and that scale economies, when they are present, are economically small.

6 Conclusions

The debate over Too-Big-To-Fail banks in the U.S. has brought about a renewed interest in the study of economies of scale in the U.S. banking industry. If large banking organizations enjoy economies of scale that benefit society as a whole, breaking up the biggest banks into smaller institutions may have deleterious consequences for the economy and the development of the banking industry. Our work sheds light in the nature and existence of scale economies at the biggest banking organizations

[9] By the LeChatelier Principle, assuming that physical and financial capital are quasi-fixed inputs, total cost will respond slowly to changes in other variables compared to the case when all inputs are variable. Thus, measured elasticities are lower than they would be if all inputs were variable. Correspondingly, measured RTS will tend to be higher since they are computed as the inverse of a sum of elasticities (as shown in equations (16) and (17)).

in the U.S. We use nonparametric methods to estimate returns to scale for large Bank Holding Companies (BHC) and Commercial Banks (CB) in the U.S. Our model includes as a special case the model of Wheelock and Wilson (2012) and offer a robustness check regarding their surprising finding that all BHC and CB in the U.S. enjoy sizable economies of scale.

Our method offers several advantages over Wheelock and Wilson (2012) that make our evidence more compelling. First, instead of using a cost function, we use an input distance function (IDF) which requires no information on input prices. Input price data are not readily available and one needs to construct them based on expenditure and input data; adding additional noise to the estimation. Second, since our approach dispense from input prices, our model of bank production can include physical and financial capital as proper input quantities. These items are traditionally excluded from the analysis or treated as quasi fixed inputs given that their associated prices are not readily available (e.g., Wheelock and Wilson 2012 and Berger and Mester 2003). Third, our sample period is more relevant for the current policy debate regarding the existence of economies of scale at large banking organizations since it covers a more recent time period. Fourth, the shorter sample period we use allows us to dispense from data dimensionality reduction techniques that may lead to noisier estimates of returns to scale without sacrificing precision and interpretability of the results.

Contrary to Wheelock and Wilson (2012), but in line with conventional wisdom, we find that not all BHC and CB exhibit increasing returns to scale (IRTS). In addition, economies of scale for those banking organizations experiencing IRTS seems small. Thus, breaking up large banking organizations into smaller institutions is unlikely to pose significant costs to the economy as a whole.

We find that our results are robust to alternative ways to estimating RTS. Namely, we estimate nonparametric measures of RTS using the traditional elasticity-based approach and find that our results are unchanged. In particular, both approaches give essentially the same results concerning the distribution of RTS for the biggest banking institutions. Thus, claims that elasticity-based nonparametric return to scale measure are inadequate seem overblown.

References

Berger, Allen N. and Loretta J. Mester, "Explaining the Dramatic Changes in Performance of US banks: Technological Change, Deregulation, and Dynamic Changes in Competition," *Journal of Financial Intermediation*, 2003, *12* (1), 57 – 95.

Boyd, J.H. and A. Heitz, "The Social Costs and Benefits of Too-Big-To-Fail Banks: A "Bounding" Exercise," Technical Report, University of Minnesota 2012.

Caves, Douglas W., Laurits R. Christensen, and Joseph A. Swanson, "Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955-74," *The American Economic Review*, 1981, *71* (5), pp. 994–1002.

Das, Abhiman and Subal C. Kumbhakar, "Productivity and efficiency dynamics in Indian banking: An input distance function approach incorporating quality of inputs and outputs," *Journal of Applied Econometrics*, 2012, *27* (2), 205–234.

Färe, R. and D. Primont, *Multi-output Production and Duality: Theory and Applications*, Springer, 1995.

Färe, Rolf., *Fundamentals of production theory*, Springer-Verlag Berlin, 1988.

Feng, Guohua and Apostolos Serletis, "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*, 2010, *34* (1), 127 – 138.

_ **and Xiaohui Zhang**, "Productivity and efficiency at large and community banks in the US: A Bayesian true random effects stochastic distance frontier analysis," *Journal of Banking & Finance*, 2012, *Forthcoming* (0).

Haldane, Andrew, "On being the right size," Speech given at Institute of Economic Affairs' 22nd Annual Series, The 2012 Beesley Lectures at the Institute of Directors, Pall Mall. October 25, 2012.

Hall, P., Q. Li, and J.S. Racine, "Nonparametric estimation of regression functions in the presence of irrelevant regressors," *The Review of Economics and Statistics*, 2007, *89* (4), 784–789.

Hardle, W. and E. Mammen, "Comparing Nonparametric Versus Parametric Regression Fits," *The Annals of Statistics*, 1993, *21* (4), pp. 1926–1947.

Harrison, William B., "In Defense of Big Banks," August 22, 2012.

Hayfield, Tristen and Jeffrey S. Racine, "Nonparametric Econometrics: The np Package," *Journal of Statistical Software*, 2008, *27* (5).

Hughes, Joseph P. and Loretta J. Mester, "Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function," Working Papers 11-27, Federal Reserve Bank of Philadelphia 2011.

Johnson, Simon, "Tarullo Telegraphs Fed's Plans to Cap Bank Size," Bloomberg News' Column December 2012.

Li, Qi. and Jeffrey S.. Racine, "Cross-validated local linear nonparametric regression," *Statistica Sinica*, 2004, *14* (2), 485–512.

Racine, Jeff and Qi Li, "Nonparametric estimation of regression functions with both categorical and continuous data," *Journal of Econometrics*, 2004, *119* (1), 99 – 130.

Restrepo-Tobón, Diego, Subal C. Kumbhakar, and Kai Sun, "Are U.S. Commercial Banks Too Big?," Technical Report, Binghamton University, EAFIT University, and Aston University 2012.

Rosenblum, Harvey, "Choosing the road to prosperity: why we must end too big to failŮnow," *Annual Report*, 2011, pp. 3–24.

Sealey, C.W. and J.T. Lindley, "Inputs, outputs, and a theory of production and cost at depository financial institutions," *Journal of Finance*, 1977, *32* (4), 1251–1266.

Tarullo, Daniel K., "Financial Stability Regulation," Speech at the Distinguished Jurist Lecture, University of Pennsylvania Law School, Philadelphia, Pennsylvania. October 10, 2012.

_ , "Industry Structure and Systemic Risk Regulation," Speech at the Brookings Institution Conference on Structuring the Financial Industry to Enhance Economic Growth and Stability, Washington, D.C. December 4, 2012.

_ , "Regulation of Foreign Banking Organizations," Speech at the Yale School of Management Leaders Forum, New Haven, Connecticut. November 28, 2012.

Wack, Kevin, "Big-Bank Breakup Popular with Rank and File of Both Parties," September, 5, 2012.

Wallison, Peter J., "Breaking Up the Big Banks: Is Anybody Thinking?," August-September 2012.

Wang, M.C. and J. Van Ryzin, "A class of smooth estimators for discrete distributions," *Biometrika*, 1981, *68* (1), 301–309.

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

_ **and** _ , "Do Large Banks Have Lower Costs? New Estimates of Returns to Scale for U.S. Banks.," *Journal of Money, Credit and Banking*, 2012, *44* (1), 171 – 199.

Table 1: Data summary statistics

	Mean	sd	Min	Percentiles					Max
				5^{th}	25^{th}	50^{th}	75^{th}	95^{th}	
				Sample I: BHC					
TA	16,200,000	115,000,000	500,000	539,000	697,000	1,060,000	2,510,000	39,400,000	2,290,000,000
C	782,000	5,490,000	5,049	23,065	33,122	51,366	120,000	1,920,000	140,000,000
y_1	1,460,000	12,200,000	20	2,166	11,677	31,935	93,191	1,620,000	238,000,000
y_2	4,320,000	25,600,000	224	217,000	352,000	548,000	1,220,000	11,200,000	553,000,000
y_3	2,150,000	13,700,000	7	22,042	64,335	129,000	310,000	4,520,000	232,000,000
y_4	6,280,000	55,400,000	8,608	82,488	163,000	283,000	664,000	8,800,000	1,250,000,000
y_5	333,000	2,630,000	21	979	2,900	6,179	19,120	545,000	56,700,000
x_1	8,740,000	75,700,000	7,056	86,885	166,000	277,000	686,000	13,900,000	1,910,000,000
x_2	5,880,000	35,400,000	557	328,000	452,000	683,000	1,510,000	15,900,000	874,000,000
x_3	2,834	18,325	33	113	184	276	588	8,026	410,000
x_4	169,000	911,000	325	6,160	13,812	23,305	48,774	505,000	18,000,000
x_5	1,400,000	9,640,000	374	36,647	59,035	93,168	219,000	3,600,000	234,000,000
w_1	0.0370	0.0140	0.0050	0.0170	0.0280	0.0360	0.0460	0.0590	0.2280
w_2	0.0200	0.0100	0.0000	0.0060	0.0120	0.0180	0.0260	0.0380	0.2600
w_3	66.45	26.56	7.21	43.85	53.18	61.03	73.16	102.18	979.55
				Sample II: Commercial Banks					
TA	9,080,000	69,000,000	500,000	523,000	649,000	961,000	2,050,000	18,000,000	1,790,000,000
C	389,000	2,860,000	6,000	19,942	28,919	43,208	90,839	791,000	70,200,000
y_1	691,000	5,860,000	5.100	1,966	10,843	29,114	80,906	981,000	158,000,000
y_2	2,810,000	17,800,000	163.9	195,000	323,000	481,000	955,000	6,120,000	484,000,000
y_3	1,780,000	14,100,000	11.10	28,998	71,212	131,000	283,000	3,240,000	395,000,000
y_4	3,890,000	36,900,000	18,355	94,204	178,000	296,000	658,000	6,380,000	1,110,000,000
y_5	176,000	1,420,000	36.00	1,755	4,851	8,771	22,695	318,000	37,600,000
x_1	3,400,000	33,800,000	10,695	65,802	130,000	214,000	465,000	4,320,000	1,000,000,000
x_2	5,420,000	40,700,000	11,177	312,000	425,000	626,000	1,300,000	10,900,000	979,000,000
x_3	1,741	10,990.20	9.200	91.80	174.2	263.3	505.1	4,311	231,000
x_4	98,598.80	559,000.00	0.000	4,231	11,252	19,210	38,820	269,000	14,300,000
x_5	867,000.00	6,210,000.00	958.0	41,965	59,080	90,513	196,000	1,780,000	171,000,000
w_1	0.0400	0.0190	0.0050	0.0170	0.0280	0.0370	0.0480	0.0700	0.2490
w_2	0.0180	0.0090	-	0.0050	0.0110	0.0160	0.0240	0.0350	0.1340
w_3	59.5000	23.700	5.800	36.500	46.400	55.200	66.600	95.700	833.300
				Sample III: BHC and Independent Commercial Banks					
TA	14,900,000	107,000,000	500,000	537,000	690,000	1,050,000	2,480,000	41,300,000	2,290,000,000
C	714,000	5,110,000	5,049	22,124	32,245	49,990	116,000	1,920,000	140,000,000
y_1	1,330,000	11,400,000	5	1,783	10,887	30,458	90,745	1,710,000	238,000,000
y_2	4,060,000	24,000,000	202	208,000	343,000	531,000	1,150,000	11,800,000	553,000,000
y_3	2,020,000	12,800,000	7	22,378	64,822	129,000	316,000	5,160,000	232,000,000
y_4	5,820,000	51,700,000	8,608	84,586	166,000	289,000	683,000	10,100,000	1,250,000,000
y_5	300,000	2,450,000	21	990	3,071	6,426	19,370	554,000	56,700,000
x_1	7,870,000	70,500,000	7,056	81,675	159,000	273,000	681,000	13,300,000	1,910,000,000
x_2	5,630,000	33,300,000	557	320,000	446,000	671,000	1,490,000	17,900,000	874,000,000
x_3	2,596	17,085	9	103	179	270	571	7,500	410,000
x_4	155,000	851,000	0	5,165	13,143	22,454	47,131	476,000	18,000,000
x_5	1,330,000	9,010,000	374	37,297	59,214	93,711	225,000	3,900,000	234,000,000
w_1	0.0370	0.0150	0.0050	0.0160	0.0270	0.0360	0.0460	0.0610	0.2440
w_2	0.0190	0.0100	0.0000	0.0060	0.0120	0.0180	0.0260	0.0370	0.2600
w_3	67.08	28.23	6.70	43.24	53.02	61.32	73.91	104.69	979.55

Notes: This table reports summary statistics for the banking organizations used in the estimation. Data cover the period 2001 to 2010 and correspond to annual values as of end-of-quarter of each year. Sample I includes 8,265 observations for 1,418 top-tier BHC with assets above \$500 million. Sample II includes 10,125 observations for 1,640 commercial banks with assets above \$500 million. Sample III includes sample I plus 1,287 observations for 262 independent commercial banks not owned by a BHC with assets above \$500 million. Nominal values are in thousands of 2010 dollars. Total assets (TA) correspond to balance-sheet values. Total costs (TC) equal interest and noninterest expenses from balance sheet data. The output variables are: consumer loans (y_1), real estate loans (y_2), business loans (y_3), securities (y_4), and off-balance sheet output (y_5). The input variables are: purchased funds (x_1), core deposits (x_2), and labor (x_3). x_4 and x_5 correspond to the value of physical and equity capital, respectively.

Table 2: Input distance function nonparametric regression estimates

	Sample I		Sample II		Sample III	
Variable	Bandwidth	Scale Factor	Bandwidth	Scale Factor	Bandwidth	Scale Factor
$\ln x_2$	1.1292	4.068889	0.9276	3.045643	1.3186	4.596545
$\ln x_3$	0.3261	1.059223	0.3607	1.049982	0.3378	1.059479
$\ln x_4$	0.4012	1.059223	1.2462	3.087151	0.4194	1.059535
$\ln x_5$	0.2874	1.059223	0.2980	1.059223	0.2948	1.059273
$\ln y_1$	1.8149	2.359051	0.7869	1.071292	0.8230	1.060344
$\ln y_2$	0.8967	1.950924	1.4508	3.672538	1.7069	3.850792
$\ln y_3$	0.6171	1.059281	1.7517	3.466948	0.6142	1.059332
$\ln y_4$	1.5634	3.007355	1.1242	2.351977	0.7378	1.426104
$\ln y_5$	0.7399	1.059104	0.5949	1.059223	0.7143	1.059382
t	0.5000	1	0.5000	1	0.5000	1
Residual S.E.	0.0041		0.0017		0.0076	
CVOF	0.1700		0.0122		3.7996	
R^2	0.9983		0.9991		0.9969	
Observations	8,265		10,125		9,550	

Notes: This table shows the Local Linear Least-Squares non-parametric regression bandwidths estimates using Least-Squares Cross validation. All variables are treated as continuous variables except for t which is treated as an ordered variable. We use second-order Gaussian kernels for the continuous variables and an ordered categorical kernel for the ordered variable. Estimations are done using the parallel implementation of the np-package in R, Hayfield and Racine 2008. S.E. means standard error, CVOF means cross-validation objective function. The R^2 values are computed as the squared of the correlation coefficient between the left-hand-side variable in (15) and its nonparametric estimate.

Table 3: Ray-scale economies (RSE) summary statistics

	Sample I: BHC					Sample II: Commercial Banks					Sample III: BHC and Ind. Comm. Banks				
Quartile	Obs.	Median	IRTS	CRTS	DRTS	Obs.	Median	IRTS	CRTS	DRTS	Obs.	Median	IRTS	CRTS	DRTS
$\gamma = 0.95$															
1	2,064	0.9988	47.00	40.84	12.16	2,528	0.9972	72.11	23.66	4.23	2,383	0.9991	40.91	36.26	22.83
2	2,064	0.9993	40.41	41.52	18.07	2,528	0.9979	65.82	28.96	5.22	2,382	0.9995	38.08	35.05	26.87
3	2,064	0.9995	34.69	40.70	24.61	2,528	0.9985	52.69	37.10	10.21	2,383	0.9997	34.49	37.18	28.33
4	2,064	0.9995	43.51	26.89	29.60	2,527	0.9987	50.49	27.74	21.76	2,382	0.9996	44.67	25.06	30.27
	8,256		41.40	37.49	21.11	10,111		60.28	29.36	10.36	9,530		39.54	33.39	27.07
$\gamma = 0.97$															
1	2,064	0.9993	46.27	41.04	12.69	2,528	0.9984	71.36	24.01	4.63	2,383	0.9995	40.83	36.09	23.08
2	2,064	0.9996	39.92	41.52	18.56	2,528	0.9988	65.15	29.71	5.14	2,382	0.9997	36.57	35.85	27.58
3	2,064	0.9997	34.59	40.65	24.76	2,528	0.9991	52.69	37.18	10.13	2,383	0.9998	34.20	37.39	28.41
4	2,064	0.9997	43.07	27.33	29.60	2,527	0.9992	50.26	28.29	21.45	2,382	0.9997	44.25	25.15	30.60
	8,256		40.96	37.63	21.40	10,111		59.87	29.80	10.34	9,530		38.96	33.62	27.42
$\gamma = 0.99$															
1	2,064	0.9998	45.83	41.47	12.69	2,528	0.9995	71.12	24.25	4.63	2,383	0.9999	39.99	36.30	23.71
2	2,064	0.9999	38.71	42.10	19.19	2,528	0.9996	64.99	29.59	5.42	2,382	0.9999	36.06	35.77	28.17
3	2,064	0.9999	34.06	40.94	25.00	2,528	0.9997	52.10	37.58	10.32	2,383	0.9999	34.28	37.43	28.28
4	2,064	0.9999	43.51	26.99	29.51	2,527	0.9998	50.22	28.22	21.57	2,382	0.9999	43.79	25.65	30.56
	8,256		40.53	37.88	21.60	10,111		59.61	29.91	10.48	9,530		38.53	33.79	27.68
$\gamma = 1.01$															
1	2,064	1.0002	45.30	41.62	13.08	2,528	1.0005	70.89	24.37	4.75	2,383	1.0002	39.19	36.72	24.09
2	2,064	1.0001	37.94	42.34	19.72	2,528	1.0004	64.16	30.50	5.34	2,382	1.0001	35.47	35.94	28.59
3	2,064	1.0001	34.06	40.26	25.68	2,528	1.0003	52.06	37.42	10.52	2,383	1.0001	33.95	37.26	28.79
4	2,064	1.0001	43.12	27.42	29.46	2,527	1.0002	50.34	27.86	21.80	2,382	1.0001	43.95	25.06	30.98
	8,256		40.10	37.91	21.98	10,111		59.36	30.04	10.60	9,530		38.14	33.75	28.11
$\gamma = 1.03$															
1	2,064	1.0006	44.53	41.57	13.91	2,528	1.0015	70.09	24.88	5.02	2,383	1.0004	38.65	36.30	25.05
2	2,064	1.0004	37.55	42.10	20.35	2,528	1.0011	64.00	30.66	5.34	2,382	1.0002	34.93	36.06	29.01
3	2,064	1.0003	34.06	40.16	25.78	2,528	1.0008	52.57	36.63	10.80	2,383	1.0002	33.53	37.31	29.16
4	2,064	1.0003	43.12	27.23	29.65	2,527	1.0007	50.14	28.45	21.41	2,382	1.0003	44.29	24.60	31.11
	8,256		39.81	37.77	22.42	10,111		59.20	30.16	10.64	9,530		37.85	33.57	28.58
$\gamma = 1.05$															
1	2,064	1.001	43.99	41.76	14.24	2,528	1.0024	69.66	25.20	5.14	2,383	1.0007	37.94	36.59	25.47
2	2,064	1.0006	37.11	41.62	21.27	2,528	1.0018	63.25	31.45	5.30	2,382	1.0003	34.01	36.15	29.85
3	2,064	1.0004	34.16	39.87	25.97	2,528	1.0013	52.37	36.55	11.08	2,383	1.0003	33.40	37.35	29.25
4	2,064	1.0005	42.78	27.71	29.51	2,527	1.0012	50.02	28.33	21.65	2,382	1.0004	44.12	24.56	31.32
	8,256		39.51	37.74	22.75	10,111		58.83	30.38	10.79	9,530		37.37	33.66	28.97

Notes: This table shows the median values of our return to scale (RTS) estimates for different values of γ by size quartiles based on total assets and the percentage of observation exhibiting increasing, constant, and decreasing returns to scale (IRTS, CRTS, and DRTS, respectively). For $\gamma < 1$ ($\gamma > 1$), a median value of RTS less (greater) than one indicates IRTS. For $\gamma > 1$ ($\gamma < 1$), a median value of RTS greater (less) than one indicates DRTS. If the median value equals one, it indicates CRTS. The last three columns for each sample show the percentage of observations in each quartile for which the bias corrected wild bootstrap confidence intervals indicates IRTS, CRTS, or DRTS.

Table 4: Summary statistics for the factor by which inputs would increase or decrease when all output quantities are multiplied by the factor $\gamma > 0$.

				Percentiles					
γ	mean	sd	min	5^{th}	25^{th}	Median	75^{th}	95^{th}	max
Panel A. Sample I: BHC									
0.95	0.9511	0.0066	0.6185	0.9480	0.9498	0.9507	0.9517	0.9550	1.1808
0.97	0.9707	0.0029	0.9178	0.9688	0.9699	0.9704	0.9710	0.9730	1.0572
0.99	0.9902	0.0012	0.9788	0.9896	0.9899	0.9901	0.9903	0.9910	1.0549
1.01	1.0098	0.0011	0.9652	1.0090	1.0097	1.0099	1.0101	1.0104	1.0234
1.03	1.0293	0.0036	0.8668	1.0270	1.0290	1.0296	1.0302	1.0313	1.0521
1.05	1.0488	0.0073	0.6915	1.0449	1.0483	1.0494	1.0503	1.0521	1.0866
Panel B. Sample II: Commercial Banks									
0.95	0.9521	0.0042	0.8575	0.9486	0.9507	0.9518	0.9530	0.9556	1.0879
0.97	0.9712	0.0025	0.9124	0.9691	0.9704	0.9711	0.9718	0.9733	1.0305
0.99	0.9904	0.0008	0.9701	0.9897	0.9901	0.9904	0.9906	0.9911	1.0100
1.01	1.0096	0.0008	0.9902	1.0089	1.0094	1.0096	1.0099	1.0103	1.0307
1.03	1.0291	0.0156	0.6327	1.0266	1.0282	1.0289	1.0296	1.0309	1.8444
1.05	1.0482	0.0164	0.6446	1.0443	1.0470	1.0482	1.0493	1.0515	1.8795
Panel C. Sample III: BHC and Independent Commercial Banks									
0.95	0.9511	0.0050	0.8850	0.9472	0.9494	0.9505	0.9518	0.9558	1.0518
0.97	0.9706	0.0030	0.9454	0.9683	0.9696	0.9703	0.9711	0.9735	1.0305
0.99	0.9902	0.0010	0.9816	0.9894	0.9899	0.9901	0.9904	0.9912	1.0100
1.01	1.0098	0.0010	0.9903	1.0088	1.0097	1.0099	1.0101	1.0106	1.0186
1.03	1.0294	0.0031	0.9713	1.0265	1.0290	1.0297	1.0304	1.0318	1.0562
1.05	1.0490	0.0054	0.9503	1.0441	1.0483	1.0496	1.0507	1.0531	1.0947

Notes: This table shows the factor by which inputs quantities increase or decrease when all output quantities are multiplied by the factor $\gamma > 0$ appearing in column one. Panel A, B, and C, show results for samples I, II, and III, respectively. For instance, for sample I, multiplying all outputs by $\gamma = 1.05$ requires an average increase in all input quantities by 1.0488 which is less than $\gamma = 1.05$, indicating increasing returns to scale. Likewise, multiplying all output quantities by $\gamma = 0.95$ requires an average decrease in all input quantities by 0.9511 which is greater than $\gamma = 0.95$, again, indicating increasing returns to scale.

Table 5: Cost function nonparametric regression estimates

Variable	Bandwidth	Scale Factor
$\ln w_1$	3.8656	16.5420
$\ln w_2$	1.2083	3.8636
$\ln y_1$	0.8149	1.0592
$\ln y_2$	1.4868	3.2350
$\ln y_3$	0.6164	1.0580
$\ln y_4$	0.5507	1.0592
$\ln y_5$	0.7400	1.0592
$\ln x_4$	0.4950	1.0592
$\ln x_5$	0.5141	1.0592
t	0.5000	1
Residual S.E.	0.0031	
R^2	0.9982	
CVOF	0.0141	
Observations	8,265	

Notes: This table shows the Local Linear Least-Squares nonparametric regression bandwidths estimates using least-squares cross validation for Wheelock and Wilson (2012)'s cost model using data for BHC (SampleI). The left-hand side variable equals $\ln C/w_3$. All variables are treated as continuous variables except for t which is treated as an ordered variable. We use second-order Gaussian kernels for the continuous variables and an ordered categorical kernel for the ordered variable. Estimations are done using the parallel implementation of the np-package in R, Hayfield and Racine 2008. S.E. means standard error, CVOF means cross-validation objective function. The R^2 value is computed as the squared of the correlation coefficient between the left-hand-side variable in (15) and its nonparametric estimate.

Table 6: Cost function ray-scale rconomies (RSE) estimates for Sample I

Quartile	Obs.	Median	IRTS	CRTS	DRTS	Quartile	Obs.	Median	IRTS	CRTS	DRTS
		$\gamma = 0.95$						$\gamma = 1.01$			
1	2066	1.0071	72.17	23.91	3.92	1	2066	0.9933	73.72	22.41	3.87
2	2065	1.0078	79.66	17.53	2.81	2	2065	0.9925	80.48	16.66	2.86
3	2066	1.0073	76.43	18.44	5.13	3	2066	0.9933	76.09	18.39	5.52
4	2065	1.0055	64.02	21.79	14.19	4	2065	0.9949	63.78	21.36	14.87
	8262		73.07	20.42	6.51		8262		73.52	19.70	6.78
		$\gamma = 0.97$						$\gamma = 1.03$			
1	2066	1.0042	72.60	23.43	3.97	1	2066	0.9959	73.33	22.75	3.92
2	2065	1.0047	79.61	17.48	2.91	2	2065	0.9955	79.61	17.48	2.91
3	2066	1.0043	76.23	18.54	5.23	3	2066	0.9959	75.99	18.30	5.71
4	2065	1.0032	64.02	21.55	14.43	4	2065	0.9969	63.73	21.26	15.01
	8262		73.12	20.25	6.63		8262		73.17	19.95	6.89
		$\gamma = 0.99$						$\gamma = 1.05$			
1	2066	1.0014	72.80	23.28	3.92	1	2066	0.9933	73.48	22.41	4.11
2	2065	1.0016	79.66	17.48	2.86	2	2065	0.9925	79.56	17.34	3.10
3	2066	1.0014	76.14	18.49	5.37	3	2066	0.9933	76.19	18.05	5.76
4	2065	1.0011	63.92	21.55	14.53	4	2065	0.9949	63.58	21.31	15.11
	8262		73.13	20.20	6.67		8262		73.20	19.78	7.02

Notes: This table shows the median values of return to scale (RTS)estimates based on the cost function for different values of γ by size quartiles based on total assets. IRTS, CRTS, and DRTS stand for increasing, constant, and decreasing returns to scale. For $\gamma < 1$ ($\gamma > 1$), a median value of RTS greater (less) than one indicates IRTS. For $\gamma > 1$ ($\gamma < 1$), a median value of RTS less (greater) than one indicates DRTS. If the median value equals one, it indicates CRTS. The last three columns for each sample show the number of observations in each quartile for which the bias corrected wild bootstrap confidence intervals indicates IRTS, CRTS, or DRTS. We use 99 bootstrap replicates to construct the confidence intervals.

Table 7: Summary statistics for the factor by which total costs would increase or decrease when all output quantities are multiplied by the factor $\gamma > 0$

				Percentiles					
γ	mean	sd	min	5^{th}	25^{th}	Median	75^{th}	95^{th}	max
0.95	0.9434	0.0076	0.8746	0.9345	0.9407	0.9433	0.9461	0.9525	1.0823
0.97	0.9660	0.0046	0.9247	0.9606	0.9644	0.9659	0.9677	0.9716	1.0508
0.99	0.9887	0.0016	0.9744	0.9868	0.9881	0.9886	0.9892	0.9906	1.0173
1.01	1.0087	0.0016	0.9948	1.0068	1.0081	1.0086	1.0092	1.0106	1.0384
1.03	1.0259	0.0048	0.9781	1.0203	1.0242	1.0258	1.0277	1.0318	1.1212
1.05	1.0432	0.0082	0.8975	1.0338	1.0402	1.0430	1.0461	1.0531	1.2123

Notes: This table shows the factor by which total costs would increase or decrease when all output quantities are multiplied by the factor $\gamma > 0$ appearing in column one. For instance, for sample I, multiplying all outputs by $\gamma = 1.05$ leads to an average increase in total costs by 1.0432 which is less than $\gamma = 1.05$, indicating increasing returns to scale. Likewise, multiplying all output quantities by $\gamma = 0.95$ leads to an average decrease in total costs by 0.9434 which is less than $\gamma = 0.95$, again, indicating increasing returns to scale.

Table 8: Scale economies from elasticity-based RTS estimates

Panel A: IDF Estimates

	Sample I: BHC					Sample II: Commercial Banks				
Quartile	Obs.	Median	IRTS	CRTS	DRTS	Obs.	Median	IRTS	CRTS	DRTS
1	2,041	1.0235	52.13	38.22	9.65	2,521	1.0506	77.67	20.67	1.67
2	2,056	1.0131	39.88	45.43	14.69	2,512	1.0372	73.61	24.32	2.07
3	2,060	1.0059	24.47	49.27	26.26	2,518	1.0257	60.76	35.15	4.09
4	2,024	1.0089	41.11	30.09	28.80	2,465	1.0234	52.94	29.21	17.85
Total	8,181	1.0119	39.36	40.80	19.84	10,016	1.0340	66.31	27.33	6.36

	Sample III: CB and BHC				
Quartile	Obs.	Median	IRTS	CRTS	DRTS
1	2,373	1.0154	42.60	39.15	18.25
2	2,377	1.0062	28.44	44.09	27.47
3	2,378	1.0009	23.17	40.45	36.38
4	2,326	1.0044	40.46	26.74	32.80
Total	9,454	1.0060	33.63	37.67	28.71

Panel B: Cost Function Estimates

	Sample I: BHC					Sample I - Adjusted : BHC				
Quartile	Obs.	Median	IRTS	CRTS	DRTS	Obs.	Median	IRTS	CRTS	DRTS
1	2,049	1.1718	79.01	19.03	1.95	2,045	1.0546	35.94	56.77	7.29
2	2,062	1.1754	80.26	17.94	1.79	2,057	1.0487	30.77	64.46	4.76
3	2,063	1.1581	76.49	19.39	4.12	2,060	1.0319	25.87	66.21	7.91
4	2,007	1.1152	64.42	22.47	13.10	2,019	1.0154	31.70	49.13	19.17
Total	8,181	1.1614	75.11	19.69	5.19	8,181	1.0401	31.06	59.20	9.74

Notes: This table shows the median values of elasticity-based return to scale (RTS) estimates by size quartiles based on total assets and the percentage of observation exhibiting increasing, constant, and decreasing returns to scale (IRTS, CRTS, and DRTS, respectively). Panel A and Panel B report information for elasticity-based RTS estimates based on the nonparametric input distance function (IDF) and the nonparametric cost function, respectively. The last three columns for each sample show the percentage of observations in each quartile for which the bias corrected wild bootstrap confidence intervals indicates IRTS, CRTS, or DRTS.

Table 9: Summary statistics for elasticity-based RTS Estimates

Quartiles	mean	sd	min	Percentiles 5^{th}	25^{th}	Median	75^{th}	95^{th}	max	N
Panel A: Elasticity-based RTS using the IDF.										
Sample I										
Total	1.0195	0.0527	0.8127	0.9699	0.9987	1.0119	1.0273	1.0928	1.5708	8,181
1	1.0302	0.0496	0.8194	0.9781	1.0077	1.0235	1.0432	1.1027	1.5214	2,041
2	1.0181	0.0404	0.8773	0.9786	1.0012	1.0131	1.0249	1.0706	1.5136	2,056
3	1.0083	0.0412	0.8127	0.9708	0.9954	1.0059	1.0159	1.0439	1.4854	2,060
4	1.0216	0.0716	0.8127	0.9537	0.9923	1.0089	1.0265	1.1445	1.5708	2,024
Sample II										
Total	1.0398	0.0449	0.8833	0.9897	1.0194	1.0340	1.0534	1.0981	1.5415	10,016
1	1.0563	0.0460	0.8833	1.0105	1.0333	1.0506	1.0688	1.1153	1.5045	2,521
2	1.0439	0.0388	0.8851	1.0099	1.0259	1.0372	1.0518	1.0937	1.5231	2,512
3	1.0303	0.0293	0.9034	1.0012	1.0166	1.0257	1.0388	1.0764	1.5088	2,518
4	1.0283	0.0556	0.8909	0.9570	1.0033	1.0234	1.0456	1.0999	1.5415	2,465
Sample III										
Total	1.0206	0.0943	0.7694	0.9536	0.9912	1.0060	1.0265	1.1127	2.2334	9,454
1	1.0278	0.0890	0.7769	0.9571	0.9980	1.0154	1.0385	1.1094	2.0604	2,373
2	1.0136	0.0675	0.7694	0.9586	0.9926	1.0062	1.0211	1.0782	2.1560	2,377
3	1.0080	0.0717	0.7715	0.9582	0.9899	1.0009	1.0138	1.0628	2.1596	2,378
4	1.0332	0.1332	0.7726	0.9382	0.9871	1.0044	1.0321	1.2032	2.2334	2,326
Panel B: Elasticity-based RTS using the the cost function.										
Sample I: Unadjusted RTS estimates										
Total	1.1731	0.1816	0.5613	0.9641	1.0975	1.1614	1.2220	1.3985	3.5603	8,181
1	1.1953	0.1834	0.5656	1.0369	1.1293	1.1718	1.2232	1.3989	3.5603	2,049
2	1.1848	0.1407	0.5786	1.0317	1.1213	1.1754	1.2302	1.3550	3.1433	2,062
3	1.1633	0.1313	0.5842	1.0021	1.0948	1.1581	1.2173	1.3537	2.3761	2,063
4	1.1486	0.2465	0.5613	0.8575	1.0268	1.1152	1.2158	1.5300	3.5067	2,007
Sample I: Adjusted RTS estimates										
Total	1.0504	0.1112	0.5602	0.9231	1.0075	1.0401	1.0768	1.1988	2.2932	8,181
1	1.0744	0.1254	0.6164	0.9558	1.0228	1.0546	1.0970	1.2453	2.2932	2,045
2	1.0562	0.0852	0.5707	0.9786	1.0217	1.0487	1.0751	1.1725	2.1336	2,057
3	1.0413	0.0747	0.5602	0.9630	1.0091	1.0319	1.0641	1.1477	1.9407	2,060
4	1.0293	0.1413	0.5685	0.8552	0.9666	1.0154	1.0696	1.2385	2.2364	2,019

Notes: Panel A shows summary statistics for elasticity-based RTS for samples I, II, and III by size quartiles based on total assets. IDF denotes estimates obtained from the input distance function using (17) and Cost denotes estimates from the cost function using (16). Panel B shows summary statistics for Technical Change estimates using the derivative of the nonparametric (log) IDF or (log) Cost function with respect to time. The last column presents the number of observations included in the computation of the summary statistics. We exclude values of RTS and Technical Change estimates below the 0.5% percentile and above the 99.5% which we considered outliers or economically implausible.