Staff Studies

Federal Deposit Insurance Corporation
Staff Studies

Report No. 2020-06
Economies of Scale in Community Banks

December 2020

FDIC

www.fdic.gov/cfr • @FDICgov • #FDICCFR • #FDICResearch

Economies of Scale in Community Banks*

Stefan Jacewitz,† Troy Kravitz,‡ and George Shoukry§
December 2020

Abstract:

Using financial and supervisory data from the past 20 years, we show that scale economies in community banks with less than $10 billion in assets emerged during the run-up to the 2008 financial crisis due to declines in interest expenses and provisions for losses on loans and leases at larger banks. The financial crisis temporarily interrupted this trend and costs increased industry-wide, but a generally more cost-efficient industry re-emerged, returning in recent years to pre-crisis trends. We estimate that from 2000 to 2019, the cost-minimizing size of a bank's loan portfolio rose from approximately $350 million to $3.3 billion. Though descriptive, our results suggest efficiency gains accrue early as a bank grows from $10 million in loans to $3.3 billion, with 90 percent of the potential efficiency gains occurring by $300 million.

JEL classification: G21, G28, L00.

* The views expressed are those of the authors and do not necessarily reflect the official positions of the Federal Deposit Insurance Corporation or the United States. FDIC Staff Studies can be cited without additional permission. The authors wish to thank Noam Weintraub for research assistance and seminar participants for helpful comments.

† Federal Deposit Insurance Corporation, sjacewitz@fdic.gov, 550 17th St. NW, Washington, DC 20429

‡ Federal Deposit Insurance Corporation, tkravitz@fdic.gov, 550 17th St. NW, Washington, DC 20429

§ Federal Deposit Insurance Corporation, gshoukry@fdic.gov, 550 17th St. NW, Washington, DC 20429

1 Introduction

Economies of scale occur when the per-unit cost of production falls as the number of units produced increases.[1] In the context of banking, scale economies exist when the cost per dollar of loans (or assets) declines as the number of loans (or assets) increases. An efficient bank is operating at the lowest cost per dollar of assets or loans. This study examines economies of scale in community banks—specifically, those institutions with less than $10 billion in assets (hereafter, “banks”)—over the past two decades. Our estimation period spans the 2008 financial crisis (hereafter, “the crisis”), allowing us to observe trends during the run-up to the crisis, its immediate aftermath, and the subsequent return to more normal economic conditions. The study’s main contributions are descriptive estimates of the shape and dynamics of the industry’s average cost curves over time. To the extent that the sample of banks with similar characteristics (e.g., size range and asset quality) is unchanged in different years, our analysis also sheds light on how the crisis affected the efficiency of banks across different segments of the industry (e.g., agricultural versus residential mortgage specialists).

To ensure that our results are not method-dependent, we use both nonparametric and parametric methods to analyze the economics of scale in banking.[2] Using nonparametric methods, we find that costs as a percentage of assets have been declining at both small and large banks.[3] The decline in costs at banks has been largely driven by a decrease in interest expenses and provisions, rather than changes to underlying noninterest expenses.[4]

The crisis seems to have increased costs for banks with weak assets more severely than banks with strong asset portfolios. After the crisis, however, costs generally decreased for banks across size and asset quality distributions. Trends in economies of scale in recent years are similar to those in the pre-crisis years.

We address bank efficiency by examining financial and supervisory data for banks and thrifts with less than $10 billion in assets. Measuring the output of a bank by its total loans and leases, we find that in 2000, per-unit costs were minimized when a bank’s loan portfolio was approximately $350 million.[5] Scale economies emerged during the run-up to the crisis, and the cost-minimizing loan portfolio size rose to approximately $800 million by 2006, before declining to $400 million during the crisis. The crisis hit large community banks hard, but those that survived seem more efficient than they were pre-crisis, as the decrease in average costs across the industry confirms. We estimate that the cost-minimizing loan portfolio size increased to more than $2.5 billion by 2014 and to just under $3.3 billion in 2019, though the confidence intervals around these point estimates are large. Our estimates are not causal and do not predict how a bank’s costs would change were it to change in size. We find evidence, however, that the overwhelming majority of any gains from increasing a bank’s loan production from $10 million to the cost-minimizing loan portfolio size of $3.3 billion accrue early in the growth process. Our nonparametric results suggest that once a loan portfolio reaches approximately $300 million, a bank has achieved about 90 percent of the potential efficiencies from increased scale; by $600 million, a bank has achieved about 95 percent of potential efficiencies.

Our parametric results—using loans rather than assets to measure bank output—show that although the estimated efficient bank size increased steadily during the run-up to the financial crisis, the industry was so severely affected by the crisis that the estimated efficient bank size during the crisis returned to nearly its 2000 level. This proved to be a temporary phenomenon, however, as the efficient size of a bank increased again post-crisis and through 2019. The difference in the

[1] “Diseconomies of scale’’ refers to the opposite case.

[2] We also consider alternative definitions of bank output.

[3] Statements about different classifications of banks (e.g., large versus small) are about the distribution of banks; the actual bank sample is variable year to year.

[4] This finding was also shown in Jacewitz and Kupiec (2012). There is some disagreement in the literature on the precise effects of post-crisis regulations on banks’ noninterest expense (for example, McCord and Prescott (2014), GAO (2015), and Hogan and Burns (2019)). Our results suggest that the majority of aggregate changes in cost were not primarily driven by noninterest expense. Although noninterest expense has fluctuated over time, increasing for some banks and decreasing for others (depending somewhat on bank size), the fluctuations in noninterest expense were generally smaller in magnitude than the fluctuations in the other components of our cost measures.

[5] In our sample, the loan portfolio is about 65 to 70 percent of a bank’s assets. Banks with strong asset portfolios tend to be on the lower end of this range, and banks with weak asset portfolios are generally on the upper end.

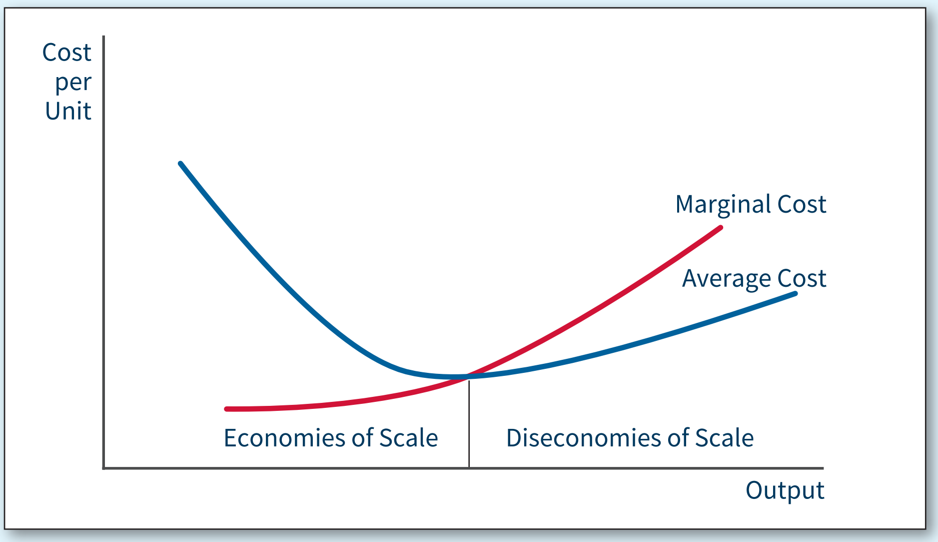
experiences of banks with strong and weak asset portfolios before and during the crisis suggests that asset strength plays an important role in the emergence of scale economies throughout the industry.

Our analysis focuses on community banks—banks with less than $10 billion in assets—as these banks comprise the vast majority of banking organizations. Approximately 97 percent of all banks in the United States have less than $10 billion in assets, and roughly 90 percent of those have less than $1 billion in assets. The consolidation trend in the industry has differentially affected community banks. The number of small institutions—those with less than $100 million in assets—has declined by 92 percent since 1985. Much of the debate about bank consolidation centers on the largest financial institutions, primarily those some argue are "too big to fail." But as consolidation in the industry has persisted in recent years, some have begun to turn the "too big to fail" designation on its head and question whether small community banks are "too small to succeed."[6] Conceptual arguments that support this notion are often based upon the economics of scale. Some

What Is a Cost Curve?

Cost curves are graphical representations of the relationship between a firm's output and its average total costs. Although total costs increase with production, a firm's average cost is typically expected to decrease before eventually increasing: that is, the average cost curve is U-shaped. In the short-run, fixed costs and diminishing marginal productivity dictate this shape; in the long-run, economies (and diseconomies) of scale determine the shape of the cost curve.

Marginal cost is the change in total cost from an additional unit of production. The marginal cost curve crosses the average cost curve at its minimum. Before this point of intersection, the marginal cost of production is below the average cost. This means that increases in output lower per-unit costs. Further increases in production raise per-unit costs, since the marginal cost exceeds the average cost.



Economists distinguish between short-run and long-run costs. Some inputs are fixed in the short run, whereas all inputs are variable in the long run. Thus, long-run costs are always lower than short-run costs.

The minimum point of the long-run average cost curve has important conceptual implications. Consider an industry composed of different-sized firms that produce an identical good from the same production technology. If firms are not at the minimum point, then society can produce the same number of goods using fewer resources by shifting production from firms operating beyond the minimum point to those operating at scales below the minimum point. An industry in which all firms are operating at the scale corresponding to the minimum point of the long-run average cost curve is using society's scarce resources efficiently and producing output at the lowest per-dollar cost.

[6] See, for example, Reckard (2013), Alloway (2015), Kamen (2010), and Schaeffer (2014).

have suggested that increased regulatory burden affects small banks in particular because regulatory compliance cost is a relatively larger item in a small bank's finances.[7] Likewise, banks that operate in limited geographical areas may find expansion into new product lines less profitable. Another possibility is that technological investments, for example in credit scoring and model-based lending, may not offer enough upside to justify the investment cost for small banks to transition from slower, more cost-intensive business practices (i.e., relationship lending).

Consolidation that shifts assets from small to large banks is more than just a rearrangement of resources. Small and large banks are not interchangeable; a single $1 trillion bank is not the same as one thousand $1 billion banks. Small banks are often built around a relationship-lending business model. Bankers acquire costly but valuable private information about their customers and make lending decisions using this expertise.[8] In contrast, large, remote banks often lack personal relationships with customers and knowledge about the local community, instead relying on a standardized approach to lending.[9] Customers that are good credit risks to a small bank may be unable to obtain credit from a large bank that lacks local knowledge.

Our analysis of scale economies within banking has implications for the future of the industry. Economies of scale can lead to consolidation within an industry as smaller firms have difficulty competing with larger and, therefore, more efficient institutions. Although the forces prompting consolidation are subject to debate, consolidation within the industry has been widely observed for the past three decades. Indeed, there were more than 18,000 insured institutions in the 1980s compared with approximately 5,100 today. This consolidation has been fairly consistent over time, averaging around 4 percent per year, but its effects across the size distribution of banks are uneven.

As the number of small banks has declined, concern about the future of small banks has extended to the future of small businesses. Small businesses generally obtain loans from small banks, especially when the businesses are in their infancy.[10] The report of findings from the FDIC's Small Business Lending Survey states that large banks are more than five times more likely than small banks to require minimum loan amounts for the primary loan products provided to small businesses and eight times more likely to use standardized small business loan products.[11] Small banks are also roughly five times more likely than large banks to underwrite loans to start-up small businesses differently.[12] These businesses are sometimes described as the engine of economic growth in the United States, so a decline in credit availability to such businesses could affect the real economy.[13]

The fate of small banks also portends that of the communities in which they operate: Kandrac (2014, p. 23) finds meaningful feedback from the failure of a bank and local economic performance, stating, "The disruption of banking and credit relationships is an important channel through which bank failures affect economic performance." Scale economies in banking thus transcend the domain of business policy into that of public policy.

The remainder of the paper proceeds as follows: we briefly describe prior research examining economies of scale in banking. We then introduce our empirical approach and describe the data used in the analysis. Next, we discuss the main findings from our nonparametric and parametric analyses. The final section offers conclusions based on those findings. An appendix describing translog cost estimation follows the references.

[7] See Grant (2012), Rapoport (2014), and Lux and Greene (2015, pp. 22–25).

[8] See Tarullo (2014b) and the references therein. Lee and Williams (2013) discuss research on community banks and lending.

[9] See Lux and Greene (2015, p. 2n) and Tarullo (2014b).

[10] Lux and Greene (2015, pp. 10-13) report that 77 percent of agricultural loans and more than 50 percent of small business loans come from community banks. DeYoung (2013, p. 50) states small businesses "typically rely on small banks for credit." See, also, Lee and Williams (2013), Tarullo (2014a p. 10), and Brennecke, Jacewitz, and Pogach (2020).

[11] FDIC (2018, p. 44).

[12] FDIC (2018, p. 46).

[13] See, for example, Neumark, Wall, and Zhang (2011), Ro (2013), and the webpage on "Jobs & the Economy: Putting America Back to Work," https://obamawhitehouse.archives.gov/economy/jobs. Labeling small businesses as the engine of economic growth can be traced back to, at least, Birch (1979). Claims that small businesses are the key to growth are contentious. Haltiwanger, Jarmin, and Miranda (2010) find that after controlling for age, there is no relationship between firm size and growth. They claim that new businesses are more important to employment growth than small businesses.

2 Related Literature

Our analysis builds upon previous work. Jacewitz and Kupiec (2012) find that community banks declined in efficiency relative to noncommunity banks from 1984 to 2011. Focusing only on community banks, as we do here, they find no indication of significant scale benefits beyond about $500 million in asset size for most lending specializations. DeYoung (2013) cites broad evidence showing substantial scale economies for banks with less than $500 million in assets. DeYoung (2013) also describes the history of empirical work in studying economies of scale in banking. The central theme is that of using increasingly complex econometric techniques that find evidence of economies of scale extending further into the size distribution.

Wheelock and Wilson (2018) consider scale economies in cost, revenue, and profit and find that the financial crisis had little impact on the returns to scale. They find that most banks faced increasing returns to scale in both 2006 and 2015. Davig, Kowalik, Morris, and Regehr (2015) find that post-crisis mergers have produced a more efficient and sounder banking system. Restrepo-Tobon and Kumbhakar (2015) estimate input distance functions and find scale economies to be economically small, while Kumar (2018) emphasizes the role of market power in studying scale economies. Anolli, Beccalli, and Borello (2015) and Pacelli and Pampurini (2016) look at scale economies for European banks.

DeYoung (2013) and Wheelock and Wilson (2018) take issue with the complicated econometric techniques used in recent studies. They question the conclusions about scale economies for the largest banks drawn from such estimation procedures. These criticisms prompted us to limit our analysis to banks with less than $10 billion in assets, where data are less sparse and banks are more comparable. Compared to some studies in which the largest banks have more than $2 trillion in assets, our sample focuses on banking institutions with similar business models. (This size cutoff still captures approximately 97 percent of banks.) In effect, we sacrifice the ability to comment on economies of scale at banks larger than $10 billion in assets for increased validity of our conclusions about smaller banks.

3 Empirical Approach

Measuring economies of scale at banks is far from straightforward. At a basic level, it is unclear what banks produce. Is the output of a bank best measured by the value of its assets, its loans and leases, or its deposits? Further, are a bank's costs best measured in terms of noninterest expenses, total expenses, or something else entirely? Finally, what are a bank's inputs? Although labor and physical capital are clear inputs for banks and industrial firms, should deposits be included as well? If not, should costs be limited to salary and physical capital expenses?

We remain agnostic about the answer to these questions and instead use several measures and multiple estimation approaches to address scale economies. We use both the nonparametric empirical approach of Jacewitz and Kupiec (2012) and parametric estimation techniques similar to the work of Wheelock and Wilson (2018). Our nonparametric approach uses assets to measure bank output, and either noninterest expense or the sum of interest expense, noninterest expense, and provisions to measure bank costs. For our parametric approach, we adopt a standard translog specification to estimate each bank's cost function. This approach views a bank's output as the value of its loans and leases. Deposits are included with labor and physical capital as inputs, and costs remain the sum of interest expense, noninterest expense, and provisions.

Our analysis assumes that the banks and thrifts in our sample produce homogeneous products so that we can compare costs across different institutions. This is likely a reasonable assumption in that a bank's loan portfolio can be produced by other banks, albeit at different cost. If loans across different lending categories are less comparable, we segment the analysis by lending specialization. This empirical approach assumes loans are comparable within lending specializations but not necessarily across specializations. These analyses assume loans by agricultural lenders, for example, are similar and can be made by other agricultural lenders but not by banks with other specializations such as mortgage lending.

4 Data

We use Call Report data from 2000 through 2019 to construct our data set. These data are supplemented by non-public safety-and-soundness examination results and FDIC-defined bank lending specialties. These data enable us to check the robustness of our main findings while providing additional insight about the role played by bank financial health.

Table 1 provides summary statistics for banks with less than $10 billion in assets for several years of interest.[14]

We perform our analysis at the holding company level. The vast majority of our institutions are single-bank holding companies. For holding companies with multiple certificate numbers, we sum certificate-level accounting variables to the regulatory high-holder. Lending specializations are defined by the FDIC's internal specialization group variable. Specializations are determined at the certificate level. For analyses performed on individual lending specialization, only certificates matching the lending category are included when aggregating to the holding company level. Finally, we divide banks into financially strong and weak categories based upon their CAMELS Asset Quality rating (CAMELS-A).[15] Holding companies in which the weighted average (by assets) CAMELS-A rating is below 2.5 are considered financially strong, while those with a rating of at least 2.5 are considered financially weak.

We place several restrictions on the data used in the analysis. First, our data include only banks and thrifts with less than $10 billion in assets. We exclude credit card institutions, as their business model is unlikely to be comparable to that of "traditional banks" that take deposits and make loans. We include banks that report positive amounts of total interest expense, total noninterest expense, total loans and leases, total interest income, and total noninterest income. We follow Jacewitz and Kupiec (2012) by limiting estimation samples to banks with total loans and leases of no more than twice their total deposits, which ensures that our sample is composed of traditional banks. Finally, to address outliers, we restrict analysis to community banks with costs less than one dollar per dollar of bank assets.

[14] These years correspond to the beginning of our sample period, the end of the run-up before the 2008 financial crisis, the crisis, the post-crisis economic recovery, and the end of our sample period, respectively.

[15] Banking supervisory regulators assign CAMELS ratings based on evaluations of a bank's managerial, operational, financial, and compliance performance. The six components of the ratings are capital adequacy, asset quality, management capability, earnings quantity and quality, the adequacy of liquidity, and sensitivity to market risk (CAMELS).

Table 1

Summary Statistics for Select Years

	Summary Statistics									
	2000		2006		2009		2014		2019	
	Asset Portfolio		Asset Portfolio		Asset Portfolio		Asset Portfolio		Asset Portfolio	
	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak	Strong	Weak
Assets	$255,983	$205,436	$345,287	$238,850	$317,899	$495,258	$487,365	$312,420	$606,197	$301,612
[thousands]	(685705)	(470181)	(767349)	(601890)	(790956)	(859416)	(996896)	(616545)	(1125859)	(757157)
Earned Interest Expense	9,365	8,881	8,896	6,835	4,806	9,130	1,985	1,640	5,201	2,887
[thousands]	(26136)	(22979)	(21869)	(22306)	(13133)	(16391)	(4984)	(4653)	(11778)	(7351)
Noninterest Expense	7,359	9,419	9,775	6,849	9,369	16,314	14,154	11,228	17,139	9,015
[thousands]	(21977)	(37680)	(23542)	(11336)	(25282)	(31849)	(34174)	(27564)	(38367)	(19270)
Provisions	629	3,510	561	869	1,633	10,315	598	1,040	976	1,459
[thousands]	(4438)	(23900)	(2609)	(2118)	(5900)	(28552)	(5558)	(9159)	(9772)	(7459)
Loans and Leases	165,344	141,437	237,600	174,686	204,214	347,063	322,582	208,281	428,768	215,532
[thousands]	(452199)	(351991)	(529616)	(509273)	(516752)	(588107)	(702007)	(422856)	(838559)	(550678)
Salary and Benefits	3,608	3,616	5,112	3,684	4,529	7,045	7,448	5,419	9,505	4,537
[thousands]	(9667)	(9461)	(10565)	(6151)	(10458)	(12266)	(16097)	(11876)	(17978)	(7549)
Full-Time Employees	82	79	89	64	73	112	101	78	106	62
[number]	(207)	(199)	(183)	(91)	(167)	(200)	(212)	(148)	(191)	(159)
Transactions Account Interest Expense	348	274	267	183	123	182	79	69	286	161
[thousands]	(697)	(501)	(677)	(330)	(280)	(547)	(243)	(341)	(967)	(488)
Savings Account Interest Expense	1,591	985	1,711	1,178	657	1,102	400	264	1,545	522
[thousands]	(4921)	(3108)	(4938)	(3532)	(2109)	(2156)	(1063)	(902)	(4569)	(1268)
Foreign Deposit Interest Expense	64	27	46	0	8	8	2	0	16	3
[thousands]	(2121)	(314)	(1448)	(0)	(258)	(291)	(73)	(0)	(520)	(35)
Interest-Bearing Deposits	171,485	140,552	230,839	164,503	217,018	348,963	310,846	212,944	388,857	201,027
[thousands]	(436295)	(297011)	(494319)	(397340)	(518538)	(576889)	(606054)	(387930)	(718210)	(480320)
Premises and Fixed Assets Expense	1,056	1,214	1,347	1,006	1,165	2,129	1,699	1,386	1,881	1,005
[thousands]	(2953)	(4462)	(3000)	(1577)	(2749)	(4284)	(3916)	(3249)	(3739)	(2133)
Premises and Fixed Assets	3,963	3,330	5,654	3,766	4,925	8,858	7,455	5,945	9,314	4,978
[thousands]	(10473)	(7550)	(12017)	(5365)	(11175)	(16241)	(15876)	(10963)	(17770)	(10607)
All Certs	10,021		8,784		8,104		6,587		5,252	
Mortgage	1,267		819		767		553		394	
Commercial	3,969		4,721		4,462		3,229		2,747	
Agriculture	1,977		1,634		1,568		1,515		1,291	

Source: FDIC.

Note: This table shows summary statistics at the holding company level for bank holding companies with total consolidated assets of less than $10B. Banks are classified as "strong" or "weak" based on a holding-company-level, asset-weighted average of the certificate-level asset quality component of the CAMELS rating. Banks are classified as strong if their weighted average asset quality rating is less than 2.5, and they are classified as weak otherwise. The bottom four rows of the table show aggregate counts at the certificate level before any sample restrictions are applied. Standard deviations are shown in parentheses below the relevant means.

5 Analysis and Discussion

5.1 Nonparametric Kernel Regression

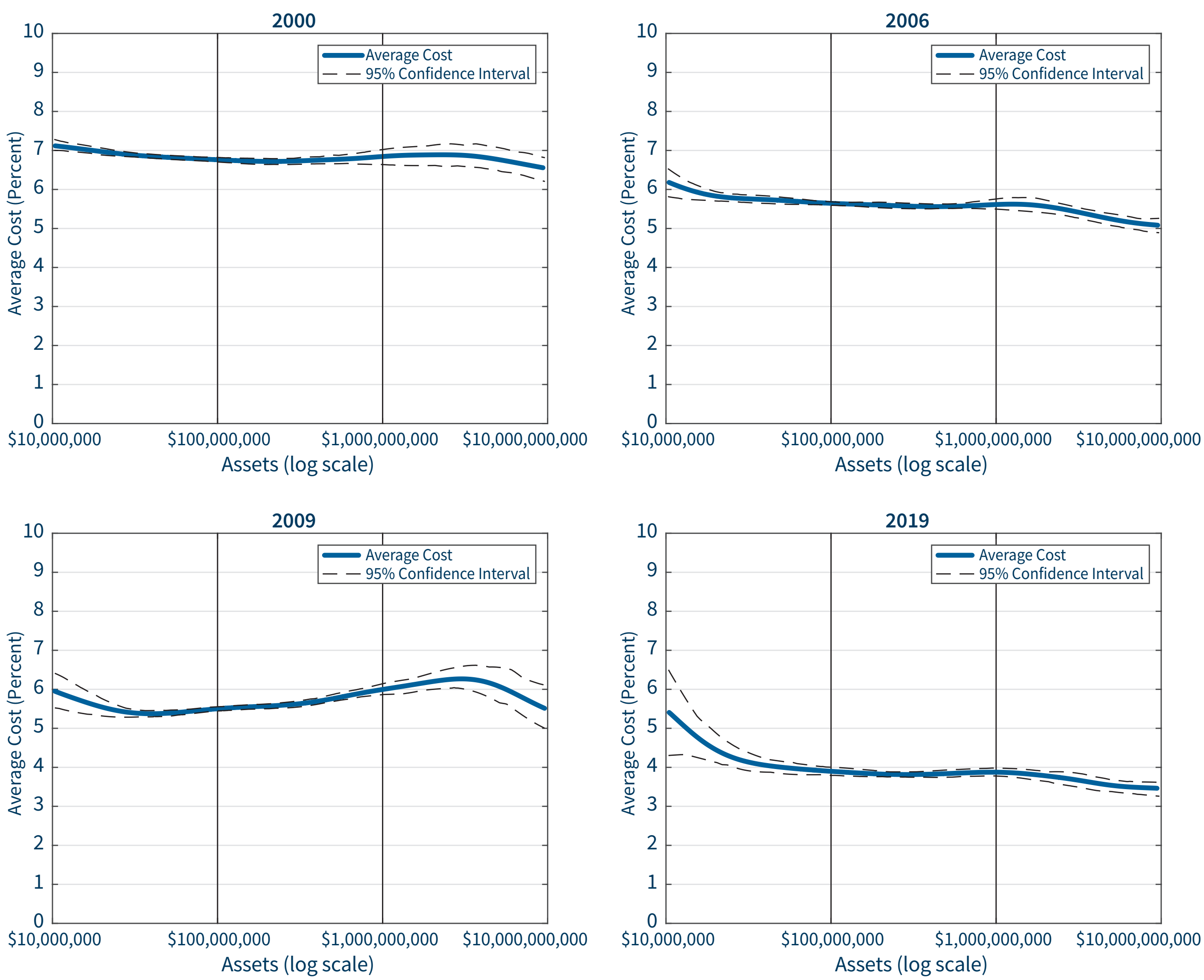
We begin our investigation into scale economies in banking with nonparametric analysis of the data. We measure a bank's output in terms of its total assets. Costs are measured in two ways: the sum of total interest expense (TIE), noninterest expense (NIE), and provisions for loan and lease losses (Provisions); and NIE alone. We refer to the sum of TIE, NIE and Provisions as "Total Costs." In both cases, costs are divided by assets so that they denote per-unit (of asset) costs.

Total Costs, measured as TIE+NIE+Provisions and expressed as a share of assets, have generally been declining for both large and small banks (Figure 1). No trend is apparent for NIE alone, suggesting that the temporal decline in costs is being driven by decreases in interest expense and provisions.

We estimate the cross-sectional regressions for select years corresponding to the start of our sample, the end of the run-up to the crisis, the crisis, the post-crisis recovery, and the end of our sample period. Costs increased during the crisis for large

Figure 1

Total Costs, All Banks, 2000–2019



Source: FDIC.

Nonparametric Kernel Regression

Nonparametric kernel regression is among the most flexible ways to estimate a relationship between two or more variables in the data. It imposes few assumptions and, instead, lets the data drive the estimated relationships. For the case of one explanatory variable (x), the nonparametric regression assigns a dependent variable (y) value at a particular point x_0 that is essentially a weighted average of y's at x's close to x_0. The weights are determined based on how far the neighboring x's are from x_0; that is, the weight assigned to a particular point (x_1, y_1) is typically a function of $|x_0 - x_1|$. That function is called the kernel and it is often chosen to resemble a probability density function. In our analysis, we use the Gaussian kernel with a fixed bandwidth. The bandwidth determines how sensitive the kernel function is to neighboring and distant observations.

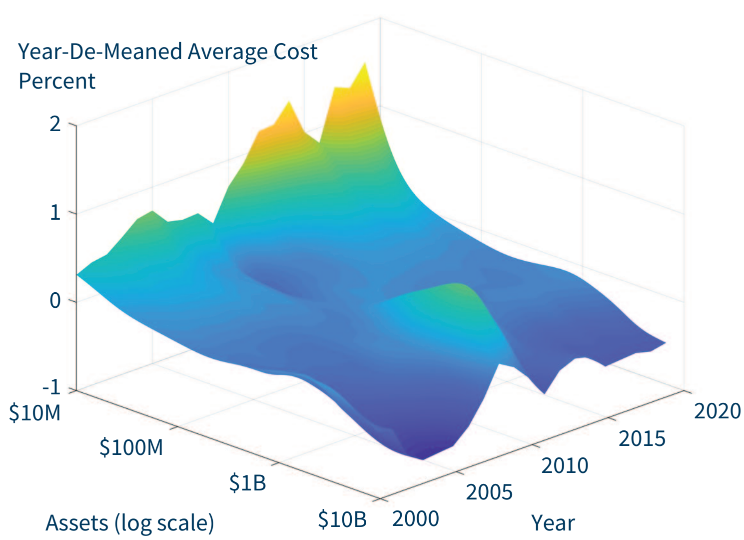
The main advantage of nonparametric regression is that it imposes only minimal assumptions (e.g., in the choice of a kernel function and a bandwidth) to reflect the relationships in the data. This flexibility does not come without a drawback: the estimated relationships may not satisfy certain regularity conditions implied by economic theory. The violations of regularity conditions may be due to either noise in the data or aspects of reality that are beyond the theory. To answer specific theoretical questions, such as studying movements in minimum points on average cost curves, it is often helpful to complement nonparametric analysis with a parametric approach that ensures the regularity conditions are satisfied, as we do in Section 5.2.

banks, and diseconomies of scale emerged for banks with assets in the $40 million to $600 million range. This increase in costs was short-lived, however, as banks reported a decrease in costs post-crisis (Figure 1). The decrease in costs—present across the board but more pronounced for large banks—suggests that a more efficient banking industry emerged from the financial crisis.

The increase in costs for large banks during the crisis is apparent in a three-dimensional plot depicting de-meaned average costs by year (Figure 2).[16] The declining trend in costs for large banks during the run-up to the crisis reverses direction during the crisis.

Figure 2

De-Meaned Total Costs, All Banks, 2000–2019



Source: FDIC.

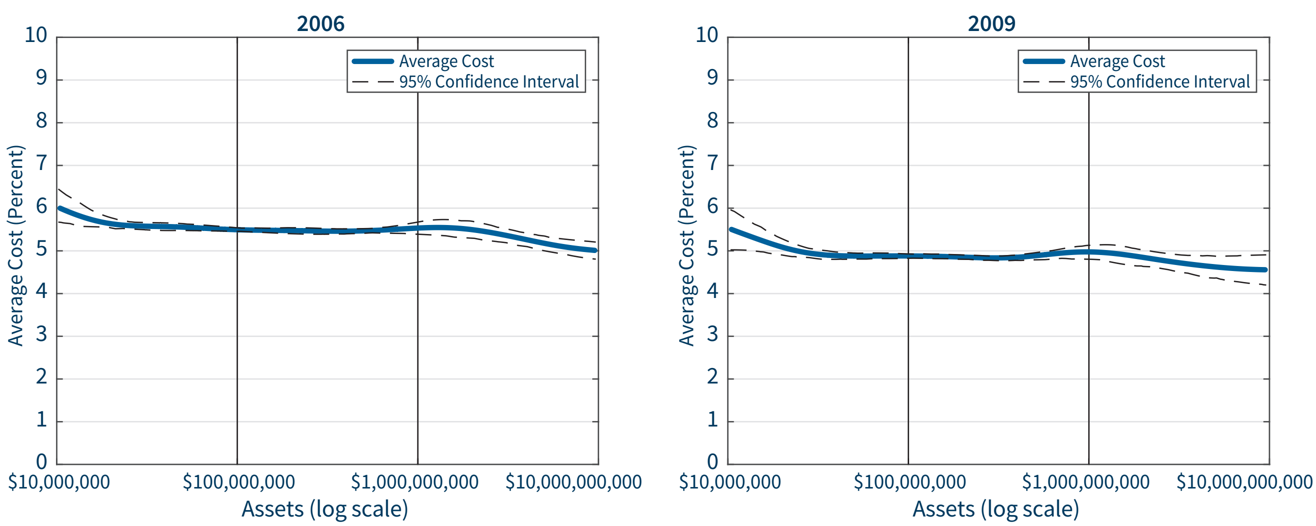
[16] In all three-dimensional plots, the mean of costs for each year is removed from that year's values to de-mean the data by year and highlight the changes in curvature.

Banks with weaker asset quality seem to have been more adversely affected by the crisis (Figures 3 and 4). Between 2006 and 2009, average costs generally decreased for banks with stronger-quality assets and increased for banks with weaker-quality assets.

Banks with weak asset portfolios exhibited marked economies of scale during the run-up to the crisis. Post-crisis, strong diseconomies of scale remained apparent for several years before returning in recent years to pre-crisis trends in economies of scale. This suggests that relatively large, weak banks grew by accumulating poor-quality assets. The conclusions are the same whether costs are measured by TIE+NIE+Provisions (Figure 4) or by NIE alone (not shown). These larger banks suffered considerably when the crisis began. As the distance from the crisis increased, banks on the other end of the size distribution—the smallest banks—saw their costs rise significantly.

Figure 3

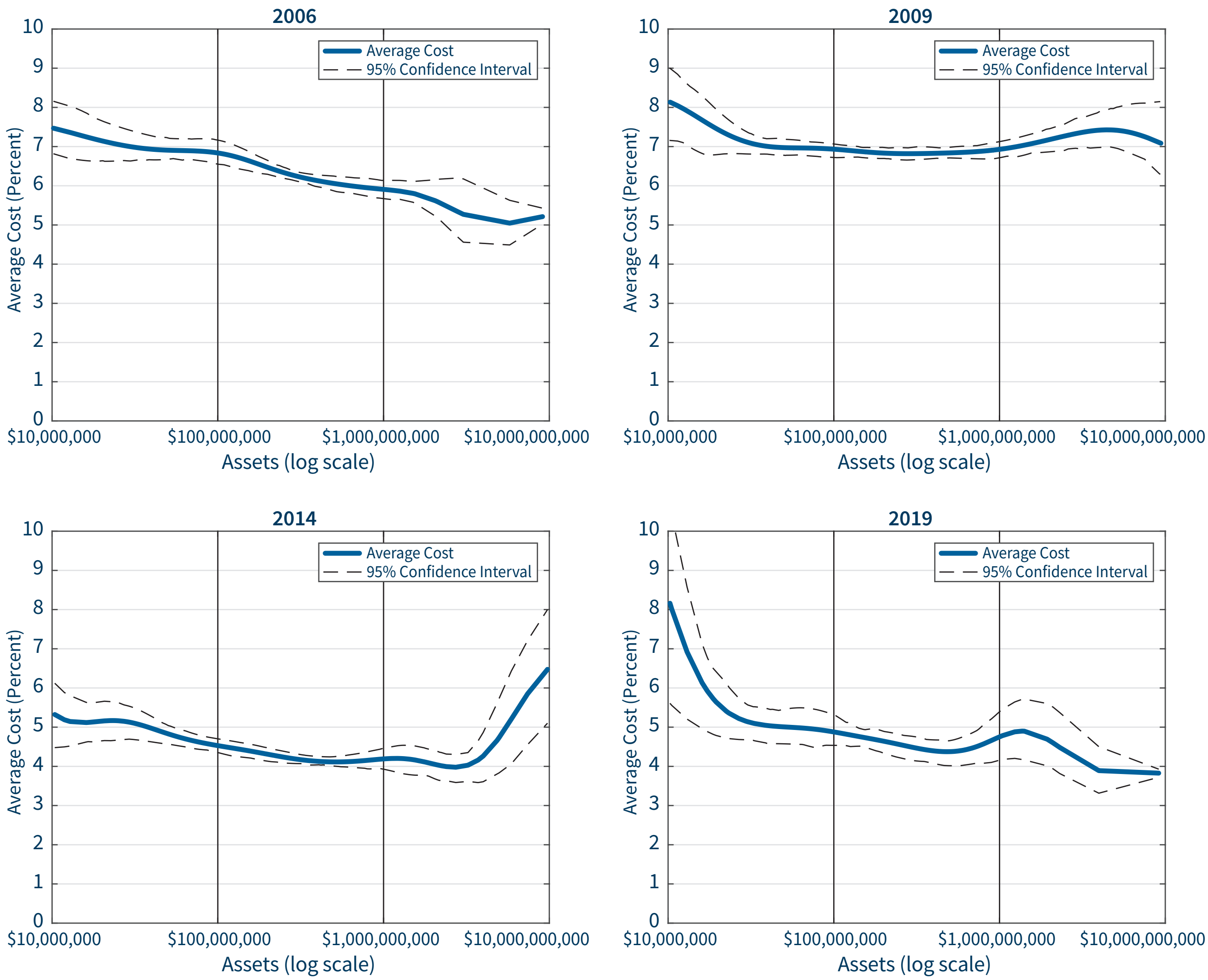
Total Costs, Strong Asset Quality Banks, 2006 and 2009



Source: FDIC.

Figure 4

Total Costs, Weak Asset Quality Banks, 2006–2019



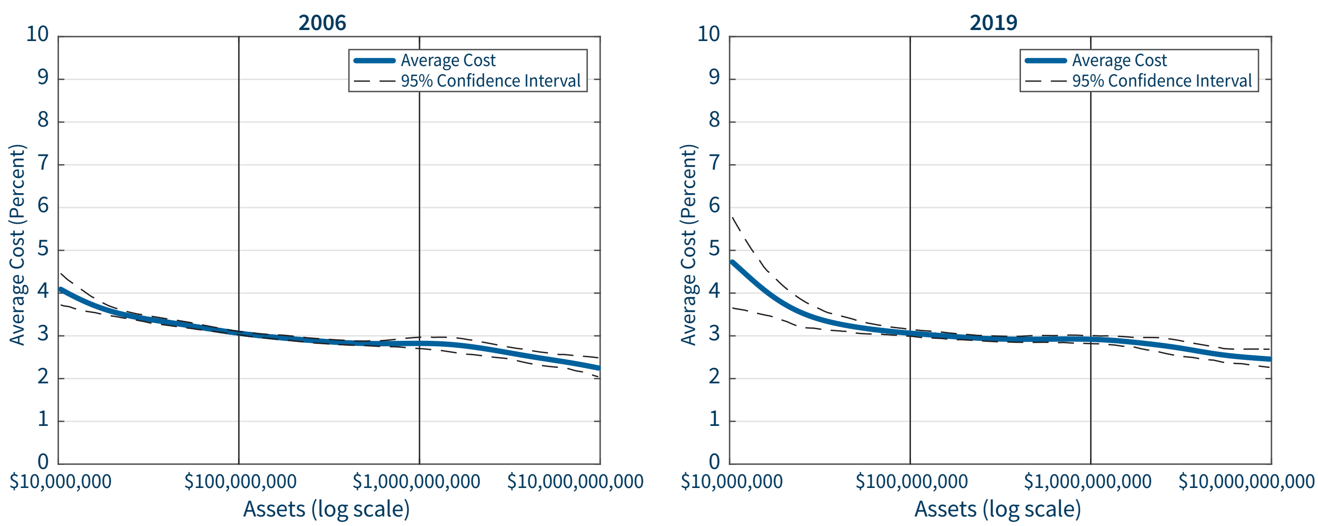
Source: FDIC.

Modest economies of scale are evident during the run-up to the crisis for NIE. These scale economies disappeared during the crisis and did not return until many years later (Figure 5).

Economies of scale and the relative stability of noninterest expense also are evident in a three-dimensional plot (Figure 6). The plot depicts average cost curves for each year after removing the yearly mean. The increase in economies of scale at large banks is clear during the run-up to the crisis.

Figure 5

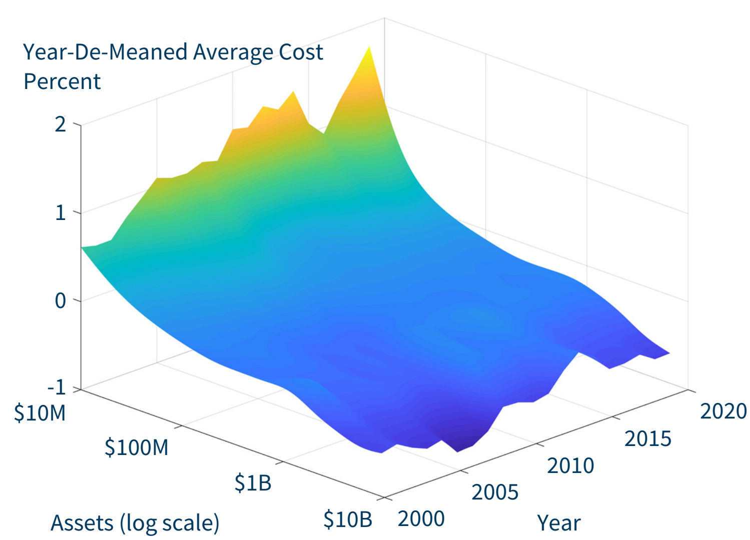
Noninterest Expense, All Banks, 2006 and 2019



Source: FDIC.

Figure 6

De-Meaned Noninterest Expense, All Banks, 2000–2019

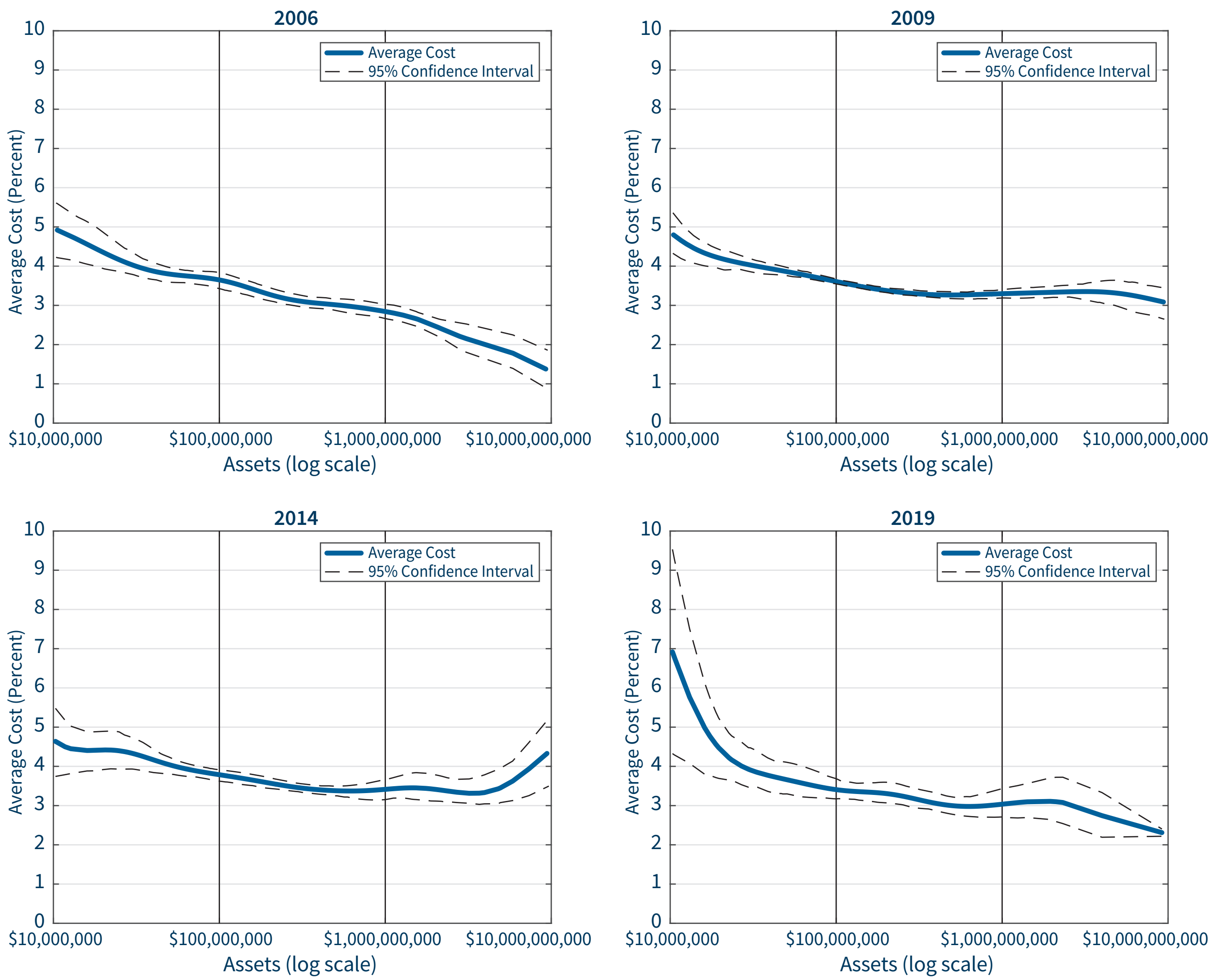


Source: FDIC.

Noninterest expenses are higher at banks with relatively weaker asset quality positions. NIE divided by assets remained between 3 percent and 4 percent for most banks over the sample period. Noninterest expenses appear considerably more variable at banks with poor asset quality ratings (Figure 7).

Figure 7

Noninterest Expense, Weak Asset Quality Banks, 2006–2019

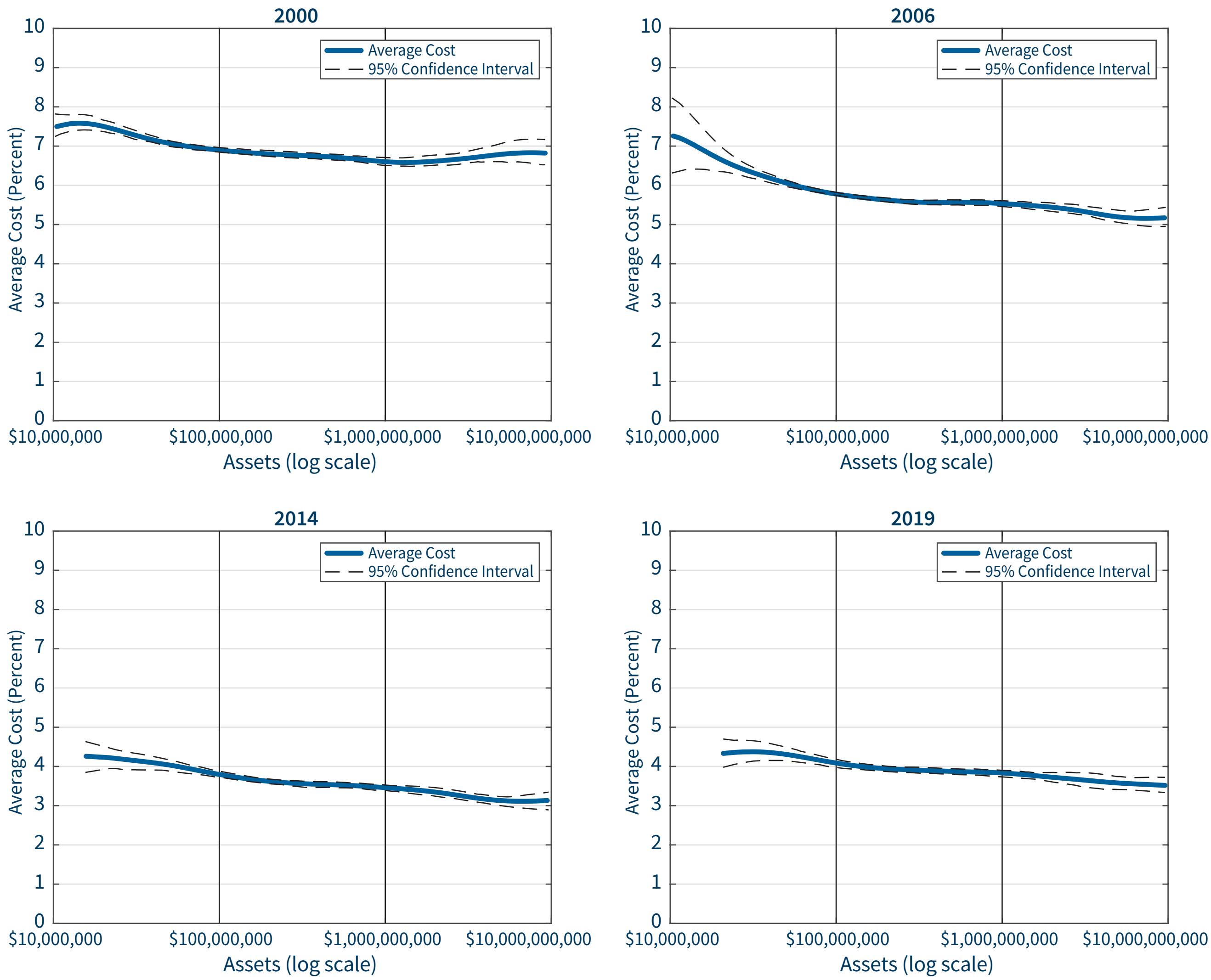


Source: FDIC.

Conducting the analysis separately for lending specializations underscores the heterogeneity of the industry's response to the crisis. During the run-up to the crisis, total costs for large commercial banks with strong asset portfolios declined considerably. This trend continued post-crisis before reverting slightly in recent years. Banks with less than $100 million in assets made up ground during the crisis and have seen comparatively smaller cost increases recently (Figure 8). NIE, conversely, shows almost no movement.

Figure 8

Total Costs, Strong Asset Quality Commercial Banks, 2000–2019

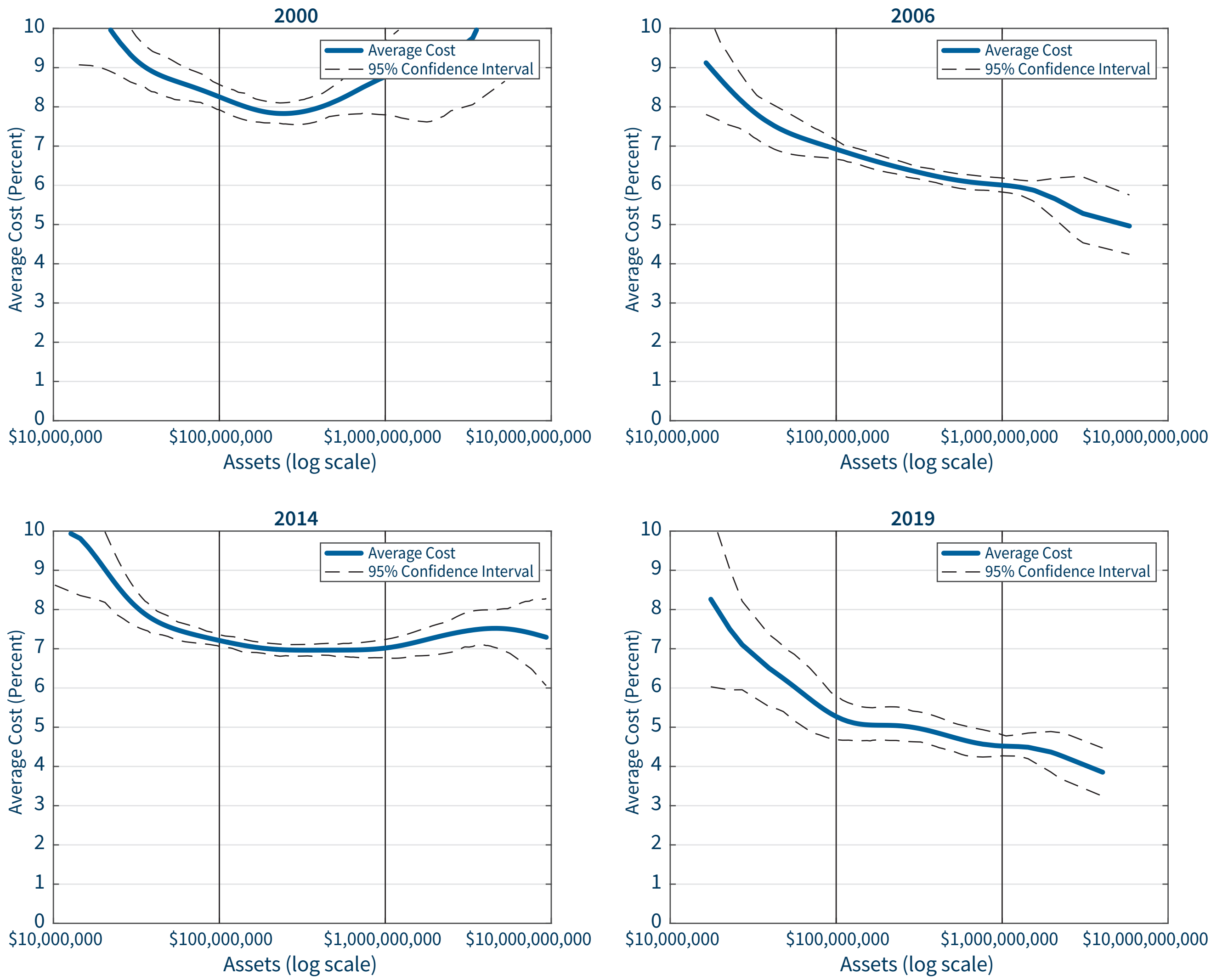


Source: FDIC.

Commercial banks with relatively weak asset portfolios displayed more extreme responses before, during, and after the crisis. Larger banks exhibited evidence of diseconomies of scale in 2000. However, large confidence intervals around the point estimates preclude drawing sharp conclusions. Economies of scale become clear immediately before the crisis. The crisis hit larger banks hard, and the scale economies at this end of the distribution disappeared and did not return until almost a decade post-crisis (Figure 9).

Figure 9

Total Costs, Weak Asset Quality Commercial Banks, 2000–2019

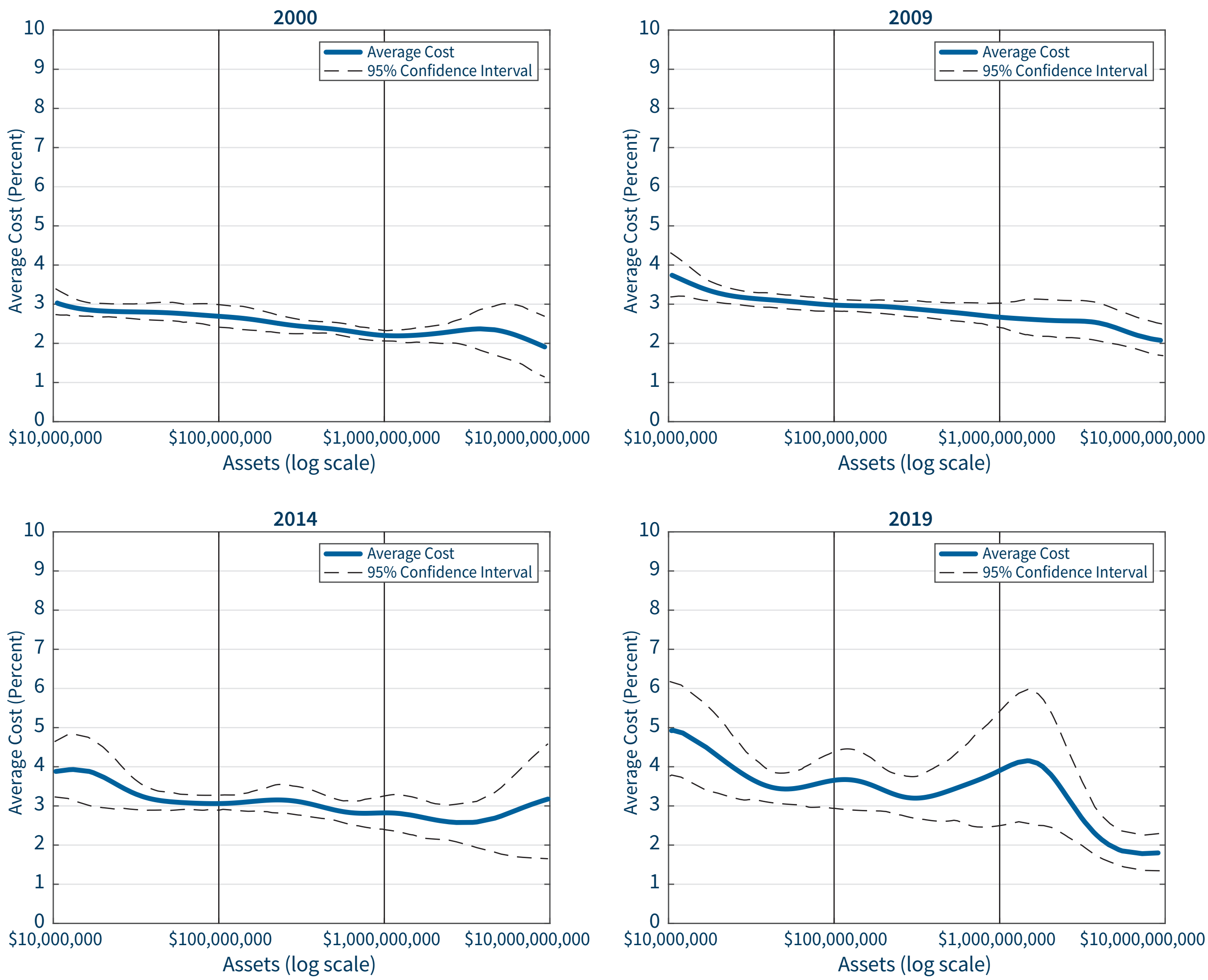


Source: FDIC.

NIE for mortgage banks was relatively stable for many years, as shown in Figure 10, but the mortgage lending industry seems to be undergoing a transition. Costs have recently increased and become more variable for mortgage banks with less than about $4 billion in total assets while lenders above this size have seen lower costs and less variability.

Figure 10

Noninterest Expense, Mortgage Banks, 2000–2019



Source: FDIC.

5.2 Parametric Translog Regression

Although nonparametric analysis enables us to bypass selecting the appropriate functional form for the cost function, it limits the conclusions we can draw. We therefore use parametric analysis based upon a translog specification of the cost function. The main advantage of such an approach is that we can obtain point estimates, however imprecise, of the efficient size for a bank over time.

The empirical approach estimates, for each bank and at each point in time, an average cost curve. This curve depicts the bank's average cost for all sizes. That is, points on the curve depict per-unit costs for any possible size of the bank. The bank's observed size-cost point is somewhere along the curve. For several reasons (e.g., adjustment costs and lags), it need not be at the exact minimum point of the cost curve.

According to economic theory, the estimated average cost curve will be U-shaped. The minimum point on the curve will be the most efficient size for the bank. Per-unit costs of production are minimized. As our empirical strategy estimates an average cost curve for each bank at each point in time, we construct an industry cost curve by using the median of the estimated costs curves at each point along a grid of output measures. The resulting curve can be thought of as depicting the cost curve of the industry as a whole.

We also plot the evolution of the cost-minimizing point within the banking industry over time. Constructing a plot of efficient bank size in this manner allows us to show confidence intervals around the estimated points.

As before, we treat total costs as TIE+NIE+Provisions. Inputs are labor, deposits, and physical capital, and output is total loans and leases. Estimating a cost function requires that we specify prices for the relevant inputs. The price of labor is salaries and employee benefits divided by the number of full-time employees. The price of deposits is calculated as the sum of interest expenses for transaction accounts and non-transaction savings accounts and foreign deposit interest expense divided by interest-bearing deposits. Premises and fixed asset expense divided by the sum of premises and fixed assets is used as the price of physical capital. These prices are obtained from Call Reports.

Parametric Translog Regression

A parametric approach produces results that more closely conform to theoretical frameworks and can be used to address theoretical questions with more precision. The drawback of a parametric approach is that theoretical assumptions fall short of capturing all aspects of reality, and conclusions drawn from such an approach may not hold in reality whenever the theoretical assumptions are not satisfied.

Translog cost estimation is a popular approach that assumes the cost function follows a translog specification of several input prices used in production. Cost functions can take many forms, even in theory, and the translog specification is a robust Taylor approximation of the cost function that can accommodate a variety of shapes. We impose several regularity conditions suggested in the literature (Christensen and Greene (1976), Serletis and Feng (2015), and Ryan and Wales (2000)) including homogeneity, positivity, monotonicity, and concavity.

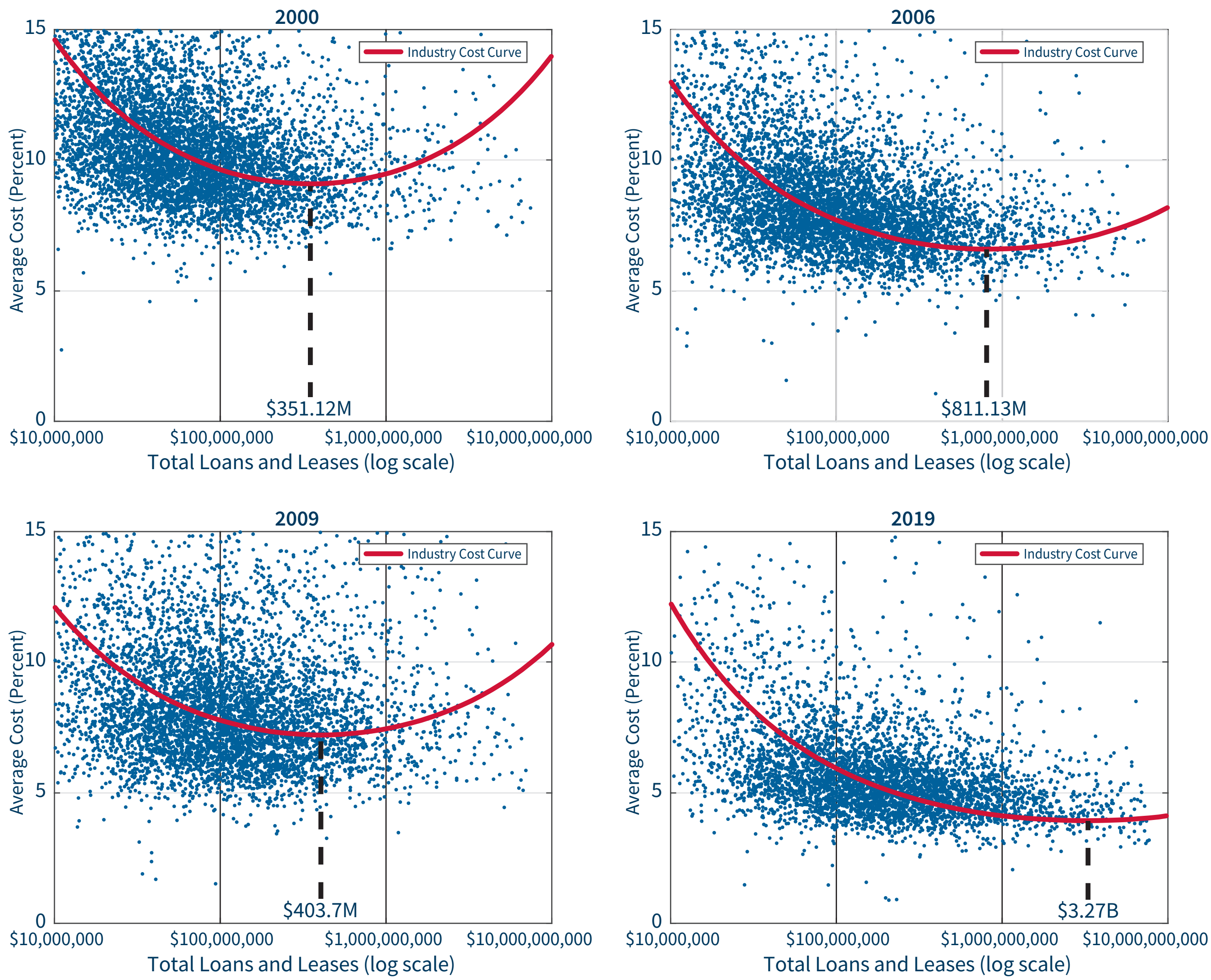
Translog cost estimation is motivated by production theory in which several inputs are used as factors in production of a certain output. This requires us to assume appropriate notions of "output" and "inputs" for banks. Based on earlier analysis in the literature (Gropper (1991)), we consider labor, physical capital, and deposits as inputs in production and total loans and leases as the output measure.

The translog specification is derived from a Taylor expansion of the cost function around the means of the data. As such, it is a poor choice when data are variable or skewed. (This is one of the criticisms of Wheelock and Wilson (2018).) We are able to avoid this problem by focusing on banks with less than $10 billion in assets.

The narrative seems similar for strong and weak banks (with the caveat that the data for weak banks seem somewhat unreliable). Costs decreased and scale economies grew during the run-up to the crisis. The estimated efficient bank size increased from about $350 million in loans and leases in 2000 to about $800 million by 2006. When the crisis began, large banks saw a sizable increase in their costs. Scale economies contracted and the estimated efficient bank size in 2009 decreased to approximately $400 million. Banks recovered post-crisis. The decrease in costs at large banks was enough to offset a smaller decrease in costs at small banks, and the estimated efficient size increased markedly to more than $2.5 billion by 2014, rising to approximately $3.3 billion today (Figure 11).

Figure 11

Translog Estimation, All Banks, 2000–2019



Source: FDIC.

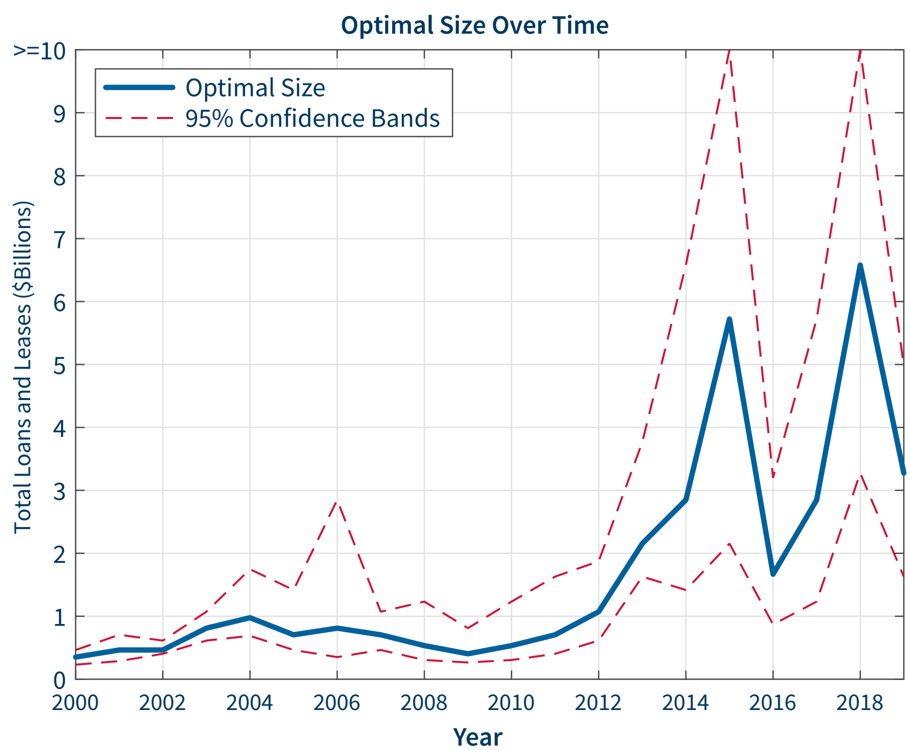
The cost-minimizing loan portfolio size is estimated to be $3.3 billion in 2019. Average costs for a bank of this size are approximately 3.9 percent, compared with 12.2 percent for a bank with a loan portfolio size of $10 million. Thus, we estimate that the difference in costs between a small bank at the end of our sample range and a bank operating at the efficient scale to be approximately 8.3 percentage points.

Although our estimates are only descriptive and not causal, they suggest that most of the cost savings that accrued while loan portfolio sizes increased are captured early in the growth process. Banks with loan portfolios of around $300 million have estimated costs of 4.76 percent, which means they have captured about 90 percent of any cost savings associated with increasing their loan portfolio size from $10 million to $3.3 billion. Banks with double the loan portfolio (around $600 million) have estimated costs of 4.33 percent and have accrued 95 percent of the potential cost savings from increased size.

The trend in efficient size can be seen in a graph depicting the evolution of the cost-minimizing loan portfolio size over time (Figure 12). Large, statistically significant increases in efficient size are only apparent post-crisis.

Figure 12

Cost-Minimizing Size, All Banks, 2000–2019

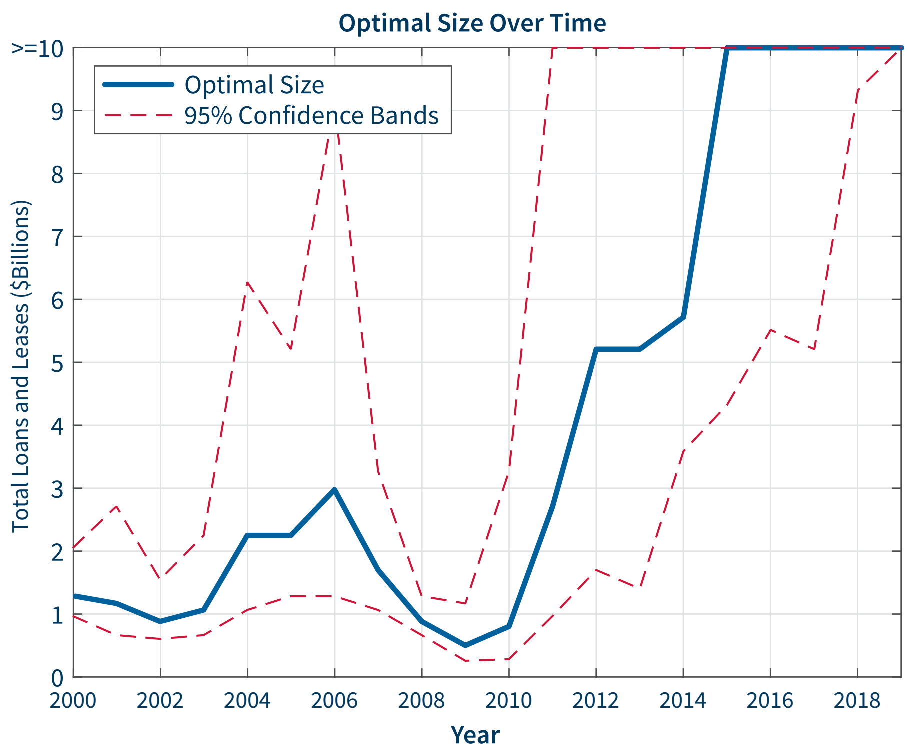


Source: FDIC.

Results are similar for commercial banks (Figure 13): the efficient size of a bank increased during the run-up to the crisis from about $1 billion in total loans at the start of the sample period to more than $2 billion by the onset of the crisis.[17] At the peak of the crisis in 2009, the efficient bank size is estimated to have been about $500 million. The estimated efficient size rose precipitously from this nadir, climbing to almost $6 billion by 2014 and reaching the high end of our truncated data sample shortly thereafter. Since our sample includes only banks with less than $10 billion in assets, we are only able to say that the estimated commercial banking industry cost curve was still declining by 2019.

Figure 13

Cost-Minimizing Size, Commercial Banks, 2000–2019



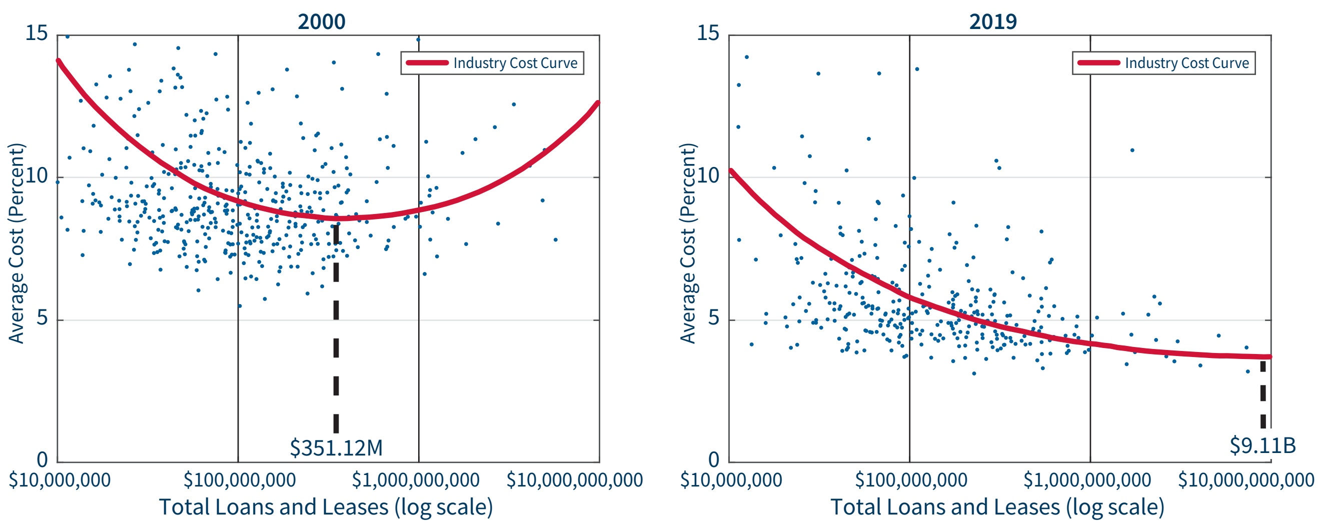
Source: FDIC.

[17] A bank is classified as a commercial bank if the sum of the following loan categories is more than 25 percent of assets: construction and land development loans, commercial and industrial loans, multifamily (five or more) residential properties secured by real estate loans, and non-farm nonresidential properties secured by real estate loans.

At the start of our sample period (2000), mortgage banks had an estimated cost-minimizing scale of about $350 million in total loans and leases.[18] This indicates that banks operating at scales greater than this amount are subject to decreasing returns to scale. This conclusion holds only briefly, however (Figure 14). By 2019, it is no longer clear that mortgage banks are subject to decreasing returns through the top of our data range ($10 billion in assets).

Figure 14

Translog Estimation, Mortgage Banks, 2000 and 2019



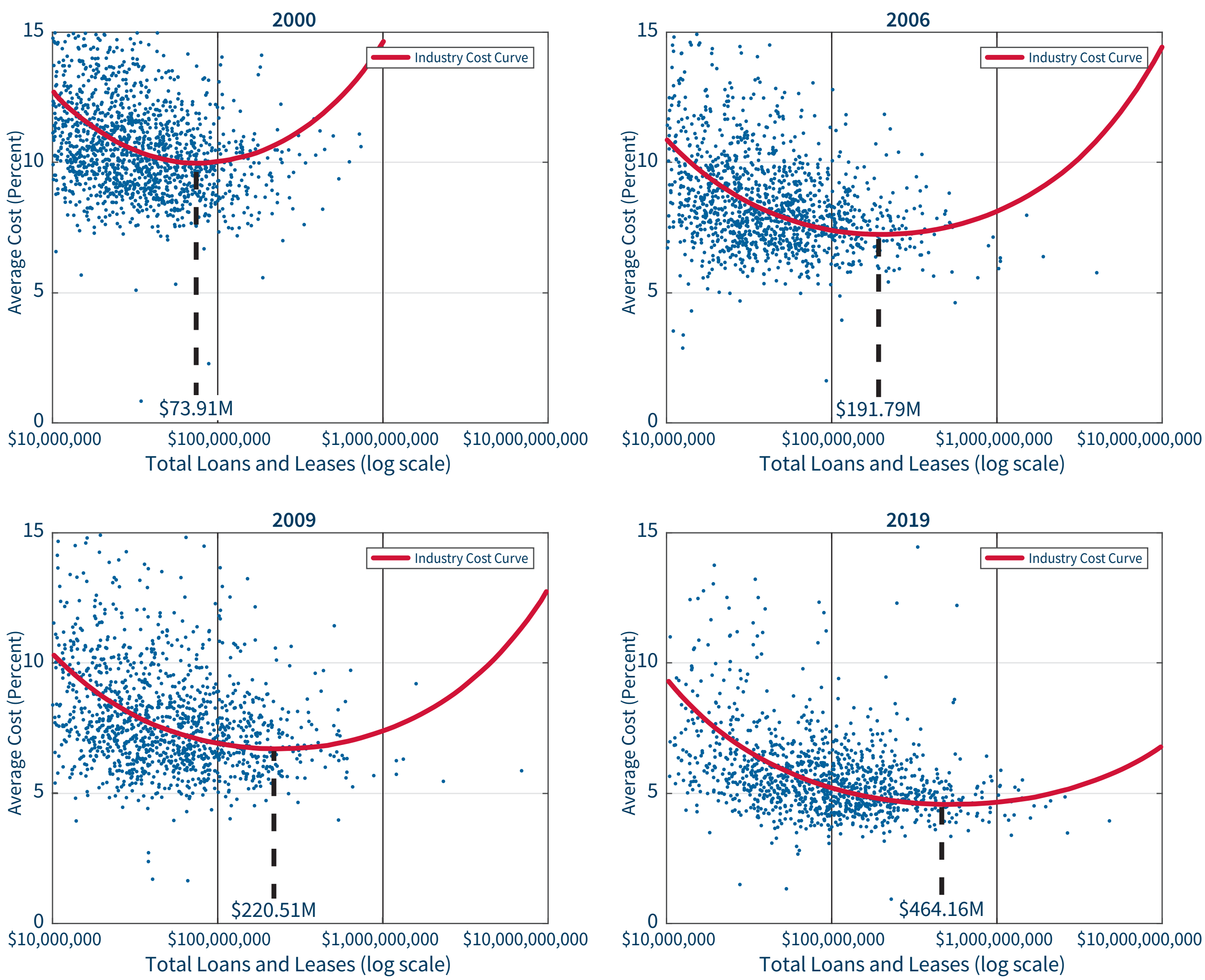
Source: FDIC.

[18] A bank is classified as a mortgage bank if the sum of the following asset classes is more than 50 percent of assets: residential mortgage-backed securities and loans secured by 1–4 family residential properties.

Large agriculture banks consistently showed decreasing returns to scale at a wide range of loan sizes throughout the sample period (Figure 15).[19] The cost-minimizing scale increased almost monotonically throughout much of our sample—it decreased slightly during the crisis—but it has remained relatively unchanged since 2014 at approximately $500 million.

Figure 15

Translog Estimation, Agriculture Banks, 2000–2019



Source: FDIC.

[19] A bank is classified as an agriculture bank if the sum of the following asset classes is more than 25 percent of total loans and leases: loans secured by farmland (including farm residential), loans to finance agricultural production and other loans to farmers.

6 Conclusion

Consolidation and growth have been hallmarks of the banking industry since the 1980s. The number of institutions has decreased by more than two-thirds while the size of the remaining institutions has increased. Although the problem of "too big to fail" has been frequently discussed within the corridors of government, academia, and the media, community bankers have begun to question if a "too small to succeed" problem also exists. Such concerns are commonly motivated by notions of economies of scale, whether due to cost efficiencies, expanded business opportunities, or the allocation of regulatory costs across a wider asset base.

Using financial and supervisory data on banks and thrifts with less than $10 billion in assets, we study economies of scale within the banking industry using nonparametric kernel regression and translog cost estimation. Our estimation period spans both sides of the financial crisis, enabling us to distinguish pre-crisis trends from post-crisis trends. We find that total costs have generally been declining over time. The crisis temporarily halted this trend, at least for some institutions, but the trend resumed in force post-crisis. With economies of scale, lending specializations matter: agriculture banks show less evidence of scale economies than commercial banks, while mortgage banks display the strongest signs of economies of scale.

Increases in the efficient bank size over time suggest an impetus for continued growth of comparatively small banks. We estimate that the cost-minimizing loan portfolio size for the industry as a whole rose from $350 million in 2000 to approximately $800 million in 2006. The efficient bank size fell to $400 million during the crisis before increasing markedly to $3.3 billion by 2019. We find evidence that almost all gains from increased size accrue early in the size distribution: by approximately $300 million in loan portfolio size, banks have achieved about 90 percent of the potential efficiencies estimated to occur by increasing in size from $10 million to $3.3 billion; by $600 million, they have achieved about 95 percent of potential efficiencies.

References

Alloway, Tracy, 2015, "Regulations Hit Smaller US Banks Hardest," *Financial Times*.

Anolli, Mario, Elena Beccalli, and Giuliana Borello, 2015, "Are European Banks Too Big? Evidence on Economies of Scale." *Journal of Banking & Finance* 58, 232–246.

Birch, David L., 1979, "The Job Generation Process," MIT Press.

Brennecke, Claire, Stefan Jacewitz, and Jon Pogach, 2020, "Shared Destinies? Small Banks and Small Business Consolidation," Working Paper 4, FDIC.

Christensen, Laurits R., and William H. Greene, 1976, "Economies of Scale in US Electric Power Generation." *The Journal of Political Economy* 84, 655–676.

Davig, Troy, Michal Kowalik, Charles Morris, and Kristen Regehr, 2015, "Bank Consolidation and Merger Activity Following the Crisis," *Economic Review* Q I, 31–49.

DeYoung, Robert, 2013, "Modelling in the Financial Services Industry," in Fotios Pasiouras, ed., *Efficiency and Productivity Growth*, chapter 3 (John Wiley & Sons, Ltd).

FDIC, 2018, Small Business Lending Survey, https://www.fdic.gov/bank/historical/sbls/full-survey.pdf.

Grant, William B., 2012, Testimony before United States House of Representatives, Testimony of William B. Grant on Behalf of the American Bankers Association before the Subcommittee on Financial Institutions and Consumer Credit.

Gropper, Daniel M., 1991, "An Empirical Investigation of Changes in Scale Economies for the Commercial Banking Firm, 1979–1986," *Journal of Money, Credit and Banking* 23, 718–727.

Haltiwanger, John C., Ron S. Jarmin, and Javier Miranda, 2010, "Who Creates Jobs? Small vs. Large vs. Young." Working paper 16300, National Bureau of Economic Research.

Hogan, Thomas L., and Scott Burns, 2019, “Has Dodd-Frank Affected Bank Expenses?” *Journal of Regulatory Economics* 55, 214–236.

Jacewitz, Stefan, and Paul Kupiec, 2012, “Community Bank Efficiency and Economies of Scale,” Regulatory Report, Federal Deposit Insurance Corporation.

Kamen, Ken, 2010, “Too Big To Fail, Too Small To Succeed,” *Forbes*.

Kandrac, John, 2014, “Bank Failure, Relationship Lending, and Local Economic Performance,” Working Paper 41, Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series.

Kumar, Pradeep, 2018, “Market Power and Cost Efficiencies in Banking,” *International Journal of Industrial Organization* 57, 175–223.

Lee, Yan Y., and Smith Williams, 2013, “Do Community Banks Play a Role in New Firms’ Access to Credit?” Updated 2014.

Lux, Marshall, and Robert Greene, 2015, “The State and Fate of Community Banking,” Working Paper 37, Harvard.

McCord, R., and E. S. Prescott, 2014, “The Financial Crisis, the Collapse of the Banking Industry, and Changes in the Size and Distribution of Banks,” *Economic Quarterly* 100, 23–50.

Neumark, David, Brandon Wall, and Junfu Zhang, 2011, “Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series,” *Review of Economics and Statistics* 93, 16–29.

Office, Government Accountability, 2015, *Dodd-Frank Regulations: Impacts on Community Banks, Credit Unions and Systemically Important Institutions* (Government Accountability Office, Washington DC).

Pacelli, Vincenzo, and Francesca Pampurini, 2016, “An Analysis of Efficiency and Scale Economies of the European Banking Groups During the Crisis,” *Bancaria* 72, 26–52.

Rapoport, Michael, 2014, “Small Banks Look to Sell as Rules Bite,” *Wall Street Journal*.

Reckard, Scott E., 2013, “Forget Too Big To Fail: Some Banks Now Too Small To Succeed,” *Los Angeles Times*.

Restrepo-Tobon, Diego, and Subal C. Kumbhakar, 2015, “Nonparametric Estimation of Returns to Scale Using Input Distance Functions: An Application to Large U.S. Banks,” *Empirical Economics* 48, 143–168.

Ro, Sam, 2013, “The Decline of America’s Job-Creating Small Businesses,” *Business Insider*.

Ryan, David L., and Terence J. Wales, 2000, “Imposing Local Concavity in the Translog and Generalized Leontief Cost Functions,” *Economics Letters* 67, 253–260.

Schaeffer, Brad, 2014, “The Dodd-Frank Effect: ‘Too Small to Succeed’,” *Wall Street Journal*.

Serletis, Apostolos, and Guohua Feng, 2015, “Imposing Theoretical Regularity on Flexible Functional Forms,” *Econometric Reviews* 34, 198–227.

Tarullo, Daniel K., 2014a, “A Tiered Approach to Regulation and Supervision of Community Banks,” speech delivered at the Federal Reserve Bank of Chicago Community Bankers Symposium, Chicago, November 7.

Tarullo, Daniel K., 2014b, “Rethinking the Aims of Prudential Regulation,” speech delivered at the Federal Reserve Bank of Chicago Bank Structure Conference, Chicago, May 8.

Wheelock, David C., and Paul W. Wilson, 2018, “The Evolution of Scale Economies in US Banking,” *Journal of Applied Econometrics* 33, 16–28.

Appendix: Translog Estimation

The general form of the translog cost function is as follows (see Christensen and Greene (1976) and Gropper (1991)):

$$\ln \mathrm{C} = \alpha_0 + \alpha_1 \ln Y + \frac{1}{2}\sigma(\ln Y)^2 + \sum_i \beta_i \ln P_i + \frac{1}{2}\sum_i \sum_j \delta_{ij} \ln P_i \ln P_j + \sum_i \tau_i \ln Y \ln P_i \tag{1}$$

where C is the cost measure, Y is the output measure, P_i is the price of input $i \in \{1, \ldots, \mathrm{m}\}$, and m is the number of inputs. The parameters in equation (1) can be estimated by Ordinary Least Squares (OLS), but the resulting cost function may violate theoretical regularity conditions. We incorporate several theoretical regularity conditions in the estimation. Such conditions impose constraints on the specification, and equation (1) is then estimated by constrained optimization where the objective function is the sum of squares of the difference between the two sides of equation (1). The following regularity conditions were imposed.

Homogeneity of Degree 1

Homogeneity of degree 1 is satisfied if the following constraints hold (Christensen and Greene (1976):

$$\sum_i \beta_i = 1$$

$$\sum_i \tau_i = 0$$

$$\sum_i \delta_{ij} = \sum_j \delta_{ij} = \sum_i \sum_j \delta_{ij} = 0$$

Positivity

Positivity of the estimated cost is automatically satisfied in the translog specification because the log of cost is used as the dependent variable.

Monotonicity

Monotonicity requires that cost increases when the price of an input increases. It is imposed by ensuring the non-negativity of each input's "share" (see Christensen and Greene (1976) and Gropper (1991)):

$$\frac{\partial \ln C}{\partial \ln P_i} = \beta_i + \sum_j \delta_{ij} \ln P_j + \tau_i \ln Y \tag{2}$$

Because the monotonicity constraints depend on an observation's input and output values (P_j and Y), there will be a set of m constraints for each observation. The non-negativity of shares is imposed for one bank in the data.

Concavity

Theoretical regularity also specifies that the cost function is concave in prices. That is, the Hessian matrix of the cost function with respect to prices is negative semidefinite (Serletis and Feng (2015)). Following Serletis and Feng (2015), this is true if the following matrix is negative semidefinite:

$$\mathrm{G} = \mathrm{B} - \underline{\mathrm{s}} + \mathrm{ss}' \tag{3}$$

where B is a matrix with element $\mathrm{B}_{ij} = \delta_{ij}$, $\underline{\mathrm{s}}$ is a diagonal matrix with share i from equation (2) on diagonal element i, and s is a column vector of shares with share i from equation (2) in element i. As noted in Serletis and Feng (2015), a necessary and sufficient condition for negative semidefiniteness of G is that all of its eigenvalues are nonpositive. Matrix G can be evaluated at every observation because the shares (from equation (2)) are observation-dependent. Following Ryan and Wales (2000), we impose concavity at only one observation. This approach often results in concavity being satisfied at most other observations and avoids imposing concavity globally (which can severely reduce the flexibility of the translog functional form).