

Productivity and Economies of Scale in the Production of Bank Service Value Added*

J. Christina Wang **

Research Department
Federal Reserve Bank of Boston
September 2003

Abstract:

This paper uses a new measure of bank service output to estimate various specifications of production and cost functions for Bank Holding Companies (BHCs) over the period 1986 to 1999. The new output series is a true flow measure of bank service value added, and it follows from a unified model of bank operation that integrates risk and the measurement of bank services. The model also establishes separability between the production function of bank services and the funds borrowed and lent—a special intermediate input for banks. The preferred specification of the cost function estimates a large dispersion in productivity levels across BHCs. This implies that one potential benefit of bank mergers is the acquisition by more productive banks of less productive banks, thereby improving the targets' productivity. The cost function also yields an estimate of increasing returns to scale in banking, contrasting with the typical finding in the existing literature of constant returns to scale. On the other hand, the production function estimates decreasing returns to scale. But these estimates are shown to be potentially biased downward, whereas the estimates from the cost function are likely to be biased upward because of measurement errors in the imputed output. The paper tentatively concludes that there is likely to be a modest degree of increasing returns to scale in the production of bank services. If confirmed, this tentative conclusion implies that the cost savings from increased scale of banking institutions should be taken into account in the analysis of mergers and antitrust policy.

Keywords: bank holding company, service output, risk premium, returns to scale, productivity
JEL Classification: G21, D24, O47

I am grateful to Jim Adams and Matthew Shapiro for excellent advice and generous support, and to Gary Solon, Bob Barsky, Susanto Basu, and Lutz Kilian for valuable comments and advice. All errors remain my sole responsibility, and the views expressed here do not necessarily represent those of the Federal Reserve System.

* This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at <http://www.bos.frb.org/economic/wp/index.htm>.

** E-mail: christina.wang@bos.frb.org

I. Introduction

Commercial banks continue to play a central role in the economy, and they constitute a growing share of GDP.¹ Moreover, the structure of the banking industry has undergone sweeping changes in the past two decades. In response to heightened competition from non-bank financial firms enabled by technological progress among other factors, banks have been expanding both the scale and scope of their operations, largely through consolidations. This merger wave coincides with extensive deregulation, which has removed restrictions on product offerings and interstate banking. These changes have motivated many studies. The estimation of bank productivity and returns to scale is of particular interest because of its broad practical applications and important policy implications. For instance, a proper antitrust policy depends on accurate estimates of the impact of mergers on the efficiency of the banking industry.

There are conceptual problems in the existing measure of bank output used in the literature on this subject, as a new model of bank operation in Wang (2003a) shows. The model coherently accounts for portfolio risk in measuring bank value added and proves that bank value added comprises only banking services but not risk premium. The model also distinguishes bank value added from the special intermediate input of loanable funds, and it establishes separability between the funds and a bank's production function for value added, thus rigorously justifying estimating the latter (and its dual cost function) as a function of capital and labor inputs alone. The model implies a new measure of bank value added that uses a flow variable, imputed and properly deflated revenue of bank services, as opposed to the balance sheet—*stock*—value of financial instruments used in the existing measure. This new output measure is then implemented for a sample of Bank Holding Companies (BHCs) in Wang (2003b) and is shown to have quite different magnitude and time series properties compared with the two existing measures of bank output. Therefore, results obtained in the existing literature (regarding issues such as bank efficiency and the impact of mergers) need to be reexamined using this new output measure that is consistent with basic theories of production, financial intermediation, and asset pricing. Such a re-examination is the objective of this empirical study.

I first summarize the model of bank operation developed in Wang (2003a) and the properties of the implied output series constructed in Wang (2003b). I then apply the new output measure to the important issue of estimating productivity differentials and scale economies in the banking sector, and I compare the results with the estimates in the existing literature. Parameter estimates based on the new method provide some evidence for resolving some longstanding puzzles in the micro banking research,

¹ In this paper, unless otherwise specified, “banks” refers to commercial banks, which include independent banks, and (one-bank as well as multi-bank) bank holding companies, and “banking” refers to commercial banking.

such as the contrast between managers' claims of large economies of scale in bank operations to justify mergers and persistent findings of near-constant returns to scale in previous academic studies.

To shed more light on the overall cost structure of banking organizations, I analyze BHCs, since I expect important decisions concerning operation and risk to be made by holding companies instead of banks, which are examined in most previous studies. Besides, the new output measure has an advantage in principle in analyzing BHC data, because, unlike the existing output measures, it can account for portfolio risk of diverse classes of financial instruments in calculating bank service output. Relative to banks, BHCs have more complex portfolios, as they typically comprise multiple banks engaging in diverse activities. Finally, the sample covers 1986 to 1999, making the analysis useful for understanding current developments and potential future trends in the banking industry.

The paper is structured as follows. Section II summarizes the theory underlying the new measure of bank output and its empirical properties. Section III presents the theory of the production function and the dual cost function, discusses the associated returns to scale and productivity measures, and explains the estimation methods. Section IV estimates the production and the cost functions. It also compares the new results with estimates based on the existing output measures. Section V summarizes the conclusions and suggests directions for future research.

II. Bank Value Added, Loanable Funds and Risk, and Their Separability

2.1 The New Model of Bank Operation and the Implied Measure of Bank Output

Wang (2003a) has developed a unified model of bank operation. By integrating theories of production, financial intermediation, and asset pricing, it shows that the return banks earn on the funds lent is compensation for bearing the risk of borrowers' cash flows and should not be considered bank value added, which comprises only services. The quantity of bank service output is affected by risk only to the extent that portfolios of different risk characteristics require different amounts of processing to resolve asymmetric information problems. The model also recognizes that the loanable funds banks borrow and lend are merely an intermediate input in the production of banking services, just as purchased merchandise is for retailers. The model thus classifies conventional banking into three qualitatively distinct functions: (1) mitigating the asymmetric information problems associated with uncertain investment returns, (2) financing loans with borrowed funds, and (3) providing transaction services. Both (1) and (3) consume real resources such as labor and physical capital, and they create the value added of a bank. In contrast, function (2) merely transfers the property and the income generated—"pure interest"—between ultimate users and suppliers of funds, creating no value beyond the return that depositors and bank shareholders with given preferences would demand for bearing borrowers' risk according to asset pricing theories. The model in turn partitions the gross interest income of a bank into pure interest and

implicit revenues of the two kinds of bank services, as summarized in Diagram 1 (reproduction of Figure 3 in Wang, 2003a). Pure interest is not bank value added, but it is still part of (nominal) *gross* output. The residual interest income net of risk-related returns, plus explicit fees, then equals nominal bank value added. It is shown that this imputation of value added remains valid even when banks cannot fully resolve borrowers' private information, or when banks have market power in supplying the services. Last, the model establishes separability between the funds and the production function of bank value added and shows that the latter should be specified as a function of labor and physical capital only.

2.2 Empirical Differences between the New and the Two Existing Measures of Bank Output

Wang (2003b) empirically implements the new output measure for BHCs. Following the model's structure, it divides a BHC into three hypothetical divisions A, B, and C that perform the functions 1, 2 and 3 above, respectively, and then constructs the output of each division by partitioning a bank's gross interest receipts according to Diagram 1. Table 1 (reproduction of Table 1 in Wang, 2003b) summarizes the imputation of A's and C's outputs. A bank's aggregate value added then equals the sum of these two outputs. By comparison, the first of the two existing measures of bank output uses book values (BVs) of financial instruments on bank balance sheets as output (henceforth referred to as the BV-based output measure). The second existing measure is used in the National Income Accounts (according to the 1993 System of National Accounts and its latest revision (Moulton and Seskin, 2003) and is referred to henceforth as the SNA93 measure). It regards the entire gap between bank interest receipts and payments as the nominal value of bank service output that is priced implicitly. Its sum with explicit fees then equals total nominal output.

There are three major differences between the new output measure and the existing ones. First, the new measure is more than 25% smaller than the SNA93 measure, and it is almost two orders of magnitude smaller than the BV-based measure. More importantly, the new output series has different time series properties. It is more volatile over time—perhaps partly the consequence of greater measurement errors due to insufficient data—and it appears to be affected more by the last recession. This may mean that banking is cyclical, or it may be that banks have countercyclical markups on services so that the fluctuations in real output would be smaller if more precise price deflators were available. Second, the new series shows a higher growth rate than the two existing measures, especially in the latter half of the sample period (1993 to 1998). Finally, one interesting result is that the nominal value of the new output equals banks' non-interest cost on average, implying zero profit in the long run. This may be viewed as informal support for the new measure of bank output under certain circumstances.

2.3 Contrast with the Literature

An extensive empirical literature estimates bank efficiency and returns to scale from the cost and profit functions, and all but three studies use the BV-based measure of bank output.² Specifically, Rogers (1998) is the only one that uses revenue from OBS (off-balance-sheet) activities to measure that particular kind of output.³ Mixing OBS revenue with the BV-based measure of loans and securities, he finds that the inclusion of OBS activities produces a marked increase in the efficiency estimates of large banks. This is not surprising given that OBS incomes are concentrated in a score of large banks. However, he mentions only in passing the rationale for using revenues to measure the output of OBS activities, and he provides no explanation for combining the two measures. Instead of using revenues, Hughes and Mester (HM, 1998) include interest rates (regarded as output prices), along with the stock of equity capital, in the cost function regressors. The interest rates are not introduced on the basis of cost minimization, but as the outcome of utility maximization (with respect to profit and risk) by risk-averse bank managers. HM find the interest rates statistically significant, and large banks exhibit a higher degree of scale economies under the new formulation. They interpret the results as proof that bank managers choose risk levels lower than those that would otherwise minimize costs. In contrast, the new banking model demonstrates that one component of loan interest rates should enter the cost function implicitly through the new output measure; this is consistent with shareholder value maximization without resorting to an arbitrary specification of managers' preferences.

III. Bank Production and Cost Functions and Their Estimation

The value-added production functions for Divisions A and C are specified as in Wang (2003a). The cost function is dual to the production function; it summarizes all the economically relevant characteristics of the technology. I estimate only the production and the cost functions, but not the supposedly more general profit function, in order to focus on uncovering the technical properties of bank service provision and to circumvent the difficult task of specifying the product market factors that affect output pricing. Previous studies have documented imperfect competition in the markets for depositor services and lending to small firms. This means that a bank's output price will be set by all banks' strategies, which may depend on numerous market parameters of both the present and the past. Few of these parameters are observed, making them impossible to quantify. So the standard bank profit function taking output price as given is not correctly specified. In contrast, as long as banks are price takers in *factor* markets, which is a more plausible assumption, the standard production (and cost) function can

² See, for example, Berger *et al.* (1987), Clark (1996), Hughes and Mester (1993), and Hughes *et al.* (1996). Berger and Humphrey (1997) survey a variety of efficiency definitions and estimation techniques.

³ Stiroh (1999) applies Rogers's approach in one of the specifications in his efficiency analysis.

generate valid estimates of parameters of the technology without observing product market conditions (Hall, 1986).⁴ Between the production and the cost functions, previous research almost exclusively chooses the latter, arguing that input prices for an individual bank are more likely to be exogenous than input quantities. (Marschak and Andrews (1944) first analyze the endogeneity problem with inputs in the production function.) But output is also endogenous, chosen to maximize profit, and duality between the two functions leads to a high correlation between output and the cost function's error term (shown in section 3.2). This biases estimates of the output coefficient and can outweigh the advantage of using input prices. Moreover, the use of many proxies to overcome data limitations means that the imputed new output series in this study may be subject to greater measurement errors than the inputs, exacerbating the bias in the cost function parameters. Therefore, I choose the production function to be the primary specification to minimize parameter bias. But I also estimate the cost function so that I can evaluate the net impact of the new output measure with the same specification as used in existing studies.

3.1 The Production and Cost Functions, Returns to Scale, and Productivity

Given that only total labor and total capital inputs are observed, I adopt the single-output specification of the production function (i.e., the aggregate for Divisions A and C). For an individual bank b , it can be written (in logarithms) as:

$$\ln Y_b = F(\ln \mathbf{X}_b) + \xi_b^Y, \quad (1)$$

where Y is the output, \mathbf{X} the vector of inputs, and ξ_b^Y the error term. The next section discusses the possible components of ξ_b^Y and their implications for the parameter estimates.

The degree of returns to scale is defined as the sum of (N) input elasticities, i.e.,

$$\gamma^Y = \sum_{i=1}^N \frac{\partial \ln Y}{\partial \ln X_i}. \quad (2)$$

Dual to the production function, the variable (or restricted) cost function can be defined conditional on the quantity of quasi-fixed inputs (e.g., physical capital) that are costly to adjust within a period. That is, it can be written as:

$$\ln C_b | \mathbf{K}_b = f(\ln \mathbf{w}_b, \ln \mathbf{Y}_b, \ln \mathbf{K}_b, \mathbf{v}) + \xi_b^C, \quad (3)$$

where C_b is the variable cost of bank b and \mathbf{K}_b the vector of fixed inputs, and \mathbf{v} are control variables often added to proxy for unobserved factors. ξ_b^C is the counterpart of ξ_b^Y and is also discussed in the next section. The efficiency study literature treats ξ_b^C as a composite term: $\xi_b^C = u_b^C + \varepsilon_b^C$, where u_b^C (called the cost-based X-efficiency) reflects a bank's efficiency level relative to the best performing bank, and ε_b^C is a random error. But this means that some banks operate within the possibility frontier—inconsistent

⁴ The cost function can indirectly yield the estimate of markup in addition to estimates of technical parameters such as the degree of returns to scale.

with the principle of optimal firm behavior. So, I view u_b^C as the bank-specific, long-run average productivity, without exploring whether there are substantive distinctions between the two different interpretations.

Following the existing literature, the relative productivity level is defined as the ratio between predicted actual cost and the predicted minimum cost that could be achieved if the bank had the same long-term productivity as the best bank in the sample. That is, compared with the most productive bank (*min*, i.e., having the lowest estimate of u^C), the relative productivity of bank b is computed as

$$C_b^{\text{PRD}} = \exp(\hat{u}_{\min}^C - \hat{u}_b^C). \quad (4)$$

The hat variables denote estimated values. $C_b^{\text{PRD}} \in (0, 1]$, and equals 1 only for banks with the best cost-based productivity in the sample. Since C_b^{PRD} is sample-dependent, its rankings, instead of actual levels, may be more useful in evaluating relative productivity by reducing the impact of random errors.

The cost-based returns to scale equal the inverse of the sum of all (M) output elasticities of cost, i.e.,:

$$\gamma^C = 1 / \left[\sum_{j=1}^M \frac{\partial \ln C}{\partial \ln Y_j} \right]. \quad (5)$$

For the self-dual Cobb-Douglas (C-D) function, when $M = 1$ and all the inputs are variable, γ^Y is equivalent to γ^C . Appendix A illustrates the relationship. But the equivalence breaks down when some inputs are quasi-fixed, or $M > 1$, and for other functional forms (e.g., the Translog).

3.2 Estimation of the Production and the Cost Functions

In the existing efficiency literature, parametric and non-parametric estimation techniques are used with about equal frequency (see Berger and Humphrey, 1997). The former has the advantage of explicitly accounting for random errors in measured cost, but it imposes restrictions on the functional form; the opposite is true for the latter. I adopt the parametric approach in this study, because the non-parametric approach is likely to yield less accurate estimates by failing to allow for random errors in a period characterized by rapid changes, such as the past decade. I consider both the Cobb-Douglas and the Translog functional forms. The simplicity of the former facilitates the interpretation of the technical parameters, whereas the latter is more flexible and more commonly used in existing studies. Comparing results from the two allows comparability with the existing literature and also helps one gauge the effect of allowing flexibility in the functional form.

How the production function should be estimated depends on the properties of its error term (ξ_b^Y). In general, ξ_b^Y has four components: (1) technology shocks, (2) classical measurement errors in inputs, (3) unobserved input variations (such as labor effort and capital utilization), and (4) classical measurement errors in Y . Only (4) causes no bias in the parameter estimates. Error (2) will bias the input coefficients

toward zero and in turn bias downward the estimate of returns to scale. Error (3) will bias upward the coefficients if unobserved and observed input components are positively correlated. If both errors (2) and (3) are present, the net bias is ambiguous. In the banking context, bias due to both (2) and (3) may be low owing to the quasi-fixity of observed labor and capital inputs, unless there is high persistence in (1) or in demand shocks. The technology shocks (1) can be further decomposed into two components: a permanent part—reflecting the long-term average level—and a stochastic part. When longitudinal data are available, partitioning ξ^Y into u^Y and ε^Y reflects one effort to mitigate the endogeneity problem stemming from possible correlations between output and long-term productivity differentials across banks. u^Y can be treated as bank-specific “fixed effects” and can be estimated using a within estimator. But the stochastic technology shocks may still be correlated with input, and the bias can go in either direction, depending on the time horizon. In banking, over the short run we may expect little or even negative correlation, because bank services cannot be stored, while observed inputs are quasi-fixed. (For the same reason, observed and unobserved inputs may also be negatively correlated.) Over the long run, inputs will be positively correlated with persistent shocks. See Appendix A for a more detailed illustration of possible biases caused by the error components. In short, the net bias is likely to be negative, but it is essentially undetermined.

The corresponding components of the cost function error term (ξ^C) are (1) technology shocks, (2) derived measurement errors in input prices, (3) classical measurement errors in output, and (4) measurement errors in cost. Both (1) and (3) are likely to bias downward the output coefficient and in turn bias upward the returns to scale estimate. Moreover, the use of many proxies to impute the new output series means that measurement errors (3) can be serious in this study. The two combined may explain why the estimate of returns to scale from the dual is typically much greater than that from the primal. Again, partitioning ξ^C into u^C and ε^C aims to use longitudinal data to mitigate part of the endogeneity associated with (1). See Appendix A for an in-depth discussion of bias in the cost function parameter estimates attributed to the error components. On net, the output coefficient is most likely biased toward zero, so an estimate of increasing returns to scale (IRS) should be discounted as evidence of true IRS, especially if the estimate from the production function indicates otherwise.

In summary, the production function parameters are likely to suffer from smaller biases than the cost function parameters, especially without natural or proven instrumental variables (IVs) for the new output series. I therefore choose the production function to be the primary specification. I estimate bank-specific long-term productivity (i.e., u_b^Y and u_b^C) using the fixed effects model and the within estimator, which offer clear statistical inferences for the parameter estimates that are not available to the common

approaches used in existing efficiency studies with panel data.⁵ I first estimate the following Translog value-added production function:⁶

$$\ln Y = (\alpha^Y + u^Y) + \sum_{i=1}^N \beta_i \ln X_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \beta_{ij} \ln X_i \ln X_j + \varepsilon^Y. \quad (6)$$

Y is aggregate output, and X_i is the i^{th} input. u^Y represents the time-invariant bank-specific productivity, and ε^Y the random error. (6) simplifies to a C-D function with the additional constraint: $\beta_{ij} = 0, \forall i$ and j . The within estimator estimates the composite intercept ($a_b^Y \equiv \alpha^Y + u_b^Y$) for each bank b from the bank-mean equation: $\hat{a}_b^Y = \ln Y_b - \mathbf{X}_b \cdot \hat{\mathbf{B}}$, where $Y_b = \frac{1}{T} \sum_{t=1}^T Y_{bt}$ and $\mathbf{X}_b = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_{bt}$. \hat{u}_b^Y is then derived as:

$$\hat{u}_b^Y = \hat{a}_b^Y - (\min_i \hat{a}_i^Y), \quad \forall b. \quad (7)$$

I next estimate the following Translog variable cost function:⁷

$$\begin{aligned} \ln C = & (\alpha^C + u^C) + \sum_{i=1}^N \beta_i \ln w_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \beta_{ij} \ln w_i \ln w_j + \sum_{k=1}^M \gamma_k \ln Y_k \\ & + \frac{1}{2} \sum_{k=1}^M \sum_{m=1}^M \gamma_{km} \ln Y_k \ln Y_m + \sum_{i=1}^N \sum_{k=1}^M \eta_{ik} \ln w_i \ln Y_k + \varepsilon^C, \end{aligned} \quad (8)$$

with the standard parameter restrictions:

$$\beta_{ij} = \beta_{ji}, \text{ and } \gamma_{km} = \gamma_{mk}, \quad (9)$$

$$\sum_{i=1}^{N^V} \beta_i = 1, \quad (10)$$

$$\sum_{i=1}^N \beta_{ij} = \sum_{i=1}^N \beta_{ji} = 0, \quad (11)$$

$$\sum_{i=1}^{N^V} \eta_{ik} = 0, \quad (12)$$

C is a bank's variable cost; w_i is the price or quantity of the i^{th} input, and N^V the number of variable inputs; Y_k is the k^{th} output. u^C is the constant bank-specific productivity term. The additional constraint for the C-D functional form is: $\beta_{ij} = \gamma_{km} = \eta_{ik} = 0, \forall i, j, k$ and m . The cost function is subject to certain parameter restrictions. Equation (9) imposes symmetry on the own cross-products of input prices and outputs. (10) to (12) ensure that the cost function is homogeneous of degree one in all the input prices. In contrast, the production function has free parameters. This is because the cost function represents only the economically relevant subset of the production technology, given input prices. The restrictions should

⁵ Most existing studies use cross-section data and estimate u_b^C and ε_b^C based on assumptions of their respective distributions (e.g., u_b^C follows a half normal and ε_b^C a normal distribution), which are often rejected by test results.

⁶ Here, for clarity of expression, I omit the bank superscript (b) for all the variables except α^Y , and the time subscript (t) for the regressors, which are all time-variant.

⁷ I do not impose the input share equations because cross-equation equality of common parameters is rarely found to be satisfied, see for example Dixon et al. (1987). Besides, the share equations become irrelevant with only one variable input (labor). See, e.g., Berndt (1991, chapter 3), for a derivation of the standard parameter restrictions.

improve the efficiency of the estimates. The composite intercept ($a_b^C = \alpha^C + u_b^C$) for a given bank b is again estimated from the bank-mean equation, and then $\hat{u}_b^C = \hat{a}_b^C - (\min_l \hat{a}_l^C)$.

The degree of returns to scale corresponding to (6) and (8), respectively, equals:

$$\gamma^Y = \sum_{i=1}^N \frac{\partial \ln Y}{\partial \ln X_i} = \sum_{i=1}^N (\beta_i + \sum_{j=1}^N \beta_{ij} \ln X_j). \quad (13)$$

$$\gamma^C = 1 / \left[\sum_{k=1}^M \frac{\partial \ln C}{\partial \ln Y_k} \right] = 1 / \sum_{k=1}^M (\gamma_k + \sum_{m=1}^M \gamma_{km} \ln Y_m + \sum_{i=1}^N \eta_{ik} \ln w_i). \quad (14)$$

In the Translog functional form, γ^C (and γ^Y) depends on not only the cost (and the production) function parameters, but also on actual data. But in the C-D case, the technology is homogenous of a constant degree:

$$\gamma^C = 1 / \sum_{k=1}^M \gamma_k, \text{ and } \gamma^Y = \sum_{i=1}^N \beta_i.$$

IV. Empirical Results

4.1 The Data and Sample Selection

All the BHC balance sheet and income statement data (on a consolidated basis, but excluding non-bank subsidiaries) come from the quarterly Reports of Condition and Income (form FR Y-9C) filed by all FDIC-insured BHCs. The Federal Reserve Bank of Chicago maintains these data, along with data of all BHC mergers since 1986. The sample runs from the second quarter of 1986 to the second quarter of 1999. All balance sheet data are end-of-quarter figures. Last, all interest rates are monthly secondary-market rates from the Federal Reserve Economic Database (FRED) at the Federal Reserve Bank of St. Louis.

In the sample panel, each combined BHC after a merger is considered a new entity. The first quarter following each merger is removed to avoid potential accounting anomalies incident to business combinations.⁸ For multi-tier BHCs, only the top-tier BHC is included in the sample. Observations with implausible values for some revenue or cost variables (e.g., negative expenses on fixed assets and employee salaries) are also removed.⁹ In addition, extreme values (defined as those negative or beyond the 99th percentile) of the imputed nominal value added and missing values for any other variables used in the production and the cost functions are excluded from the sample. To prevent BHCs with heavy non-bank operations from skewing the results, I also exclude the BHCs whose non-bank assets exceed 10% of the consolidated total assets.¹⁰ Last, a BHC must have at least 12 quarters of time series data to be

⁸ The first observation of a BHC's time series starting from the second quarter or later are also removed, since the quarterly values cannot be derived for such quarters given that all the flow variables reported in the income statement are year-to-date figures.

⁹ The only pattern in the occurrence of such negative values seems to be that they appear most often in the fourth quarter and concentrate around the first year a banking organization is formed.

¹⁰ When those BHCs are included in the same set of regressions of the value-added cost function, they cause only

included in the regressions.¹¹ This lower bound is imposed to balance the benefit of a large sample against that of a long time series (for estimating the BHC productivity fixed effects). It changes none of the parameters' significance, except to lower the estimate of γ^C , possibly because of greater variances and thus the errors-in-variables bias. Overall, the selection criteria trim the initial data set of 77,961 observations belonging to 4,527 BHCs down to a final sample of 36,616 observations belonging to 1,505 BHCs.¹² It is an unbalanced panel with BHC time series of up to 44 quarters. Some BHCs have missing observations in the middle of the time series, but their occurrence exhibits no regularity and is thus assumed to be random.

4.2 Estimation of the Production and the Cost Functions

The single-output value-added production function is estimated only for the two flow measures of output—the new and the SNA93 measures—because no proper aggregate output is available for the BV-based measure. Three inputs—labor, physical capital, and cash reserves—enter as regressors, and the dependent variable is aggregate output. See Wang (2003b) for details of the computation of the output series. Aggregate output becomes a regressor in the dual variable cost function (the duality is shown in Appendix A), where capital is treated as quasi-fixed. Cash reserves are also viewed as quasi-fixed—representing an “inventory” assisting the provision of depositor services—and the quantity is calculated as the balance-sheet value divided by the GDP deflator. The last input, labor, is always treated as variable, and the wage rate as the regressor equals total salaries and benefits divided by the total number of full-time equivalent employees. So, the dependent variable—the variable cost—is just the cost of labor. (Table 2 and Appendix B summarize, respectively, the formulae and the data items used to compute cost, inputs, and input prices.) Both the production and the cost functions are estimated using the within estimator in a fixed effects model of both the C-D and the Translog functional forms.

I follow the asset approach in specifying the cost function according to the BV-based output measure in order to generate results comparable to those found in the existing literature. The dependent variable is total cost, equal to the sum of interest and non-interest expenses net of explicit charges on depositor services, and is considered an intermediate income that offsets the cost of raising funds through deposits. The regressor outputs are the real book values (calculated using the GDP deflator) of four

minor differences in the regression results, and all are statistically insignificant. The main change is that estimates of the sample mean γ^C mostly decline, and the magnitude of reduction depends on the averaging method used.

¹¹ 45,580 observations from 3181 BHCs remain before this criterion is imposed, and they are called the usable observations. This requirement decreases the sample mean total assets (by 41%) and the cross-section variance of the variables, since most of the entities excluded are large BHCs, which merged often and thus have short time series between mergers, but the relative asset compositions remain similar. See Table 2 and Figure 8 in Wang (2003b) for comparisons of the average BHC asset size, across size groups and over time, respectively, among the original sample, the subsample of all the usable observations, and the final sample.

categories of financial assets: (1) investment securities, (2) business lending, equal to C&I loans plus lease financing receivables, (3) consumer loans, and (4) real estate loans. Besides K and L , inputs also include loanable funds. Deposits are considered a variable fund input, so the actual deposit interest rate is a regressor. The federal funds rate is used as a proxy for the price of other liquid borrowings (mostly federal funds).¹³ Equity capital (E) is always treated as a quasi-fixed input.

Every equation is estimated with a full set of quarterly dummies to capture the common time factors. To choose the consistent estimator for the parameter covariance matrix, I first apply White's test. It rejects the null of homoscedasticity, which is hardly surprising given the large number of disparate BHCs. Then, the test proposed by Bhargava *et al.* (1982) for panel data strongly rejects the null hypothesis of no serial correlation (vs. the alternative of positive serial correlation).¹⁴ However, using ρ -differencing to correct for serial correlation in the parameter estimates will exacerbate the problem of errors in variables.¹⁵ Hence, I keep the fixed-effects estimates of the coefficients, but I estimate their standard errors using the most general version of the robust covariance matrix estimator proposed in Arellano (1987). According to (13), the variance of the returns-to-scale estimate from the production function should be a linear function of variances and covariances of the parameter estimates, and, in the case of the Translog function, of data. According to (14), the variance of the scale estimate from the cost function is a non-linear function of the parameter covariances and thus is computed using the delta method. For both Translog functions, the variances are estimated at (i.e., conditional on) the sample mean.

In the production function for the new output, L explains most variations in Y , and cash reserves ranks second, whereas K explains little, probably because the observed K is subject to the most severe quasi-fixity (see Table 3a).¹⁶ The estimate of returns to scale (γ^Y) equals 0.77 for the C-D function. A one-sided asymptotic test cannot reject the null of decreasing returns to scale (DRS) at the 5% level. The

¹² Summary statistics of the final sample are the same as those reported in Wang (2003b).

¹³ I omit illiquid long-term borrowing (LB, largely subordinated debt) in the estimations, because neither LB nor the return r^{LB} , calculated as the expenses on LB divided by the quarterly balance of LB, is significant.

¹⁴ The exact upper and lower bounds for the test statistics as computed by Bhargava *et al.* (1982) are not applicable to an unbalanced panel. However, since both bounds quickly converge to 2 as the sample size increases, and my sample contains a large number of BHCs, I can reject the null, as the test statistic (=0.76) is much less than 2. I use the Bhargava test because most BHCs have time series too short to estimate the individual correlation coefficients with confidence, even though it is restrictive to assume that all BHCs have the same covariance structure given the large number of disparate BHCs.

¹⁵ The unbalanced panel renders the method suggested by Bhargava *et al.* (1982) inapplicable. Moreover, like the Durbin-Watson test, their test cannot identify the definitive structure of the serial correlation.

¹⁶ The within estimator likely exacerbates the errors in inputs and thus biases downward the input coefficients, and the common-intercept OLS estimates are indeed higher: $\gamma^Y \approx 1$ (Table 3b). The parameter estimates change only slightly without the time dummies (Table 6). The production function is also estimated without the input of reserves, because of its high correlation with the new output ($\rho = 0.96$). The coefficient on labor alone absorbs most of the variation in output that is otherwise accounted for by reserves, so the estimate of γ^Y falls by only 0.05

Translog function (see Table 4) yields the same relationship, and it hardly improves the model's fit, generating the same adjusted R^2 as the C-D function.¹⁷ But an F -test rejects the null that the second-order and cross-product terms in the Translog all equal zero. An F -test (Greene, 1997, p. 617) also rejects the null $H_0: u_I^Y = \dots = u_N^Y$,¹⁸ and adjusted R^2 with the BHC dummies rises to over 0.96, indicating that a significant portion of the difference across BHCs is essentially unexplained—accounted for by the time-invariant productivity differentials. We will also see that the production function has a greater error variance relative to the cost function, consistent with the premise of large errors in output.

For the SNA93 output, the same production function yields similar within estimates of the coefficient on L , but significantly lower estimates of the coefficient on cash reserves than those using the new output series (Tables 6 and 7). Thus the SNA93 output also leads to lower estimates of returns to scale.¹⁹ Since cash reserves are an inventory held mostly for providing depositor services, the higher coefficients on reserves using the new output suggest that it is a more accurate measure of depositor services than the SNA93 output. But the latter has much smaller stochastic variations over time than the former, as indicated by the higher adjusted R^2 from the within estimation using the SNA93 output without period dummies and given that the two output series have similar variances ($\sigma_{\text{new}} = \$61$ million vs. $\sigma_{\text{SNA93}} = \88 million). This suggests that the supply of or the markup on bank services may be more volatile across time than revealed in the SNA93 output, and the new output series may be subject to greater measurement errors.

Estimates of the cost function for the new output are reported in Table 5a (C-D) and Table 6 (Translog).²⁰ The most noticeable feature is that the estimates of returns to scale (γ^C) are significantly greater than 2.5 for the C-D function (adjusted according to (A.8) in Appendix A) and 4.3 for the Translog, much higher than the typical finding of constant returns to scale in existing studies.²¹ These estimates of γ^C , however, are likely biased upward by measurement errors in the new output. The next

to 0.72.

¹⁷ The within estimate of γ^Y (0.81, DRS) is lower than the simple OLS estimate (0.99, standard error $\sigma = 0.005$, not shown), and time dummies again decrease the γ^Y estimate—from 0.87 (not shown, $\sigma = 0.02$) to 0.81.

¹⁸ The 5% critical value for all the F -statistics is 1, while the actual values range from 3.02 to 3.61.

¹⁹ By comparison, the SNA93 output results in significantly higher OLS estimates of the coefficient on L , offsetting the lower estimates of the coefficient on cash reserves to yield the same estimates of returns to scale (Table 3b).

²⁰ I report only the cost function with cash reserves, since its high correlation with the new output measure merely lowers the output coefficient from 0.164 to 0.117 in the C-D function, and the estimate of γ^C thus rises from 4.0 to 4.5. The Translog function only increases adjusted R^2 (net of the BHC dummies) from 0.82 to 0.86. Gross of the BHC dummies, adjusted R^2 rises to no less than 0.98. A similar value obtains when the cost function is estimated with cross-section data from only one quarter (with the most observations), indicating that the goodness of fit is not driven by comovements between persistent time series.

²¹ The estimate of γ^C without the adjustment equals 4.5, indicating that correcting the calculation of γ^C may enable the biases in the cost function parameter estimates to offset one another to yield a less biased estimate of returns to scale. Time dummies lower the estimates of γ^C (e.g., from 4.5 ($\sigma = 0.34$, not shown) to 2.5 in C-D), possibly because they attenuate the errors-in-variable bias in the cost function more than in the production function.

section will test the bias and examine the feasible remedies.²² Restrictions on the cost function parameters are all rejected, suggesting either biases due to errors in the explanatory variables or a misspecified functional form. Relaxing these restrictions, however, causes no qualitative changes to γ^C estimates, the average productivity level, or the significance of the parameter estimates.²³ Possibly for the same reason, the C-D function generates a wrong (positive) sign for K 's coefficient (representing the cost elasticity of K), which should be negative given that K and L are substitutable in the chosen functional form.²⁴ In fact, an upward bias in K 's coefficient is consistent with a downward bias in the output coefficient, and hence with high estimates of γ^C . The F -test again rejects the null that all the BHC fixed effects are the same, i.e., $H_0: u_i^c = \dots = u_N^c$. The sample average of the within estimates of relative productivity levels (based on u_i^c) is 0.43 for both functional forms (all output-weighted, and excluding observations outside the 5th and the 95th percentiles to minimize the impact of outliers).²⁵

The same cost function is then estimated for the SNA93 output measure (see Table 5 for the C-D estimates and Table 6 for the Translog). The output coefficient rises to 0.53 (C-D), thus lowering the estimate of γ^C substantially to 1.3.²⁶ This relationship mirrors the pattern uncovered in the production function estimation: either bank output exhibits large stochastic variations, or the new output measure contains greater measurement errors than the SNA93 output, or both. In either case, the new output would lead to a more severe downward bias in the coefficient on Y than the SNA93 output, and in turn to a greater upward bias in the γ^C estimate, given that all the other variables in the cost function are the same. Also from the cost function, the estimate of the average productivity level is 0.61 (0.65 for Translog)—higher than the estimate using the new output.

The Translog cost-function parameter estimates using the BV-based output measure are reported in Table 7.²⁷ The estimate of γ^C at the sample mean is 1.3, and the null of constant returns to scale cannot

²² The high estimate of γ^C from the Translog function should especially be discounted as evidence of high increasing returns to scale, since it is not even comparable with γ^Y from the production function (shown in Appendix A).

²³ The only exception is the constrained parameters, e.g., L 's coefficient is lowered from 1 to 0.63. A parameter is significant if it exceeds the critical value in a two-sided asymptotic test at the 5% significance level, unless otherwise specified.

²⁴ The cost function parameter constraints do not seem to be the cause, since the K coefficient remains positive even when the restriction $\beta_L = 1$ is relaxed. On the other hand, the capital (K) elasticity of cost is negative in the Translog function calculated at the sample mean.

²⁵ Productivity levels are decreasing in output. (Similar for the BV-based cost function.) Average productivities fall to 0.28 (C-D) when only the observations outside the 1st and the 99th percentiles are excluded, illustrating that the new-output cost function is more susceptible to outliers. The productivity level based on the within estimates of u_i^c 's is lower than the values calculated using either the distribution-free approach (0.58) or the thick-frontier approach (0.60) that is standard in previous studies of the efficiency literature (see, e.g., Berger and Humphrey, 1997).

²⁶ Similarly, in the Translog cost function, the estimate of γ^C falls from 4.3 to 1.8 (see Table 6).

²⁷ The F -test again rejects the null of equal intercepts. The adjusted R^2 is only about 0.88, even though this cost function contains many more regressors. As in Hughes and Mester (1998), the marginal effect of equity on cost is found to be positive, meaning that it behaves as a complement to instead of as a substitute for borrowed funds.

be rejected, confirming the common previous finding of near constant returns to scale. The sample average productivity level (≈ 0.70) is higher than that from the value-added cost function, but it remains lower than the levels calculated using either the distribution-free (0.81) or the thick-frontier (0.77) approach, both of which are similar to those found in the existing literature (70% – 80%). These values show that the level of average industry productivity depends on the output measure (especially when using the within estimator) as well as on the estimation method. Nonetheless, the cost-based productivity estimates associated with the three output measures are highly correlated: $\rho = 0.87$ between the estimates using the new and the SNA93 outputs (rank correlation 0.89), and $\rho = 0.72$ between the estimates from the new and the BV-based outputs (0.76). Interestingly, a BHC's productivity level estimated based on the new output depends positively on its asset size and SNA93 measure of output.

4.3 Alternative Specifications of the Cost Function and Biases in the Estimates

Three more versions of the value-added cost function are estimated for robustness checks: one treats K as a variable input (version V1), and the other two account for Division A's and C's services separately (versions V2 and F2, respectively, see Table C.1).²⁸ In V1 and V2, r^K is a regressor; it equals total expenses on premises and fixed assets divided by the stock of these assets, and the cost is the total expense on both L and K . Qualitatively similar results emerge (see Table 5b). The most salient is that all the estimates of γ^C are significantly greater than 1. But the especially high values from V2 and F2 are not comparable with γ^Y from the production function (see Appendix A). Treating K as variable lowers the average productivity level, suggesting that K is more likely to be quasi-fixed, since its level explains more of the cross-BHC cost differentials. Next, the cost function corresponding to the BV-based output is also estimated treating K as variable. Individual coefficients are little affected, but the cumulative effect is that the γ^C estimate falls by one standard deviation and the null of constant returns to scale cannot be rejected.

Last, I experiment with a “loanable-funds” cost function that contains only financial variables: the dependent variable contains only the cost of all fund inputs, i.e., it excludes labor and physical capital costs; the independent variables include the same four categories of financial assets as the fund outputs, as well as the interest rates of deposits and other borrowed funds. The new model in Wang (2003a) shows that the standard multi-product cost function is not well defined for the supply of funds. So, the purpose here is to compare the loanable-funds cost function with parameter estimates from the BV-based cost function and to explore whether coefficient values in the latter are driven by financial assets, whose magnitude dominates among the latter's regressors, which also include prices of the primary inputs (i.e., the wage rate and capital rental rate). As expected, the “loanable-funds” cost function generates sample

²⁸ V2 and F2 can be thought of as approximating (in logarithms) a reduced-form joint cost function that is the sum of

average estimates of γ^C (1.16, $\sigma = 0.07$) and productivity (60%) that are similar to the respective estimates from the BV-based cost function, suggesting that the estimates in previous studies may be largely driven by the inclusion of financial assets.

As shown in section 3.2, unobserved input variations bias the production function parameter estimates, and Sbordone (1997) suggests using the intertemporal change in the observed components of inputs to serve as a proxy for the unobserved change in utilization of quasi-fixed inputs. I use two proxies for the unobserved components in $L_t(K_t)$ in the production function (and K_t in the variable cost function): $L_t - L_{t-1}(K_t - K_{t-1})$ and $L_{t+1} - L_t(K_{t+1} - K_t)$. $L_t - L_{t-1}$ attempts to capture the concurrent movement between the number of workers (observed) and the labor hours or effort level (unobserved); $L_{t+1} - L_t$ serves as a leading indicator, since labor hours and effort often rise in response to a shock before the payroll is expanded. Adding $L_t - L_{t-1}$ (detailed parameter estimates not shown) in fact raises the within estimate of γ^Y from 0.77 to 0.80 ($\sigma = 0.02$), while adding $L_{t+1} - L_t$ raises it to 0.84 ($\sigma = 0.02$); this is virtually the same result as obtained by including both sets of proxies. Overall, conditioned on the proxies used, unobserved variations in input utilization cause only minor bias toward zero in the coefficients on inputs, and in turn downward bias in the estimate of γ^Y .

The estimates of γ^C from the cost function, on the other hand, are quite likely biased upward by measurement errors in the new output. The first evidence is that the within estimates are all considerably greater than the common-intercept OLS estimates (e.g., 2.5 vs. 1.1 ($\sigma=0.01$)) and the γ^Y estimates, since mean differencing filters out some real signals in the data and leaves the within estimator more susceptible to noise.²⁹ To verify, I examine whether the relationship among the estimates using different transformations of the data (e.g., first difference vs. mean difference) matches the pattern Griliches and Hausman (1986) find analytically. The result is inconclusive: here, opposite to their finding, the longest (possible) difference generates an output coefficient slightly smaller than the within estimate, although long difference estimators do yield coefficients greater than those from the first difference. (Estimation results of versions F1 and V1 are reported in Table C.2.) One possible explanation for the discrepancy may simply be that my sample is an unbalanced panel, with individual BHC time series ranging from 12 to 44 observations. This means even the longest possible difference operator ($T_s - T_{s-11}, \forall s > 0$) will still generate larger bias than the within estimator for BHCs with long time series. Other potential reasons include measurement errors in inputs (and in turn input prices) and stochastic technology shocks that correlate with output, all of which affect the bias in the output coefficient.

two separate cost functions of the two services individually.

²⁹ The between estimate (based on BHC means) of γ^C falls to 1.13 (version F1, C-D), close to the pooled OLS estimate (1.14), which is a weighted average of the within and the between estimates.

One way to mitigate the bias in the output coefficient is to use instrumental variables for Y . I first experiment using three lags of output as IVs, and the GMM estimate of γ^C falls significantly from 2.5 to 1.6 ($\sigma = 0.09$, not shown), with the Hausman test rejecting the null that the two estimates are equal. Lagged Y s, however, may not be valid IVs here, given the likelihood of serial correlation in the error term. So I next use the two existing output measures as IVs, which Wang (2003b) shows to be correlated with the new Y (indeed $\rho = 0.95$). Using the SNA93 output as IV in a 2SLS estimation, the coefficient on Y rises to 0.67 ($\sigma=0.01$), and γ^C in turn falls to 1.2 ($\sigma=0.02$), close to the estimate from the cost function of the SNA93 output. When the BV-based outputs are used as IVs in a GMM estimation, Y 's coefficient rises to 0.65 ($\sigma=0.03$)— γ^C falls to 1.3 ($\sigma=0.07$). In both regressions, the Hausman test rejects the null of equality between the IV and the within estimates, but the over-identifying test in the second regression also rejects the null that the BV-based IVs are independent of the error term. Nonetheless, the two IV estimates together support the conjecture that the unusually high estimates of γ^C result more from noises or large stochastic shocks in the observed new output series than from true large economies of scale.³⁰

The overall conclusion is that measurement errors or large stochastic shocks in the imputed value of the new output series lead to a severe upward bias in the fixed-effects estimate of returns to scale from the value-added cost function, whereas the estimate based on the production function may be biased downward. The true degree of returns to scale is likely to lie between the two estimates, and closer to the latter. That is, returns to scale in the production of bank services are probably only somewhat higher than constant, implying that only modest cost savings can be expected from bank mergers through size increase. On the other hand, greater cost savings may be possible through the acquisition of less efficient banks by more efficient ones, if the finding of a high degree of productivity dispersion across BHCs is confirmed by additional data.

V. Conclusions

This paper estimates the value-added production and cost functions of BHCs using a new measure of bank output implied by a unified banking model and examines how the results differ from those estimated using the two existing measures of bank output. The banking model justifies estimating a value-added function by establishing that the production of bank value added is separable from the supply of loanable funds. In addition, its implied new output measure, unlike the existing ones, can coherently

³⁰ Other potential IVs for output (e.g., state industrial output, consumer expenditures, and housing starts, all proxies for the demand for bank services) are not explored here, partly because the main objective of this paper is to evaluate how the new output measure changes the cost function estimates, but also because several potential caveats apply. First, as explained in Appendix A, demand-side variables may be correlated with the error term too. Second, IVs that are macro variables often perform poorly in industry-level analysis, and they offer even lower explanatory power in firm-level analysis. Third, without adequate information to map accurately a bank's

account for risk in measuring banks' value added services. An output series computed according to the new measure is then found to have considerably different empirical properties from the existing output measures. All these make a comparison study necessary.

Several findings emerging from the estimations are worth noting. First, the fit of the cost function using the new output series appears to be as good as that using the existing output measures. Second, the estimates of returns to scale range from as low as 0.77 for the production function to as high as 4.5 for the cost function, while the most plausible range is 0.77 to 2.5, depending on the sources of errors in variables. Assuming that the major source of error is measurement errors in the new output measure, then the production function estimates of returns to scale should be subject to less bias than those from the cost function. Among different specifications of the latter, the estimates of about 2.5 (with a standard error of 0.2) from the single-output Cobb-Douglas cost function are directly comparable to those from the production function, and likely to be less biased. In contrast, high values from the Translog and the two-output Cobb-Douglas cost functions are not comparable, and so should be discounted as evidence of large increasing returns to scale. Moreover, two IV estimates of returns to scale from the cost function equal roughly 1.3. The preferred range is still quite wide, and one can neither rule out nor confirm the existence of increasing returns to scale. Further investigation will be necessary. By comparison, the existing book-value-based measure generates estimates of about constant returns to scale in my sample, as in almost all previous studies. Last, the new output measure results in a lower estimate of the industry-average productivity level (relative to the best practice) than the two existing output measures, and the fixed effects estimator also lowers the estimate relative to the methods commonly used in previous analyses.

The degree of returns to scale has implications for bank mergers and antitrust policy. If future research can confirm the existence of small increasing returns to scale, it will mean that mergers can potentially lead to modest cost savings, which in turn may induce lower prices, higher bank profits, or both. If the finding that the average bank has a low relative productivity level is confirmed, then mergers can increase the average productivity level of the banking industry if relatively efficient banks acquire less efficient ones and reorganize the targets to increase productivity. Of course, mergers will still have the potential downside of anticompetitive effects that reduce social welfare. The ability to evaluate potential cost savings from mergers is essential for bank regulators to balance the costs and benefits.

geographical presence, the efficacy of state-level variables becomes problematic.

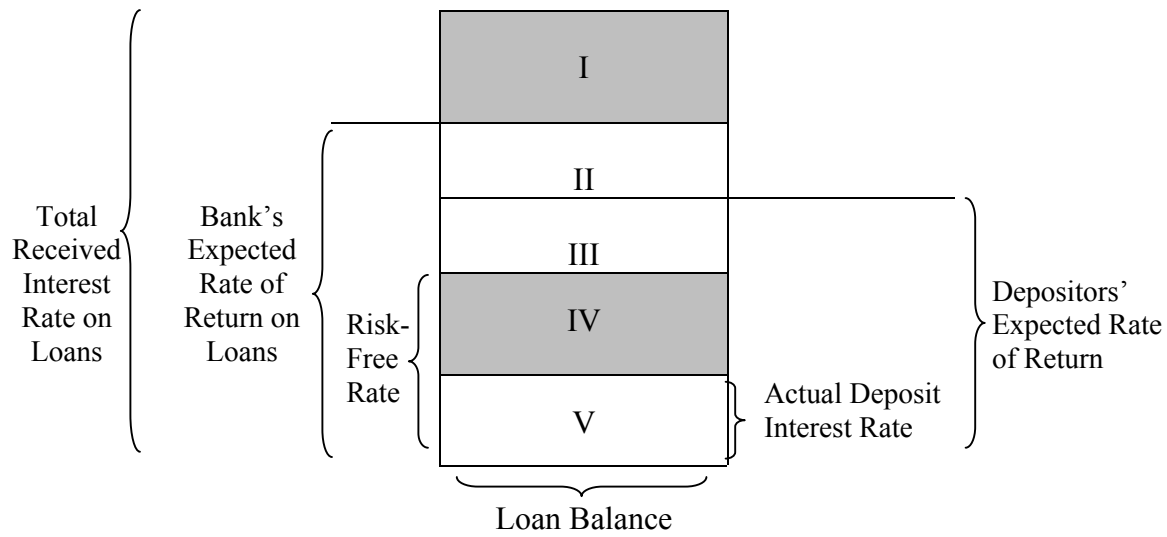


Diagram 1. Decomposition of a Bank's Total Interest Receipts

Notes:

1. The content of each area:

Area I: implicit fees for intermediation services in lending (e.g., origination and monitoring)

Area II: loan risk premium

Area III: deposit insurance premium

Area IV: implicit fees for transaction and payment services (e.g., mostly to depositors)

Area V: deposit interest payment

So,

Area (I + ... + V): a bank's total receipt of loan interest income

Area (II + ... + V): the bank's expected return on the funds given the loans' systematic risk

Area (III + IV + V): depositors' expected return on deposits, given the risk of the bank's loan portfolio (if without deposit insurance)

Area (IV + V): risk-free return \times deposit balance

N. B.: when there is equity, the two deposit-related rates on the right of the block and the risk-free rate should be adjusted by (deposit balance/loan balance).

2. The risk-free rate is the rate of return required by depositors given deposit insurance, whereas the "depositors' opportunity cost of capital" is the return they would demand without deposit insurance. The two rates should be very close (or the same) for banks with very low credit risk (say, having AAA-rated bonds outstanding).

Table 1. Output Measurement for Divisions A, B, and C of a Bank

Division	A	B	C
Function	Loan origination and monitoring services	Loanable funds	Depositor services
Attribute	Value Added	Gross Output	Value Added
Output (Y)	Nominal Value	The quarterly balance of 1. marketable securities, 2. each category of loans, such as C&I, real estate and consumer loans.	1. Explicit revenue for consumer services, 2. Implicit revenue = $(r^F \times \text{quarterly balance of demand, regular time and savings deposits}) - \text{the actual interest expenses on these deposit accounts.}$
	P. Idx.	GDP deflator	GDP deflator
Input	L: labor, variable input K: physical capital, quasi-fixed input	Loanable funds: D (various deposits) + E (equity capital, quasi-fixed input)	L: labor, variable input K: physical capital, quasi-fixed input k*D: cash reserves, quasi-fixed input

Notes:

1. For the output of Division B, I currently classify all the loans into three categories: (1) commercial loans, which include C&I loans and lease financing receivables, (2) real estate loans, and (3) consumer loans.
2. r^F : risk-free interest rate, using the 90-day Treasury bill rate.
3. The price index for Division C's output is compiled by BEA, and maintained in the I database.

Table 2. The Three Specifications of the Cost Functions, and the Corresponding Dependent and Independent Variables

Cost Function Specification	Value-Added (1)	Loanable Funds (2)	Book-Value-Based (Asset Approach)
Cost (dependent variable)	Total labor and capital cost: $wL + r^K K$	Total (imputed) interest cost: $r^F D + r^{FF}(FF) + r^B B$	Total interest and non-interest expenses: $wL + r^K K + r^D D + r^{FF}(FF) + r^B B$
Input Prices (independent variables)	w : wage r^K : rental p. of K	r^F : risk-free rate r^{FF} and r^B : the interest rates of FF and B , respectively	w, r^K, r^{FF}, r^B are the same as in (1) and (2), r^D : actual deposit interest rate paid
Outputs (independent variables)	Y^A : origination and monitoring Y^C : depositor services	RE : real estates loans CI : C&I loans and other short-term lendings $L. Ind.$: consumer loans Sec : market securities	The same as in (2)
Inputs	L : labor K : capital	D : Deposits FF : other liquid borrowings (Fed funds) B : non-liquid borrowings	L, K, D, FF and B are the same as in (1) and (2), as well as E -equity capital

Notes:

1. Inputs are included to help explain the input prices.
2. The column for the “value-added” cost function corresponds to versions V1 and V2 in Table 2. For versions F1 and F2, just replace r^K with K , and exclude capital cost ($r^K K$) from the dependent variable. Also, for the cost function according to the SNA93 output, just replace the new output Y ($= Y^A + Y^B$) with the SNA93 measure of output.

Table 3a. Fixed-Effects Within Estimates of the Cobb-Douglas Value-Added Production Function

		New Output		SNA93 Output	
		T. Dummy ^a	No T. Dummy	T. Dummy ^a	No T. Dummy
Parameter Estimates	Time		0.0016 (0.00042)		0.0023 (0.00035)
	L	0.480 (0.072)	0.455 (0.070)	0.498 (0.076)	0.500 (0.075)
	K	0.097 (0.025)	0.137 (0.027)	0.084 (0.026)	0.080 (0.026)
	Csh Res.	0.192 (0.020)	0.238 (0.020)	0.130 (0.020)	0.125 (0.019)
	Returns to Scale	0.77 (0.034)	0.83 (0.034)	0.71 (0.035)	0.70 (0.036)
Adjusted R ²		0.66	0.34	0.67	0.66

Notes:

1. Csh. Res.: cash reserves
 2. Standard errors of parameter estimates are in parentheses.
 3. A blank parameter value means it is not used in the regression.
- ^a: Estimation results with a time dummy for every quarter.

Table 3b. Common-Intercept OLS Estimates of the Cobb-Douglas Value-Added Production Function

		New Output		SNA93 Output	
		Cash Reserves as Input	No Cash Reserves	Cash Reserves as Input	No Cash Reserves
Parameter Estimates	Intercept	0.807 (0.115)	2.515 (0.085)	1.990 (0.125)	3.115 (0.073)
	L	0.551 (0.032)	0.977 (0.026)	0.733 (0.035)	1.014 (0.023)
	K	0.040 (0.015)	0.054 (0.023)	0.006 (0.016)	0.015 (0.021)
	Csh Res.	0.412 (0.024)		0.271 (0.023)	
	Returns to Scale	1.00 (0.006)	1.03 (0.007)	1.01 (0.005)	1.03 (0.005)
Adjusted R ²		0.93	0.92	0.97	0.96

Notes:

1. Csh. Res.: cash reserves
2. Standard errors of parameter estimates are in parentheses.
3. A blank parameter value means it is not used in the regression.

Table 4. Within Estimates of the Translog Value-Added Production Function

		Fixed-effects (within estimation)	
		New Output	SNA93 Output
Parameter Estimates	L	0.549 (0.023)	0.568 (0.021)
	K	-0.016 (0.055)	-0.173 (0.043)
	Csh Res.	0.053 (0.062)	-0.095 (0.050)
	L ²	-0.122 (0.010)	-0.127 (0.011)
	K ²	0.040 (0.011)	0.020 (0.007)
	CR ²	0.036 (0.011)	0.015 (0.009)
	L*K	0.031 (0.023)	0.055 (0.022)
	L*CR	0.026 (0.031)	-0.001 (0.027)
	K*CR	-0.026 (0.010)	0.008 (0.007)
	Returns to Scale	0.81 (0.015)	0.76 (0.015)
	Adjusted R ²	0.66	0.69

Notes:

1. Csh. Res. (CR): cash reserves
2. Standard errors of parameter estimates are in parentheses.

Table 5a. Within Estimates of the Cobb-Douglas Value-Added Cost Function

		Single Output, Quasi-fixed K	
		New Output	SNA93 Output
Parameter Estimates	Y	0.221 (0.012)	0.530 (0.023)
	Wage	1.000	1.000
	K	0.262 (0.014)	0.189 (0.012)
	Csh Res.	0.180 (0.010)	0.114 (0.008)
	Adjusted R ²	0.82	0.85
Returns to Scale	2.53 (0.18)	1.32 (0.08)	
Avg. Efficiency [#]	0.43	0.61	

Notes:

1. Csh. Res. (CR): cash reserves
2. Standard errors of parameter estimates are in parentheses, except for those constrained by the cost function (e.g., $\beta_L = 1$), which are denoted by a blank value.
3. A blank parameter value means it is not used in the regression.

[#]: output-weighted sample average efficiency level, excluding the observation outside the 5th and the 95th percentiles to minimize the impact of outliers (applicable to all the sample average efficiency levels reported in the following tables)

Table 5b. Alternative C-D Value-Added Cost Functions for the New Output (Within Estimates)

		Single Output	Two Outputs (Y ^A and Y ^C)	
		Variable K (Version V1)	Quasi-fixed K (Version F2)	Variable K (Version V2)
Parameter Estimates	Y ^A	0.315 (0.015)	0.058 (0.004)	0.085 (0.005)
	Y ^C		0.050 (0.004)	0.069 (0.005)
	Wage	0.830 (0.026)	1.000	0.835 (0.026)
	r^K	0.170 (0.026)		0.165 (0.026)
	K		0.291 (0.015)	
Csh Res.	0.256 (0.014)	0.215 (0.012)	0.325 (0.016)	
Adjusted R ²	0.77	0.81	0.74	
Returns to Scale	2.36 (0.13)	4.56 (0.33)	4.37 (0.25)	
Avg. Efficiency	0.32	0.38	0.27	
Cost	$wL + r^K K$	wL	$wL + r^K K$	

Notes:

1. The estimate of γ^C from versions F2 is also adjusted according to (A.8) in Appendix A.
2. All the notes for Table 5a apply.

Table 6. Within Estimates of the Translog Value-Added Cost Function

	Single Output, Quasi-fixed K	
	New Output	SNA93 Output
Y	0.182 (0.060)	0.701 (0.101)
wage (w)	1.000	1.000
K	-0.250 (0.112)	-0.175 (0.087)
Csh Res.	0.353 (0.112)	0.298 (0.102)
Y ²	0.041 (0.019)	-0.013 (0.048)
w ²	0.000	0.000
K ²	0.063 (0.012)	0.053 (0.012)
CR ²	0.027 (0.013)	0.009 (0.013)
wK	0.078 (0.062)	0.074 (0.059)
w(CR)	-0.115 (0.063)	-0.114 (0.060)
K(CR)	-0.008 (0.011)	-0.015 (0.012)
wY	0.000	0.000
KY	-0.014 (0.015)	-0.013 (0.025)
(CR)Y	-0.017 (0.015)	0.010 (0.022)
Adjusted R ²	0.86	0.90
Returns to Scale*	4.32 (0.51)	1.75 (0.20)
Returns to Scale [#]	4.35	1.75
Avg. Efficiency	0.43	0.65

Notes:

1. Standard errors of parameter estimates are in parentheses.
2. A blank parameter value means it is not used in the regression, and a blank standard error value only means the parameter value is constrained.
3. In the Translog cost function, the estimates of γ^C also depend on data. So I compute γ^C (1) at the sample mean, (2) at every observation, and then average over the sample. I also check the average of γ^C estimated at the mean of each BHC. The results are typically all very close.

*: Returns to scale calculated at sample mean

[#]: Sample average of returns to scale calculated at each observation point

Table 7. Within Estimates of the Translog Book-Value-Based Cost Function
(First-Order and Second-Order Terms)

	Quasi-fixed K		Variable K	
	Coefficient	Standard Error	Coefficient	Standard Error
Sec. (Y ₁)	0.286	0.041	0.257	0.044
L. Ind. (Y ₂)	0.041	0.038	0.043	0.043
C&I (Y ₃)	0.282	0.063	0.290	0.063
RE (Y ₄)	0.287	0.063	0.305	0.074
Eq. (X ₁)	-0.203	0.093	-0.126	0.104
Dep. R. (X ₂)	0.578	0.095	0.548	0.105
wage (X ₃)	0.047	0.095	-0.006	0.105
r^K (X ₄)			0.012	0.038
K (X ₄)	-0.022	0.065		
FF. R. (X ₅)	0.375	0.143	0.446	0.156
Sec. ²	0.086	0.006	0.090	0.006
L. Ind. ²	0.033	0.005	0.031	0.007
C&I ²	0.044	0.011	0.048	0.012
RE ²	0.098	0.014	0.108	0.015
Eq. ²	0.033	0.018	0.047	0.019
Dep. R. ²	0.280	0.045	0.264	0.046
wage ²	0.043	0.013	0.037	0.011
$(r^K)^2$			0.005	0.004
K ²	0.046	0.011		
FF. R. ²	0.040	0.065	0.018	0.059
Adjusted R ²		0.89		0.88
Returns to Scale*		1.30		1.18
		(0.09)		(0.08)
Returns to Scale [#]		1.37		1.21
		(0.07)		(0.05)
Avg. Efficiency		0.72		0.68

Notes:

1. Notations:

- Sec.: Investment securities
- L. Ind.: consumer loans (loans to individuals)
- C&I: commercial and industrial loans
- RE: real estate loans
- Eq.: equity
- Dep. R.: deposit interest rate
- FF. R.: Federal funds rate

2. A blank parameter value means it is not used in the regression.

*: Returns to Scale calculated at sample mean

#: Sample average of returns to scale calculated at each observation point, weighted by total financial assets and excluding observations outside the 5th and the 95th percentiles. Standard errors are in parentheses.

Table 7. (continued)
 Estimation Results of the Translog Book-Value-Based Cost Function (Cross-Product Terms)

	Quasi-fixed K		Variable K	
	Coefficient	Standard Error	Coefficient	Standard Error
Y ₁ Y ₂	-0.003	0.005	-0.006	0.005
Y ₁ Y ₃	-0.034	0.003	-0.037	0.004
Y ₁ Y ₄	-0.051	0.007	-0.058	0.007
Y ₂ Y ₃	-0.005	0.004	0.001	0.006
Y ₂ Y ₄	0.010	0.006	0.011	0.007
Y ₃ Y ₄	-0.039	0.011	-0.042	0.010
X ₁ X ₂	0.032	0.026	0.024	0.026
X ₁ X ₃	-0.010	0.019	-0.018	0.018
X ₁ X ₄	0.000	0.011	-0.019	0.008
X ₁ X ₅	0.016	0.022	0.012	0.023
X ₂ X ₃	-0.142	0.025	-0.124	0.026
X ₂ X ₄	-0.016	0.019	-0.036	0.014
X ₂ X ₅	-0.138	0.050	-0.104	0.047
X ₃ X ₄	0.003	0.013	0.016	0.008
X ₃ X ₅	0.098	0.027	0.071	0.027
X ₄ X ₅	-0.023	0.014	0.015	0.012
Y ₁ X ₁	0.016	0.010	0.013	0.010
Y ₁ X ₂	0.043	0.014	0.030	0.015
Y ₁ X ₃	-0.038	0.008	-0.040	0.009
Y ₁ X ₄	-0.018	0.007	0.011	0.007
Y ₁ X ₅	-0.005	0.014	-0.001	0.015
Y ₂ X ₁	-0.025	0.008	-0.033	0.009
Y ₂ X ₂	-0.016	0.009	-0.012	0.010
Y ₂ X ₃	0.000	0.006	0.004	0.007
Y ₂ X ₄	-0.006	0.006	-0.002	0.004
Y ₂ X ₅	0.015	0.008	0.010	0.009
Y ₃ X ₁	0.016	0.011	0.027	0.011
Y ₃ X ₂	0.002	0.010	0.005	0.010
Y ₃ X ₃	-0.003	0.006	-0.002	0.006
Y ₃ X ₄	0.014	0.009	-0.004	0.005
Y ₃ X ₅	0.001	0.008	0.000	0.008
Y ₄ X ₁	-0.019	0.012	-0.031	0.011
Y ₄ X ₂	-0.031	0.017	-0.027	0.018
Y ₄ X ₃	0.050	0.010	0.058	0.010
Y ₄ X ₄	-0.013	0.006	0.008	0.005
Y ₄ X ₅	-0.019	0.015	-0.038	0.014

Notes: Notations in this table match the ones in the table on the last page as follows:

Y ₁ : Sec.	Y ₂ : L. Ind.	Y ₃ : C&I
Y ₄ : RE	X ₁ : Eqt.	X ₂ : Dep. R.
X ₃ :wage	X ₄ : r^K or K	X ₅ : FF. R.

Appendix A. Properties of the Production and Cost Function Parameter Estimates

I use a simple Cobb-Douglas production function to illustrate the properties of its error term and the implications for the properties of its coefficient estimates. Assume the true production function (in logs) of bank i in period t is:

$$y_{it}^* = a_i + \beta_l^* l_{it}^* + \beta_k^* k_{it}^* + \varepsilon_{it}^{y^*}, \quad (\text{A.1})$$

where y_{it}^* is the correctly measured real output, and k_{it}^* and l_{it}^* are the true capital and labor inputs, respectively. a_i is the bank fixed effects, and $\varepsilon_{it}^{y^*}$ is the real productivity shock, assumed to be i.i.d. but possibly non-spherical. So, the true γ^Y is $\gamma_y^* = \beta_l^* + \beta_k^*$. The true inputs k^* and l^* can be expressed as³¹

$$l^* = e + h + n, \quad \text{and} \quad k^* = u + \kappa, \quad (\text{A.2})$$

where e is labor effort, h hours worked, n the number of workers, u capital utilization, and κ capital stock.³² e , h , and u are not observed in banking data, and κ is measured imprecisely. Moreover, the adjustment costs associated with n and κ are expected to be far greater than the cost of adjusting h and u . Hence, in response to a shock, banks are likely to adjust h and u before changing n and κ .

Accordingly, the production function that can be estimated with observed variables is:

$$y = a_i + \beta_l l + \beta_k k + \varepsilon_y, \quad (\text{A.3})$$

$$y = y^* + \eta_y, \quad l = n, \quad \text{and} \quad k = \kappa + \eta_k, \quad (\text{A.4})$$

where y , l , and k are the observed output, labor, and capital, respectively. η_y and η_k are measurement errors in output and capital stock respectively, all assumed to be i.i.d. (A.3) can be written as

$$y = a_i + \beta_l^* n + \beta_k^* (\kappa + \eta_k) + [\varepsilon^{y^*} + \eta_y - \beta_k^* \eta_k + \beta_l^* (h + e) + \beta_k^* u], \quad (\text{A.5})$$

where the composite error term (ε_y) contains four components: (1) technology shocks (ε^{y^*}), (2) classical measurement errors in input (η_k), (3) unobserved input variations (h , e , and u), and (4) classical measurement errors in Y (η_y). Clearly, only component (4) (η_y) is uncorrelated with inputs. Input measurement errors ($-\beta_k \eta_k$) will bias β_k toward zero. In contrast, the unobserved input variations will bias β_l and β_k upward if positively correlated with observed inputs; i.e., $\text{cov}(h, n) > 0$, $\text{cov}(e, n) > 0$, and $\text{cov}(u, \kappa) > 0$. In this case, the net bias of (2) and (3) combined is ambiguous. In the banking context, these contemporaneous correlations are likely to be low, owing to the quasi-fixity of n and κ , unless there are strong serial correlations in ε^{y^*} or demand shocks. For instance, in response to a stochastic demand shock, a bank is likely to increase just h and u within one period without changing n and κ . If it does so, the net bias caused by (2) and (3) will be negative. But if the shocks are persistent, then h (u)

³¹ In the following discussion, for clarity of exposition, I omit the subscript it where no confusion is likely.

³² I do not explicitly consider the quality of labor and capital here, because it is less a concern with high frequency

and n (κ) may co-move in the periods when n is also being adjusted, resulting in ambiguous net bias. This also implies that demand-side variables, possibly correlated with the error term, may be invalid instruments. As for ε^{y^*} , it can be further decomposed into a permanent part—reflecting the long-term average productivity level—and a stochastic part: $\varepsilon^{y^*} = u^{y^*} + \zeta^{y^*}$. When longitudinal data are available, u^{y^*} can be estimated as bank-specific “fixed effects.” This should solve the endogeneity problem stemming from possible correlations between inputs and the long-term productivity differentials across banks. But ζ^{y^*} may still be correlated with inputs, and the bias can go in either direction, depending on the time frame.³³ In banking, over the short run, we may expect a negative correlation between utilization and technology shocks, because banks supply a service that cannot be stored, and n and κ are quasi-fixed. Over the long run, if the shocks are persistent, they will be positively correlated with inputs. In summary, the net sign of the parameter bias is more likely to be negative, but it is essentially undetermined in theory and thus an empirical matter.

For the Cobb-Douglas functional form, the cost and production functions are self-duals: the former’s coefficients are exclusively functions of the latter’s, given correctly measured variables. If both K^* and L^* are treated as variable inputs, then the true cost function can be written as:

$$\ln C^*(y^*, \ln w^*, \ln r^{K^*}) = \min_{L, K} \ln(w^* L^* + r^{K^*} K^*) = A_i + \theta_y^* y^* + \theta_w^* \ln w^* + \theta_r^* \ln r^{K^*} + \varepsilon_c^*,$$

$$\text{i.e.,} \quad c^*(y^*, \ln w^*, \ln r^{K^*}) = A_i + (y^* + \beta_l^* \ln w^* + \beta_k^* \ln r^{K^*} - \varepsilon_y^*) / \gamma_y^*, \quad (\text{A.6})$$

where $A_i = -(a_i + \beta_l^* \ln \beta_l^* + \beta_k^* \ln \beta_k^*) / \gamma_y^*$, w^* is the wage rate, and r^{K^*} is the rental price of capital. The true γ^C defined by $\gamma_c^* = 1 / (\partial \ln C^* / \partial \ln Y^*)$ is equal to γ_y^* in the case of $c^*(\cdot)$. (A.6) also makes it clear that the output coefficient in the cost function is inherently biased toward zero, as y^* is negatively correlated with $(-\varepsilon_y^*)$. In fact, if $\gamma^* = 1$, the error terms in the cost and production functions are exactly opposite: $\varepsilon_c^* = -\varepsilon_y^*$. This means that if the primal and the dual functions were to be jointly estimated, the cross-equation restriction on the error terms would limit the degrees of freedom of the system to be the same as for either equation estimated individually.

When K^* is treated as a quasi-fixed input, the short-run variable cost function is

$$\ln C^{S^*}(y^*, \ln w^*, k^*) = \min_{L} \ln(w^* L^*) = (-a_i + y^* + \beta_l^* \ln w^* - \beta_k^* k^* - \varepsilon_y^*) / \beta_l^*, \quad (\text{A.7})$$

In this case, $\gamma_c^* = \beta_l^* \neq \gamma_y^*$, and the definition of γ_c^* must be adjusted to be consistent with γ_y^* :

data and a short time period, and it can be represented as one element in the BHC fixed effects.

³³ See Basu et al. (1998) for an account of the short-term negative correlation between technology shocks and inputs. In the long run, however, it is generally believed that inputs move in the same direction as technology

$$\gamma_{c^s}^* = (1 - \partial \ln C^S / \partial \ln K^*) / (\partial \ln C^S / \partial \ln Y^*) = \gamma_y^*. \quad (\text{A.8})$$

The cost function estimated using observed variables can be written as:

$$\begin{aligned} c(y, \ln w, \ln r^K) = & A_i + \theta_y(y^* + \eta_y) + \theta_w(\ln w^* + e + h) + \theta_r(\ln r^{K^*} + u + \eta_r - \eta_k) \\ & + [\eta_c + \varepsilon_c^* - \theta_y \eta_y - \theta_w(e + h) - \theta_r(u + \eta_r - \eta_k)], \end{aligned} \quad (\text{A.9})$$

where the observed input prices are computed using input expenditures and quantities, the measurement errors of which are shown in (A.4):

$$\ln w = \ln w^* + e + h, \quad \text{and} \quad \ln r^K = \ln r^{K^*} + u + \eta_r - \eta_k. \quad (\text{A.10})$$

(A.9) illustrates that the structure of the dual cost function's error term mirrors that of the production function. There are four components in the composite error term (ε_c): (1) technology shocks (ε_c^*), (2) derived measurement errors in input prices (η_r , η_k , h , e , and u), where η_r is the error in the observed capital expense, (3) classical measurement errors in Y (η_y), and (4) measurement errors in C (η_c). η_y biases the output coefficient (θ_y) to zero, and so may ε_c^* , although its correlation with y can also be zero, as discussed above. η_c may also bias θ_y downward if, instead of being classical measurement errors, η_c results from deviations between an observed input price and the true shadow price when the input (say, L) is quasi-fixed but is treated as variable in the cost function specification. It is less clear how (2) may bias θ_y if η_r arises from deviations between reported and true capital expenses—possibly downward. Overall, the fact that biases due to (1), (3), and (4) have the same sign offers a plausible reason why estimates of γ^C generally tend to be far greater than γ^Y , but the exact size of the bias in θ_y is ambiguous when there are errors in input prices (possibly to a lesser extent) as well.

A similar error structure holds for the short-run variable cost function:

$$\begin{aligned} c^S(y, \ln w, k) = & A_i + \theta_y(y^* + \eta_y) + \theta_w(\ln w^* + e + h) + \theta_k(\kappa + \eta_k) \\ & + [\eta_c + \varepsilon_{c^s}^* - \theta_y \eta_y - \theta_w(e + h) + \theta_k(u - \eta_k)]. \end{aligned} \quad (\text{A.11})$$

The preceding discussion of the error term structure and the implications for the properties of the parameter estimates applies to the Translog production and cost functions as well, except that the closed-form mapping between the two functions' error terms no longer holds, because the Translog production and cost functions are not self-duals.

Appendix B. Data Items Used to Construct Cost, Inputs and Input Prices

Cost Function Specification	Value-Added (1)	Loanable Funds (2)	Book-Value-Based (Asset Approach)	
Cost	Def'n	Total labor and capital cost: $wL + r^K K$	Total (imputed) interest cost: $r^F D + r^{FF}(FF) + r^B B$	Total interest and non-interest cost: $wL + r^K K + r^D D + r^{FF}(FF) + r^B B$
	Data Items	wL : 4135 $r^K K$: 4217	$r^{FF}(FF)$: 4180 $r^B B$: 4396+4397+4398	$wL + r^K K$: 4093 $r^D D + r^{FF}(FF) + r^B B$: 4107
Input	Def'n	L : labor K : capital	D : Deposits FF : other liquid borrowings (Fed funds) B : non-liquid borrowings	L, K, D, FF and B are the same as in (1) and (2)
	Data Items	L : 4150 K : 2145	D : 6636+6631 FF : 2800 B : 2332+2333+4062+2910	ibid.
Input Price	Def'n	w : wage r^K : rental p. of K	r^F : risk-free rate r^{FF} and r^B : the interest rates of FF and B , respectively	ibid. r^D : actual deposit interest rate paid
	Data Items	---	r^F : 90-day T-bill rate	---

Notes:

- Notations:
All the data items not listed are terms derived from the corresponding costs, inputs, or input prices.
- Finer distinctions:
 - Strictly speaking, r^F should apply only to deposits less than \$10,000, and the appropriate rate for those over \$100,000 (item 2604) should be greater than r^F , so I use actual interest payment to deposits in item 2604.
 - To be precise, the costs as defined in the asset approach also contains costs incurred in purchasing materials (such as paper), and excludes the explicit fee income from depositor services (which is considered to offset direct funding cost).
- Data Items:
 - 4135: salaries and benefits of all officers and employees of the bank and its consolidated subsidiaries
 - 4217: all non-interest expenses related to the use of premises, equipment, furniture, and fixtures
 - 4150: number of (full-time equivalent) workers
 - 2145: net stock of physical capital
 - 4093: total non-interest expenses
 - 4107: total interest expenses
 - 4180: gross expense of all liabilities included in federal funds purchased and securities sold under agreements to repurchase
 - 4396: interest expense on borrowed funds other than subordinated debt

- 4397: interest expense on subordinated debt
- 4398: all other interest expenses
- 6636: total interest-bearing deposits in domestic and foreign offices
- 6631: total non-interest-bearing deposits in domestic and foreign offices
- 2800: federal funds purchased and securities sold under agreements to repurchase
- 2332: other borrowed money with maturity less than one year
- 2333: other borrowed money with maturity greater than one year
- 4062: subordinated debt
- 2910: mortgage debt (generally of very small relative magnitude)

Appendix C. Alternative Specifications of the Cost function, and Additional Estimation Results

Table C.1 Four Different Versions of the Value-Added Cost Function

		Number of Outputs	
		1	2
Property of Physical Capital	Variable	V1	V2
	Quasi-fixed	F1	F2

Table C.2 Fixed-Effects Difference Estimates of the Cobb-Douglas Value-Added Cost Function

		First Difference		Long Difference ($Y_t - Y_{t-11}$)	
		Quasi-fixed K (Version F1)	Variable K (Version V1)	Quasi-fixed K (Version F1)	Variable K (Version V1)
Parameter Estimates	Time	0.0065 (0.00033)	0.0112 (0.00031)	0.0041 (0.00011)	0.0075 (0.00013)
	Y	0.0496 (0.00302)	0.0706 (0.00278)	0.0919 (0.00317)	0.1516 (0.00354)
	L	1.0000	0.7556 (0.00172)	1.0000	0.8484 (0.00363)
	K (or r^K)	0.2622 (0.00533)	0.2444 (0.00172)	0.2811 (0.00384)	0.1516 (0.00363)
	Csh Res.	0.0418 (0.00258)	0.0581 (0.00238)	0.1982 (0.00417)	0.2998 (0.00455)
Returns to Scale		14.03 (2.15)	13.35 (1.49)	5.67 (0.33)	4.62 (0.20)
Adjusted R^2		0.62	0.65	0.81	0.74

Notes:

1. Csh. Res.: cash reserves
2. All the standard errors of parameter estimates are in parentheses. A blank standard error value means the parameter value is constrained by the cost function (e.g., $\beta_L = 1$).
3. A blank parameter value means it is not used in the regression.

Bibliography

- Arellano, M (1987). "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Economics and Statistics* 49(4), p. 431-34.
- Basu, S., J. G. Fernald and M. Kimball (1998). "Are Technology Improvements Contractionary?" *International Finance Discussion Papers, Board of Governors of the Federal Reserve System*, Washington, DC.
- Berger, A. N., G. A. Hanweck and D. B. Humphrey (1987). "Competitive Viability in Banking: Scale, Scope and Product Mix Economies," *Journal of Monetary Economics* 20 (3), p. 501-20.
- Berger, A. N. and D. B. Humphrey (1997). "Efficiency of Financial Institutions: International Survey and Directions for Future Research," *European Journal of Operational Research* 98(2), p. 175-212.
- Berger, A. N. and L. J. Mester (1997). "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance* 21(7), p. 895-947.
- Berndt, E. R. (1991). *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley.
- Bhargava, A., L. Franzini and W. Narendranathan (1982). "Serial Correlation and the Fixed Effects Model," *Review of Economic Studies* 49(4), p. 533-49.
- Clark, J. (1996). "Economic Cost, Scale Efficiency and Competitive Viability in Banking," *Journal of Money, Credit and Banking* 28(3), p. 342-64.
- Dixon, B., P. Garcia and M. Anderson (1987). "Usefulness of Pretests for Estimating Underlying Technologies Using Dual Profit Functions," *International Economic Review* 28(3), p. 623-33.
- Froot, K. A. and J. C. Stein (1998). "Risk Management, Capital Budgeting and Capital Structure Policy for Financial Institutions: An Integrated Approach," *Journal of Financial Economics* 47(1), p. 55-82.
- Greene, W. H. (1997). *Econometric Analysis*, Prentice Hall, New Jersey.
- Griliches Z. and J. A. Hausman (1986). "Errors in Variables in Panel Data," *Journal of Econometrics* 31(1), p. 93-118.
- Hall, R. E. (1986). "Chronic Excess Capacity in U.S. Industry," *NBER Working Paper* #1973.
- Hughes, J. P. and L. J. Mester (1993). "A Quality and Risk-Adjusted Cost Function for Banks: Evidence on the Too-Big-To-Fail doctrine," *Journal of Productivity analysis* 4(3), p. 293-315.
- Hughes, J. P., W. Lang, L. J. Mester and C.-G. Moon (1996). "Efficient Banking under Interstate Branching," *Journal of Money, Credit and Banking* 28(4), p. 1044-71.
- Hughes, J. P. and L. J. Mester (1998). "Bank Capitalization and Cost: Evidence of Scale Economies in Risk Management and Signaling," *Review of Economics and Statistics* 80(2), p. 314-25.
- Klette, T. J. and Z. Griliches (1996). "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," *Journal of Applied Econometrics* 11(4), p. 343-61.
- Marschak, J. and W. Andrews (1944). "Random Simultaneous Equations and the Theory of Production," *Econometrica* 12(1), p. 143-205.
- Mitchell, K. and N. M. Onvural (1996). "Economies of Scale and Scope at Large Commercial Banks: Evidence from the Fourier Flexible Functional Form," *Journal of Money, Credit, and Banking* 28(1), p.179-199.
- Moulton, B. R and E. P. Seskin (2003). "Preview of the 2003 Comprehensive Revision of the National Income and Product Accounts, Changes in Definitions and Classifications," *Survey of Current Business*, June, p. 17-34.
- Rogers, K. E. (1998). "Nontraditional Activities and the Efficiency of US Commercial Banks," *Journal of Banking and finance* 22(4), p. 467-82.
- Sbordone, A. M. (1997). "Interpreting the Procyclical Productivity of Manufacturing Sectors: External Effects or Labor Hoarding?" *Journal of Money, Credit, and Banking* 29(1), p. 26-45.
- Stiroh, K. (1999). "How Did Bank Holding Companies Prosper in the 1990s?" *Journal of Banking and Finance*, forthcoming.

- Wang, J. C. (2003a). "Loanable Funds, Risk, and Bank Service Output," Federal Reserve Bank of Boston Working Paper, 03-4.
- Wang, J. C. (2003b). "Service Output of Bank Holding Companies in the 1990s, and the Role of Risk," Federal Reserve Bank of Boston Working Paper, 03-6.