Missing Output, User Cost Scale Economies, and Incentives to Merge.

David Humphrey
Professor of Finance (Retired)
Florida State University
Tallahassee, FL, U.S.A

Abstract

Spreading costs over a larger output is an important incentive to merge. Many banking scale studies may underestimate these benefits. We compute user cost scale economies, an approach consistent with how banking is treated in the national accounts. Banking outputs missing from academic studies are included and output is measured as value added. Our results are compared with a standard approach to scale measurement using balance sheet values. Scale effects for five major banking services are also noted. User cost scale economies are more closely associated with underlying costs by including missing output and reducing collinearity among service line outputs. This improves local identification of bank scale effects. (108 words)

Key Words: scale economies, bank output, bank mergers

JEL Classification Code: G23, D24, L25

1. Introduction

Spreading excess capital and labor over a larger output is the standard interpretation of scale economies. Most banks only grow as fast as their deposit and loan base so normal annual growth over a number of years can typically be accommodated with current factor input excess capacity. With a merger or acquisition (M&A), however, all factor inputs are suddenly variable. Here local or non-local consolidation of certain activities is implemented to reduce the excess capacity that was acquired.

Academic estimates of scale economies are often determined from a cost function where reported bank total cost is related to the balance sheet values of loan and securities services plus input prices. This approach has been used in Mester (1993); Berger, Leusner, and Mingo (1997); DeYoung, Hasan, Kirchhoff (1998); Berger and Mester (2003); Humphrey and Vale (2004); Feng and Serletis (2009); Davies and Tracey (2014) and many others. As the cost of these asset outputs is not separately reported, total cost--interest expense plus noninterest expense (composed of labor and physical capital cost)--is used instead. This specification excludes outputs from the liability side of the balance sheet even though their cost is included in the dependent variable.

The most important excluded output concerns deposit services (payments, cash acquisition, and insured savings). Deposit services use more labor and physical capital inputs (branches, ATMs) than all types of loans and securities together. Another (usually) excluded output concerns noninterest income activities (investment banking, brokerage, insurance, trading, etc.). With a complete specification of banking outputs, predicted average costs are shown to be importantly lower than those reported in the literature.

Computing user cost scale economies suggests that scale benefits from M&A activity can be larger than what many current studies show. The approach is based on the how banking is treated in the national accounts. The national accounts include deposit services and noninterest income activities that are missing from the academic studies cited above. Bank output is measured as value added, and is the standard measure of output for other industries in the national accounts.

In what follows, the importance of M&A activity in contributing to bank asset size is outlined in Section 2. Section 3 outlines our critique of the standard cost function model used to estimate bank scale economies. A different approach relies on user cost which is explained in Section 4. Scale economies using a standard approach are estimated in Section 5 and contrasted with results using all banking outputs to generate user cost scale economies. Separate scale estimates are also noted for five types of banking services, as well as how monetary policy can affect these estimates. Conclusions are in Section 6.

2. Mergers and Acquisitions: The Main Way Banks Grow

An old rule of thumb used by banking consultants is that generally 30% of an acquired bank's noninterest expense (capital and labor) may be saved in a local merger or acquisition. Given that operating expense is 45% of total cost (when interest rates were higher) and the size of the average acquired bank is about 30% of the post-merger consolidated bank, this suggested that savings from a merger may amount to 3% of the consolidated bank's total costs. Cost savings are expected to come from closing overlapping/underperforming branch offices and reducing staff.

A standard estimate of U.S. bank scale economies evaluated at the mean of the data in the past would be around .95. Doubling a bank's size, which typically occurs with from one to three mergers or acquisitions, is thus expected to increase incremental costs on average by only 95%. The main way banks become larger is through a merger or acquisition. For example, over 2001 to 2018 there were 4,284 unassisted mergers/acquisitions according to FDIC data. Total banking assets grew by 155%, profits rose by 200%, but the number of banks fell by 42%. This is not unusual. Between 1980 and 1998 there were 8,091 mergers and acquisitions and the number of banks fell by 39% (Rhoades, 2000). Over this period of bank consolidation, the number of banks fell but the number of branches doubled to over 77 thousand by 2018. Unfortunately, close to half of mergers that have been studied do not yield significant cost reductions. The most frequent problem encountered is the unexpected difficulty in integrating legacy data processing systems and operations (Rhoades, 1998). These difficulties can offset gains achieved by closing overlapping offices and reductions in staff.

Smaller banks experience the greatest scale benefit of a merger and, in fact, this is where most mergers occur.[1] Scale economies by bank service line, which is computed below, indicate which services may be contributing the most to lower costs as output expands. Of the 5,406 banks in 2018, nine banks had > $250 billion in assets and account for 49% of all banking assets. But these nine institutions represent less than 1% of all banks. Banks with > $10 billion in assets comprise 2.5% of all banks but hold 83% of all assets. The remaining 5,269 banks hold only 17% of all assets but account for 97.5% of all banks. Clearly, large banks **are** the banking system. This justifies the research focus on scale economies at large banks in particular.

3. Two Related Problems.

3.1 A Standard Approach to Scale Measurement.

A standard total cost (TC) function model used to estimate bank scale economies is expressed as: lnTC = f (lnAn,lnPi) where Anrepresents different asset balance sheet categories of loans and securities outputs while Pireflects the prices of factor and funding inputs. The problem is that costs specific to the various Anasset outputs are not reported and total cost is typically used instead. But TC is the sum of interest expense and noninterest expense. And a large share of the labor and physical capital costs in noninterest expense is not associated with any of the Anoutputs. Cost accounting information would show that deposit services use a large share of bank FTE labor and most of the space in each branch office to: open/close accounts (around 8% of accounts turn over each year), provide payment services (over 30 billion card, check, ACH transactions in 2018), accept physical check/cash deposits, and provide convenient access to cash at branch offices and ATMs. Thus lnTC should instead be specified as ln(TCx θn) where θnis the share of total cost that refers only to the Anoutputs. As θn is unknown, the scale economies associated with only Anare potentially biased.

3.2 Both sides of the Balance Sheet, Not Just Half of It.

Academic banking studies have long neglected the output associated with the liability side of the balance sheet, viewing it solely as a funding input. However, payment services associated with deposits are an important banking output. Broadly defined, they are estimated to generate between one-third to two-fifths of bank operating revenues for 25 of the largest U.S. bank holding companies (Radecki, 1999).

[1]As shown in DeYoung (2019), Figure 31.5, the number of banks fell in all size classes except for those with assets between $500 million to $2 billion. This was primarily due to mergers among banks with less than $500 million in assets crossing the $500 million threshold but not the $2 billion threshold.

After much academic and government economist debate, it was concluded that banking is more than a "pass through" of deposit and other funding for making loans, as it was treated earlier (Triplett, 1992). Deposit services are now included as a banking output and banking output is measured as value added (using user cost).

The Bureau of Economic Analysis (BEA) and national accounts statisticians in other countries have altered the treatment of banking and now include bank services from both sides of the balance sheet. The much-cited Sealey and Lindley (1977) paper is correct that deposits are an intermediate input to the production of loans. They just did not go far enough and neglected the output component of deposits.

A bank uses deposits to fund loans and depositors use their deposits for transactions and safekeeping.[2] The price of a loan is set to cover the funding cost of deposits (the input funding side of deposits). Thus the loan price includes the cost of funding and need not be separately specified on the RHS of a cost function as usually occurs. The loan price also includes the labor/capital production cost of a loan, the expected credit risk (covering possible losses), and a portion of the return to equity capital. Borrowers pay for the funding cost of deposits which is a "pass through" expense paid to depositors at the discretion of the bank. Since 2003, deposit services and noninterest income activities have been included, along with asset services (loans and securities) in measuring the contribution of the banking sector to the national accounts (Fixler, Reinsdorf, and Smith, 2003).

At the individual bank level, it would seem that the balance sheet value of liability outputs (L_j) could simply be added to the standard cost function specification giving lnTC = f(lnAn,lnLj, lnPi) where the derivative ∂lnTC/∂lnAn gives the scale effect of asset n and similarly for liability j (Lj). Simply adding the balance sheet value of liabilities to an asset-focused cost function is possible but is less workable due to the balance sheet constraint. Each $1 of balance sheet liabilities would be measured again as $1 of balance sheet assets. Total bank assets always equal total liabilities plus equity. The balance sheet value of demand deposits, savings, or time deposits would be reflected once as liabilities and a second time as assets in the same equation.[3]

For example, the intercorrelation of Lj with An in 2010 using balance sheet data for five services is higher in our sample (with an average r value of .94) versus a lower intercorrelation with the user cost approach (.83). Balance sheet data has greater multicollinearity than the user cost approach and thus is less likely to accurately locally identify scale economies in a cost function. While the user cost value of the various An and Lj banking outputs will be generally proportional to their value in the balance sheet, the proportions are not all the same.[4]

For these reasons, we use the value of user cost ($\ln V_{An}$, $\ln V_{Lj}$) in the specification lnTC = ($\ln V_{An}$, $\ln V_{Lj}$, lnPi)to determine the scale effects for five major banking services. The cost of funding loans via deposits, purchased funds, or equity is essentially a pass-through from borrowers who pay it to depositors, lenders of purchased funds, and equity holders who receive it, as suggested earlier by Sealey and Lindley (1977). Funding costs are already included in the user cost of loans V_{An} through the revenue received from setting the price of bank loans (and neednot be specified again as a funding price in lnPi). The same applies to the return on equity and a bank's assessment of expected loan risk. Thus, only the input prices of labor and capital are included in lnPi.

4. User Cost Applied to Banking.

The user cost of a financial asset (or liability) is an extension of a concept applied to a nonfinancial asset developed by Diewert (1974), Barnett (1978), and Donovan (1978). It has been applied to banking in Hancock (1985) and Fixler (1993) both using (now discontinued) Functional Cost Analysis information and by Fixler and Zieschang (1999) using Call Report data.[5]

---

[2]There is no need to "test" to see whether deposits are an input or an output (Hughes, Mester, and Moon, 2001). With user cost, deposits are both and both are included--not just one or the other.

[3]This is the same problem faced when trying to measure productivity with a sample of individual banks using a distance function (Humphrey, 2020).

[4]Wheelock and Wilson (2012, 2018) and Hughes and Mester (2013) add the balance sheet values of deposits (and other funding variables) to their cost function model. They clearly mean for them to be input quantities, not outputs. In balance sheet values they are the same. But if used as an output quantity, the cost-to-output assumption needed here would be a problem.

[5]Hancock computed user cost prices but used balance sheet data for bank output. Her purpose was to estimate price and output elasticities in response to changes in interest rates.

User cost is basically just adjusted revenue from various banking services. This adjusted revenue (output) is related to total cost in a cost function, holding constant labor and capital input prices. The cost of a banking service to its users is based on the (adjusted) revenue obtained from a service or activity. If a bank guesses wrong regarding credit or default risk, this reduces profits. It does not reduce Call Report loan revenues or user cost but does affect the reserve for loan losses. Users have already paid for expected losses in the price being charged. There is no need for users to pay for the credit risk of expected default (via the loan price) and again, after the fact, for changes in the reserve for loan losses.

The unit user cost (UC) of financial asset n is $UC_{An} = R_{An} - R_R$ where the monetary return on that asset is R_{An} and R_R is a risk-free interest rate (or reference rate) usually set equal to the short-term U.S. government Treasury security rate. The unit user cost for liability j (R_{Lj}) is determined similarly: $UC_{Lj} = R_R - R_{Lj}$. Borrower unit user cost is the observed loan borrowing rate for a private entity (R_{An}), minus the risk-free rate (R_R), so that only private credit risk and the cost of making a loan is reflected in unit cost (UC_{An}). Thus, unit user cost covers the funding cost of a loan, private credit risk, loan officer time, meeting space, obtaining credit assessment information, reserve for possible loan losses, mark-up for return on equity, and loan monitoring costs if the loan is made.[6]

The value of user cost (V_{An}), which is the value of the loan service to the borrower, is just $V_{An} = UC_{An} \times An$ where An is the value of the loan asset in the balance sheet. Thus the value user cost of a loan is considerably smaller than the balance sheet value of the loan. And while $1 of a real estate loan (An) always equals $1 of a consumer loan (Am), that is not the case for the value user cost of these two loans (V_{An} and V_{Am}) as their unit user costs (UC_{An} and UC_{Am}) differ. This is an important benefit since if balance sheet values of assets as well as liabilities were specified in the same cost function, local identification of their respective contributions to the dependent variable could be marred by a high degree of collinearity, but less so with the values of user cost in the same cost function.

The value of user cost for liability j is $V_{Lj} = UC_{Lj} \times Lj$. If this refers to deposit liabilities, this includes the greatest portion of branch and back-office staff costs, all of ATM network expenses, the cost of purchasing idle cash balances from the Federal Reserve for depositor withdrawal, the expense of making and receiving different types of payments to and from depositor accounts, daily deposit accounting, regulatory reporting, safekeeping expenses, and an unstated charge to provide a return to equity. The vast majority of these expenses are implicit. Banks could charge a price for all the separate deposit services noted above and pay interest to depositors at or just slightly above/below the risk-free rate. Instead, they pay a lower interest rate to depositors to cover most of their costs (an indirect price) and charge a direct fee only for selected depositor services.[7]

A key point is that depositors "produce" the funds placed in a deposit account, not the bank. With user cost, a bank's loan-deposit rate spread can be expressed as $SPREAD = R_{LOAN} - R_{DEPOSIT} = R_{LOAN} - R_R + R_R - R_{DEPOSIT}$. This shows that the rate spread is merely the sum of the user cost of loans and deposits together (with the proviso that R_{LOAN} and $R_{DEPOSIT}$ also include direct fees to reflect better user cost).

[6] When a bank purchases a U.S. Treasury security, a loan is made to the government. There is no need to assess credit risk, no charge is booked for expected loan losses, no creditworthiness monitoring is needed after a Treasury security is purchased, and no branch offices and staff are needed to deal with the borrower. However, these expenses apply to regular bank loans and represents the value added of the loan.

[7] The indirect pricing of deposit services typically makes the return paid to depositors (R_{Lj}) less than the reference rate (R_R). For convenience, unit user cost (UC_{Lj}) is expressed as a positive value: $UC_{Lj} = R_R - R_{Lj}$.

Table 1: Comparison of Average Cost and Balance Sheet Output Values, 2001-2005-2010

	Value User Cost		Value Balance Sheet	
Output Category:	$ Million	Share	$ Billion	Share
Business Loans	$148	30%	$6.1	29%
Consumer Loans	32	7	0.8	4
Securities Activity	112	23	3.5	16
Transaction & Savings Deposits	121	25	5.7	27
Time Deposits	74	15	5.2	24
Total	$487	100%	$23.3	100%

Table 1 compares the values and relative shares of five user cost banking outputs with their corresponding balance sheet values.[8] The average total value of user cost output over 2001, 2005, and 2010 was $487 million while the corresponding balance sheet value was almost 50 times larger at $23.3 billion. The shares of business loans, deposits, and consumer loans differ by, respectively, only 1, 2, and 3 percentage points. The largest differences occur for securities activities and time deposits, 7 and 9 percentage points, respectively. In our cost function estimation, all five user cost outputs are specified while the usual balance sheet approach specifies only the three asset-side outputs (covering 49% of the possible balance sheet outputs in Table 1).

Fee-based services (investment banking, fiduciary activities, brokerage, insurance, derivatives, etc.) as a group make up noninterest income activities. They have revenues but balance sheet values are either zero, commingled, or small and unrepresentative so an overall user cost price ($UC_{An} = R_{An} - R_R$) cannot reasonably be computed. Thus the value of user cost for noninterest income activities (V_{An}) used here is the reported revenue from these activities, minus deposit service charges (noninterest income) which are direct fees in the deposit function and affects deposit user cost. This is a variant of the approach used by Wheelock and Wilson (2018).

5. The Cost Function Model

Our five-user cost output, two input price Fourier functional form is shown below. This captures better differences in scale economies away from the mean of the data compared to the translog form. Our specification follows that in Berger, Leusner, and Mingo (1997) and DeYoung, Hasan, and Kirchhoff (1998). The translog form is shown below in the first two rows of the Fourier equation while the last three rows show the appended Fourier terms. These are a linear combination of the transformed outputs (lnVj) that are mutually orthogonal to one another (due to the cosine and sine terms).

[8]This corresponds to a bank with about $11 billion in assets over our three time periods. The table shows the average value of the five user cost outputs used in estimation. Only three of the five balance sheet outputs were used in the standard three output estimation.

$$\ln TC = \alpha_0 + \sum_{j=1}^{5} \alpha_j \ln V_j + 1/2 \sum_{j=1}^{5}\sum_{k=1}^{5} \delta_{jk} \ln V_j \ln V_k + \sum_{m=1}^{2} \beta_m \ln P_m$$
$$+ 1/2 \sum_{m=1}^{2}\sum_{n=1}^{2} \gamma_{mn} \ln P_m \ln P_n + \sum_{j=1}^{5}\sum_{m=1}^{2} \rho_{jm} \ln V_j \ln P_m$$
$$+ \sum_{j=1}^{5} [\phi_j \cos Z_j + \omega_j \sin Z_j] + \sum_{j=1}^{5}\sum_{k=1}^{5} [\phi_{jk} \cos(Z_j + Z_k)$$
$$+ \omega_{jk} \sin(Z_j + Z_k)] + \sum_{j=1}^{5}\sum_{k=1}^{5}\sum_{i=1}^{5} [\phi_{jki} \cos(Z_j + Z_k + Z_i)$$
$$+ \omega_{jki} \sin(Z_j + Z_k + Z_i)] + \ln v$$

where:

TC = total bank cost which equals total interest expense and total noninterest expense (basically the cost of labor and physical capital);

Vj =five asset and liability outputs composed of: (1) business loans and leases (the sum of real estate loans, agricultural loans, C&I loans, other loans and leases); (2) consumer loans (including credit card loans); (3) securities (held-to-maturity and available-for-sale) plus total noninterest income activities (fee-based activities such as securities trading, investment banking, brokerage, insurance, etc.) minus service charges on deposit accounts; (4) deposits (demand, other transaction, and savings deposits--including MMDAs); and (5), time deposits (plus other borrowed money, federal funds purchases, subordinated notes and debentures).

Pm = the average prices of FTE labor and physical capital (premises and fixed assets);[9] and,

Zj = a transformed value of each output (lnVj) spanning the interval [0.1 x 2PI, 0.9 x 2PI] where PI = 3.14154.

There is no need to specify additional independent variables for banks that experience cost inefficiency relative to other institutions on a cost frontier, for banks that do not optimize between risk and profits, or have a different capital structure, a different return on equity, or (as observed later) to adjust for bank errors in assessing default risk (which can go both ways). User cost is only concerned with the costs and revenues that are actually experienced. However, such additional specifications can be made to illustrate what user cost scale effects could be using $V_{An}$and $V_{Lj}$when these influences on bank cost are held constant (Kovner, Vickery, and Zhou, 2014).

A balanced panel of 337 individual U.S. banks for 2001, 2005, and 2010 (totaling 1,011 observations), all with over $1 billion in total assets (TA) in 2010, is used to estimate bank scale economies. Fourteen separate service lines in the 2010 Call Report were grouped into five, giving a total average bank user cost output value of $921 million in 2010. The corresponding average balance sheet value is $36.7 billion.[10] The output categories were chosen primarily to separate consumer from business loans and transaction and all types of deposits from noninterest income activities.

[9]As noted earlier, the price of funding is not included in Pm. Funding costs paid to depositors (and sellers of purchased funds) are effectively charged to loan borrowers and passed through to depositors and other funds suppliers.The standard symmetry and adding-up restrictions are imposed in estimation.

[10]The numbers differ from Table 1 which reports average output values for 2001, 2005, and 2010 altogether.

Including securities (which now includes noninterest income activities), this gives five service lines. To facilitate a comparison over time, the nominal values were deflated by a cost-of-living index and so are in real terms with 2001 as the base. The value of physical capital is also deflated.[11]

There are cases of poorly performing banks which experienced negative user costs. This occurs for some small banks where asset returns are less than the reference rate. Just as some banks have negative profits in times of business cycle instability and poor loan practices, they can also have negative user cost if their observed asset returns are less than the reference rate at that point in time. The Federal funds rate, which can influence the short-term U.S. government security rate (the reference rate), changed 40 times over 2001-2010. We focus on those institutions whose performance, when observed, is viable at that point in time and value added is positive.

5.1 User Cost Scale Economies Versus the Standard Model.

Table 2 shows how scale economies and predicted average cost differ depending on whether user cost (applied in the national accounts) or balance sheet values (the standard academic approach) are used to reflect the value of banking output. Both models are estimated using the Fourier functional form shown above, suitably adjusted to reflect banking services from both sides of thebalance sheet (5 outputs) or, as is standard in the literature, just outputs from the asset side of the balance sheet (3 outputs).[12]

Table 2: Scale Economies and Average Costs: User Cost Versus Balance Sheet Values

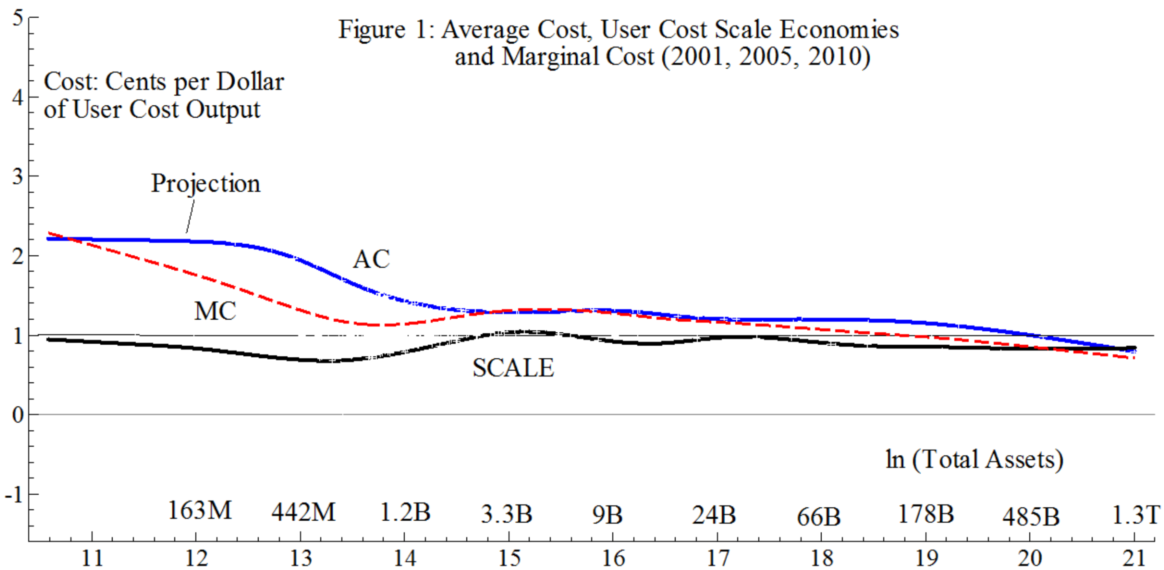
Bank Asset	User Cost 5 Outputs: UCSCE $= \partial \ln TC/\partial \ln V_j$			Balance Sheet 3 Outputs: SCE $= \partial \ln TC/\partial \ln X_j$		
Size-Class	Fourier	AC ¢	MC ¢	Fourier	AC ¢	MC ¢
< 1B	.71	1.70	1.20	.92	5.77	5.25
1 - 5B	.89	1.36	1.22	.94	5.22	4.93
5 - 10B	.93	1.30	1.29	.96	5.04	4.85
10 - 25B	.87	1.29	1.13	.95	4.90	4.66
25 - 50B	1.09	1.15	1.22	.95	4.54	4.30
50 - 100B	.83	1.26	1.06	.96	4.99	4.81
100 - 200B	.88	1.22	1.05	.99	4.85	4.81
> 200B	.84	.93	.82	1.01	4.62	4.72
Average	**.83**	**1.46**	**1.20**	**.93**	**5.37**	**5.01**

The average scale economy using balance sheet data (SCE) is .93 and corresponds well with many other studies using the standard 3 output balance sheet framework. User cost scale economies (UCSCE) average .83 and are 11% lower. A bank engaging in from one to three mergers that doubles its assets would on average expect that its additional costs may rise by only 93% using the standard academic model, reducing the associated incremental costs by 7%. When output from both sides of the balance sheet is considered, a doubling of output would reduce incremental costs by 17%, more than doubling the expected gain from doubling output via mergers or acquisitions.

[11] An Appendix, available from the author by request, contains more information on the data. The calculation process is best explained in Fixler, Reinsdorf , and Smith (2003).

[12] For the standard model, Xj substitutes for Vj in the estimating equation and covers the same three asset outputs used in the user cost approach but are measured as balance sheet values. Here securities services cover only the asset side of the balance sheet while for user cost this service also includes noninterest income activities. Finally, the price of funding is included in Pm.

A good way to illustrate the results from estimating user cost scale economies is in Figure 1.[13] The predicted values of user cost scale economies and predicted AC (the two solid lines) were fitted using a cubic spline. Marginal cost was calculated from AC x UCSCE (dotted line).[14] From Table 2, user cost scale economies are largest (.71) at banks with about $550 million in assets.[15] From that point, scale economies approach 1.0 (thin horizontal line) for banks with about $3.5 billion in assets. This occurs even though over the range of $1-$5 billion, average scale economies are only .93 while MC is just slightly less than AC. Most M&A activity occurs at these smaller banks with assets less than $3 billion. Scale values then fall and rise again and experience local scalediseconomies for banks with $25 to $50 billion in assets. After this, scale economies then fall but flatten out toward the end for the largest banks (those with assets > $100 billion) with a scale value of .88 to .84. Monetary policy can affect scale economies. Changes in the short-term treasury bill rate (used as the reference rate R_R), rates on new loans (R_{An}) as well as floating rates on existing loans, and deposit rates (R_{Lj}) determine user cost scale economies. These economies are based on revenues which, like productivity, can vary over the business cycle.



Mergers among large banks would seem to continue to lower costs, up to the point where a bank holds 10% of nationwide deposits (the legal limit for growth via M&A activity). Three of the largest banks are already at or close to this point currently. Note that throughout the fluctuation of scale economy values in Figure 1, AC is always falling, except for the time when local diseconomies are experienced and the scale estimate is 1.09 (scale diseconomy).

[13]Scale economies for banks with less than $500 million in assets are projections based on too few banks to be relied upon.

[14]The observations used for generating the curves were removed for clarity. All banks in the panel had more than $1 billion in assets in 2010, which means that some had less than $1 billion in 2001 and/or 2005 and only exceeded $1 billion in assets in 2010 due to normal growth or mergers after 2001.

[15]Following convention, and for comparison purposes, bank size is the log of the balance sheet value of total bank assets (not the total value of user cost output).

5.2 Which Services Provide the Greatest Scale Benefit?

The scale results reported in Table 2 and Figure 1 are the sum of the scale values for five banking services. Scale economies estimated for each service line separately overlap one another (not shown). Scale benefits from consumer loans are always larger (further away from 1.0) than for business loans. This reflects the fact that the cost of making a business loan, per dollar of the loan made, is higher. Consumer loans require less work. Often, just a FICO score and an auto or home assigned as collateral for the loan is about all that is needed. Securities and noninterest income activities have, overall, a greater scale benefit than either loan category for banks with less than around $100 billion in assets (which includes almost all banks in the sample). The situation is reversed after this point and both types of loans provide a greater scale benefit than securities (plus noninterest income activities). However, since average cost is the cumulative result of past local scale benefits, the reduced scale benefit of securities compared to loans for the very largest banks is likely more than offset by their contribution to profits due to relatively high prices and revenues.

Deposits cover all transaction and savings deposits. They offer the same or a slightly lower scale benefit than time deposits for banks with less than $12 billion in assets. Banks larger than this receive a reduced scale benefit from deposits while time deposits have a clear scale advantage. Overall, deposits, time deposits, and securities activities contribute more to a bank's scale benefit than business or consumer loans until a bank reaches about $10 billion in assets. After that, deposits are the least scale efficient service for large banks even though deposits are a good and consistent source of low-cost funding relative to purchased money. Time deposits and securities, which includes profitable noninterest income activities, contribute the greatest scale benefit.[16]

6. Conclusions

Past estimates of bank scale economies have focused on economies for the whole bank but were limited only to bank asset services, focusing on various types of loans and asset-based securities holdings. These studies excluded the other half of the balance sheet--banking services provided by demand, savings, and time deposits, usually along with noninterest income activities. Banks do more than just intermediate between depositors and borrowers. Bank deposits--demand, other transaction, savings and time deposits (plus other funding operations)--measured in terms of estimated value added, are larger than the value added from all loan and lease activity. The Bureau of Economic Analysis revised the banking portion of the national accounts in 2003 by including output from bank liabilities as well as assets. National accounts in Europe have done the same.

We estimate scale economies for the whole bank and five bank service lines from both sides of the balance sheet. Instead of defining bank output as the balance sheet value of loan and asset-based securities, the user cost approach used here approximates value added for all services and is used as a measure of banking output. Value added is approximated as asset revenues or liability costs to users, including fees. This is then adjusted for the assumed underlying cost of "passed-through" funds unassociated with any actual banking service (a reference rate) to obtain value added which represents banking output.

Scale estimates are presented by bank size-class and noted separately for five major bank service lines. As individual bank data are highly skewed, the Fourier functional form provides more accurate estimates across different sizes of banks compared to the translog form. Over 2001, 2005, and 2010, our average user cost scale economy value is .83 while it is .93 in a standard model using balance sheet data. If output is doubled following one to three bank mergers, the projected incremental cost savings from doubling output is 7% with the standard balance sheet approach but is 17% over our time period applying the user cost approach. This covers both sides of the balance sheet, is similar to how banking is measured in the national accounts, and suggests that the incentive to merge is larger than previous studies have found.

[16]Noninterest income, not balance sheet value, was also added to the balance sheet output mix used by Wheelock and Wilson (2018 and 2012) who found greater scale benefits for large banks than other studies.

Bibliography

Barnett, B. (1978). The User Cost of Money, *Economic Letters*, 2, 145-149.

Berger, A., Leusner, J., & Mingo, J. (1997). The Efficiency of Bank Branches, *Journal of Monetary Economics*, 40, 141-162.

Berger, A., & Mester, L. (2003). Explaining the Dramatic Changes in Performance of US Banks: Technological Change, Deregulation, and Dynamic Changes in Competition, *Journal of Financial Intermediation*, 12, 57-95.

Davies, R., & Tracey, B. (2014). Too Big to Be Efficient? The Impact of Implicit Subsidies on Estimates of Scale Economies for Banks, *Journal of Money, Credit and Banking*, 46, Supplement 1, 219-253.

DeYoung, R. (2019). Banking in the United States. In A. Berger, P. Molyneux, &J. Wilson (eds.) *The Oxford Handbook of Banking*(pp. 977-203), Third Edition, Oxford University Press.

DeYoung, R., Hasan, I.,& Kirchhoff, B. (1998). The Impact of Out-of-State Entry on the Cost Efficiency of Local Commercial Banks, *Journal of Economics and Business*, 50, 191-203.

Diewert, W.E. (1974). Intertemporal Consumer Theory and the Demand for Durables, *Econometrica*, 42, 497-516.

Donovan, D. (1978). Modeling the Demand for Liquid Assets: An Application to Canada, *IMF Staff Papers*, 25, 676-704.

Feng, G., & Serletis, A, (2009). Efficiency and Productivity of the U.S. Banking Industry, 1998--2005: Evidence from the Fourier Cost Function Satisfying Global Regularity Conditions, *Journal of Applied Econometrics*, 24, 105-138.

Fixler, D. (1993). Measuring Financial Service Output of Commercial Banks, *Applied Economics,* 25, 983-999.

Fixler, D., Reinsdorf, M.,& Smith, G, (2003). Measuring the Services of Commercial Banks in the U.S. National Accounts: Changes in Concepts and Methods, OECD National Accounts Experts Meeting, Paris, October 7-10, STD/NAES(2003)20. Also published in *Survey of Current Business,* September, Issue 9, 33-44.

Fixler, D., & Zieschang, K. (1999). The Productivity of the Banking Sector: Integrating Financial and Production Approaches to Measuring Financial Service Output, *Canadian Journal of Economics*, 32, 547-569.

Hancock, D. (1985). The Financial Firm: Production with Monetary and Non-Monetary Goods, *Journal of Political Economy*, 93, 859-880.

Hughes, J., & Mester, L. (2013). Who Said Large Banks Don't Experience Scale Economies? Evidence from a Risk-Return-Driven Cost Function, *Journal of Financial Intermediation*, 22, 559-585.

Hughes, J., Mester, L.,& Moon, C. G. (2001). Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-Taking into Models of Bank Production, *Journal of Banking and Finance*, 25, 2169-2208.

Humphrey, D. (2020). Distance Functions, Bank Output, and Productivity, *Journal of Productivity Analysis*, 54(1). 13-26.

Humphrey, D., & Vale, B. (2004). Scale Economies, Bank Mergers, and Electronic Payments: A Spline Function Approach, *Journal of Banking and Finance*, 28, 1671-1696.

Kovner, A., Vickery, J., & Zhou, L. (2014). Do Big Banks Have Lower Operating Cost, *Economic Policy Review,* 20, Federal Reserve Bank of New York, December, 1-27.

Mester, L. (1993). Efficiency in the Savings and Loan Industry, *Journal of Banking and Finance,* 17, 267-286.

Radecki, L. (1999). Banks' Payments-Driven Revenues, *Economic Policy Review*, 5 (2). 53-70.

Rhoades, S. (2000).*Bank Mergers and Banking Structure in the United States, 1980-98,* Staff Study 174, Board of Governors of the Federal Reserve System, August.

Rhoades, S. (1998). The Efficiency Effects of Bank Mergers: An Overview of Case Studies of Nine Mergers, *Journal of Banking and Finance*, 22, 273-291.

Sealey, C., & Lindley, J. (1977). Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions, *Journal of Finance,* 32, 1251-1266.

Triplett, J. (1992). Comment. In Z. Griliches (Ed.): *Output Measurement in the Service Sectors,* Studies in Income and Wealth (pp. 287-296), Volume 56, National Bureau of Economic Research, University of Chicago Press.

Wheelock, D., & Wilson, P. (2018). The Evolution of Scale Economies in U.S. Banking, *Journal of Applied Econometrics*, 33, 16-28.

Wheelock, D., & Wilson, P. (2012). Do Large Banks Have Lower Costs? New Estimates of Returns to Scale for U.S. Banks, *Journal of Money, Credit, and Banking*, 44, 171-199.