Risk Bearing, Implicit Financial Services, and Specialization in the Financial Industry

J. Christina Wang and Susanto Basu

Abstract:
What is the output of financial institutions? And how can we measure their nominal and, more importantly, real value, especially since many financial services are provided without explicit charges? This paper summarizes the theoretical result that, to correctly impute the nominal value of implicit financial service output, the "user cost of money" framework needs to be extended to take account of the systematic risk in financial instruments. This extension is easy to implement in principle: One can continue using the current imputation procedure, and the only change needed is to adjust the reference rates of interest for risk.

The paper clarifies why the risk-related income is not part of the output—or equivalently, why risk bearing is not a service—of financial institutions. The paper next argues that, to measure real output, one must first explicitly specify and define the economic services produced by financial firms, a step that is absent from the "user cost of money" theory. Once it is established that only financial services, and not instruments, should be counted as the value added of financial firms, it follows that the quantity of services provided by these institutions is not necessarily in fixed proportion to the volume of instruments. The corollary is that the implicit price of financial services bears no definitive relationship with any reference rate. Instead, price deflators for financial services should be constructed using methods similar to those used for other services.

J. Christina Wang is an Economist at the Federal Reserve Bank of Boston. Susanto Basu is a Professor in the economics department at Boston College, a research associate at the NBER, and a visiting scholar at the Federal Reserve Bank of Boston. Their email addresses are christina.wang@bos.frb.org and susanto.basu@bc.edu, respectively.

This version: November 21, 2005

1. Introduction

The definition and measurement of financial institutions' output and prices is a difficult issue that has seen much debate and little consensus. It is made even more challenging by rapid and massive changes in the form of financial organizations and by the range and features of financial instruments they offer. Hence, it is hardly surprising that conceptual disagreement remains regarding even the measurement of *nominal* output of financial services, let alone *real* output.[1]

In this overview, we summarize and extend the new measure of financial service output developed in our previous papers (Wang, 2003a; Wang, Basu, and Fernald [WBF], 2004) that resolves the extant conceptual disparities. In short, we make three points. Two of them concern the measurement of *nominal* financial output, especially banking output. First, we show that, to impute implicitly priced financial output, it is necessary to adjust each reference rate of interest (equal to "the user cost of funds") for the risk inherent in that corresponding financial transaction. Otherwise, nominal financial output will be overstated, and the bias can be large (near 25 percent, based on a data set of all bank holding companies; see Wang, 2003b). Second, we argue that, according to finance theory, the required risk correction can be implemented practically at the level of industries (for example, the banking sector as a whole). The third point concerns the construction of a financial services price index and thus applies to the measurement of *real* output. We argue that the reference rates or the related rate spreads used to impute the nominal output of financial institutions (FIs) are not the right implicit price deflators for deriving the real output of FIs.

These conclusions follow from applying two principles to analyzing the operations of FIs. The first principle is that one must understand the market context in which FIs operate, particularly how financial markets set the rates of return on (and thus prices of) risky financial products. A large modern literature on asset pricing under uncertainty gives us the tools needed to choose the variables—the "reference rates"—that are key to the measurement of nominal financial output. FIs receive income both from holding financial products and from

---

[1] Triplett and Bosworth (2004, ch. 7) provide a clear summary and critique of several of the existing measures of bank output and discuss their preferred measure.

producing financial services. Once the reference rates have been determined, we can impute the returns to FIs' asset ownership—the income to the financial products. With total income observed, the nominal value of FIs' output—financial services—is then measured as the residual income.

Knowing the reference rates for risky assets suffices for measuring nominal output. But to measure real output, one must also understand what exactly are financial firms' products. Hence, the second principle is that one must be precise about the definition of "financial services"—what it is that FIs actually produce. In our framework, a "service" produced by an FI is like any other professional service, such as accounting or consulting: It is a flow of output that is valued by customers and created through a production process using real resources. The only difference is that an FI routinely receives compensation for these services implicitly in the form of interest margins—the spread between interest received and interest paid. For example, an FI may charge a borrower for the service of loan screening via a higher interest rate than would be required for just the use of the funds. Such bundling makes it difficult to distinguish between the financial *instruments* and the financial *services*, and so makes it necessary to emphasize the distinction.[2]

In fact, it is exactly the implicit compensation and bundling that make measuring bank output so difficult, especially since many FIs rely on implicitly-priced services as a primary source of income. For instance, practically all banks have non-interest operating expenses that exceed non-interest income and rely on positive net interest income to cover the difference. Hence, it is clear that national income accounting's usual practice of treating interest income as a mere transfer of property income is not generally applicable to FIs. The System of National Accounts 1993 (SNA93) thus recommends regarding net interest as the value of "financial intermediation services indirectly measured" (FISIM). In accordance, the U.S. national income and product accounts (NIPA) equate the *entire* interest margin (net of the returns on own funds) to the nominal value of implicit financial services. This amounts to assuming implicitly that any interest rate differential exists solely because a financial firm uses the interest margin as an

---

[2] Also see, for example, Fixler et al. (2003), Fixler (2005), and Schreyer and Stauffer (2003).

indirect means of charging for services provided.

Furthermore, since the 2003 comprehensive revision, the NIPA also apply a single reference rate to divide the entire spread into two parts, and explicitly identify the two interest *rate differentials* as implicit *prices* paid for the implicit financial services. Specifically, the differential between the loan interest rate and the reference rate is viewed as the price of services to borrowers, while the differential between the reference rate and the deposit interest rate is the price of depositor services (Fixler et al. 2003).

This procedure is usually justified by reference to models of the "user cost of money," defined as the interest rate on an asset that is held solely for the return it provides and not associated with any other financial services. In practice, the U.S. NIPA choose an average Treasury rate as the reference rate—the user cost of money—because, it is argued, the Treasury securities are free from default risk and (almost) costlessly accessible to all.

While the user-cost principle is appropriate, the current criterion for selecting the reference rate cannot pin down a single interest rate that has a robust claim to being *the* reference rate. Some may justify using the interest rate on 90-day U.S. Treasury bills by arguing that one could buy newly issued 90-day bills and hold them until maturity. But one could equivalently buy and hold 10-year U.S. Treasury bonds over the same 90-day period. Backed by the credit of the same issuer, both are free of default risk over the same (short) period of time, and they are equally easy to buy and sell, that is, equally liquid. But, on average, holding 10-year bonds for 90 days will yield higher returns. This is true because the value of the bond may change over the 90-day period as long-term interest rates change, exposing the owner to the risk of capital loss. So the Treasury bonds have to offer a higher interest rate to compensate the owners for this price risk. Likewise, one could hold Microsoft stock for 90 days instead. Its expected (average) return is higher still than that on the Treasury bonds over 90 days, as its rate of return is even more volatile: It is riskier.

In general, it is clear that there are many assets held solely for their returns, but—even correcting for any probability of default—they carry different expected interest rates because they have different risk characteristics. That is, simply identifying an asset as held for its "pure return" is not enough: One needs to specify the risk of each asset to set its rate of return.

4

Formally, the issue is that to pin down any security's interest rate the user-cost framework needs to be extended to take account of risk. This issue has recently been recognized by one of the pioneers of the user-cost models; Barnett et al. (1995), for example, argue that one needs to incorporate a risk premium into the user-cost formula in order to infer the extent of liquidity services provided by a risky asset. The basic concept of a user cost of funds, and the corollary that one can infer the nominal output of FIs by looking at interest-rate spreads (multiplied by the dollar volume of transactions) provide a sound starting point. But in order to know the relevant interest-rate spread, one has to know the risk category of the "reference asset." Modern asset-pricing theories state clearly that the user cost of funds—in the terminology of asset pricing, the "required rate of return" on those funds—is a function of the (systematic) risk that investors face by holding those funds. See, for example, Cochrane (2001).

Our conclusion, presented in Wang (2003a) and WBF (2004) is that in a world with a multitude of risks there are many reference rates; each rate is chosen to equal the user cost of funds *subject to the same risk* as specific asset(s) provided by an FI, and thus each rate is appropriate for inferring the services bundled with assets of that particular risk class. This is consistent with Barnett's (1995) criterion: Each reference rate is based on an asset held solely for the (usually risky) return it provides, free of any other financial services. The only difference is that the basic user-cost principle is extended to a world with *many* "pure" rates of interest.

A thought experiment highlights the difference between our extension of the user-cost approach and the existing version. Suppose all transactions were costless, so that no financial services were needed. According to the existing method, in such an ideal world there would be no differentials between loan interest rates, the reference rate, and deposit rates. In contrast, our method would still imply different rates for financial products with different risk characteristics. Only the latter approach is consistent with modern asset-pricing theories.

In the case of one important set of financial institutions, namely, banks with deposit insurance, the new theory has an interesting implication: One should choose a different reference rate for depositor services (the risk-free rate, assuming the deposits are insured) than for borrower services (risk-adjusted rates). This means that not all of a bank's net interest income is compensation for its implicit services. This result differs from the existing

5

implementation of FISIM, which takes the *entire* net interest as the nominal value of implicit financial services. So, we conclude that the usual implementation needs to be modified.

Now, the question is, how should one interpret the gap between the appropriate reference rates for loans and deposits, that is, the risk premium on loans. We explain in Sections 3 and 4 that, as elaborated in Wang (2003a), this gap is part of a borrowing firm's cost of (risky) capital and not the lending bank's output. In particular, we argue against treating this risk premium as income for what some term the "risk-bearing service." By our definition, risk-bearing is not a real service, because it requires no real resources to produce. But beyond this definitional point, we show that classifying risk-bearing as a bank service creates various distortions, especially in measuring the output of firms that borrow from banks.

In summary, one should always use a reference rate adjusted for the relevant risk; one uses the riskless interest rate only in special cases when no risk adjustment is needed. But is it hard to implement this clear theoretical recommendation in practice, and how would one go about this? The short answer is, no, it is not difficult. Once one understands that the ideal reference rates would be multiple, and adjusts for risk, this recognition will help to focus future research efforts and data collection in the right directions. Moreover, in practice, the work of risk adjustment is made much less complex by the fact that one can apply a required rate of return to a portfolio directly (for example, to a bank as a whole, or even to the entire banking sector) and not to each asset individually. It is still a complex project, but one that is feasible and well worth undertaking. Wang (2003b) discusses a variety of practical issues and measures nominal bank output using risk-adjusted reference rates in a pilot project using data from bank holding companies.

So far, we have focused on the measurement of FIs' nominal output, and our contribution to the literature is to clarify the central role of risk and the need to account for risk when measuring financial output. Note that, as stated at the outset, to derive our nominal measure, one needs only to specify the models for pricing risky assets, but needs not spell out the *exact* services FIs supply.

In contrast to the relatively straightforward problem of measuring nominal output of FIs, when it comes to decomposing nominal output into a real quantity index and a price index,

one does need to know exactly what financial firms produce that their customers value. These are questions that the current literature does not formally consider; thus, our model-based answers are quite different from those that have been suggested recently.

In previous work (Wang 2003a; WBF 2004), we demonstrated the separability between the flow of financial services and the stock of financial products. In fact, casual observation increasingly confirms that flows of FIs' services are not inherently tied to stocks of financial instruments. For example, in securitized mortgage lending, some banks specialize in providing bookkeeping and payment services to both the borrowers and the holders of mortgage-backed securities, yet they do not hold any of the underlying mortgage loans on their balance sheets. Thus, two banks may perform the same services of screening and monitoring loans (that is, produce the same output), but hold very different stocks of financial assets and liabilities.

If financial services are not always supplied in fixed proportion to the stock of assets that an FI holds, then the interest-rate differential is not necessarily proportional to the price of the service output. Thus, the service price needs to be measured in a different way, for example, as we have argued (Wang 2003a; WBF 2004), by a price index similar to those used for other professional services. Alternatively, one can obtain an implied price index (given the measure of nominal output discussed above) by creating a real quantity index of the various activities performed by FIs. In fact, the Bureau of Labor Statistics (BLS) traditionally constructs such as an aggregate quantity index for real bank output (consisting of services to depositors and borrowers, such as ATM transactions and loan screening; see the BLS *Technical Note* 1998). In short, anyone who argues for general fixed proportionality needs substantial evidence to justify this strong assumption in the face of the numerous real-world counterexamples we and others have identified.

In conclusion, correct measures of bank output, at least, are well within our grasp. One needs only to implement the risk-adjusted measure of nominal bank output, divide by a real quantity index along the lines that the BLS now follows, and obtain a nominal price index. These quantity and price indices may be quite different from the ones currently proposed or implemented, as evidenced by the empirical findings in Wang (2003b). But the measures we propose have the virtue of being consistent with economic and finance theory, and their

underlying principles can readily be applied to measure the output and prices of other financial institutions.

2. The Economic Model of Financial Institutions

2.1 The Model and the Definition of Financial Services

One must first define a concept before attempting to measure it accurately. So, what is the output of financial institutions? Wang (2003a) and Wang, Basu, and Fernald (WBF 2004) answer this question with dynamic models of FIs' optimal operations under uncertainty.[3] By embedding FIs' operations within the backdrop of competitive financial markets, these papers show that the value added of FIs lies solely in resolving information problems and processing transactions. In accordance, Wang (2003a) and WBF (2004) formally recognize the intuitive parallel between the activities of FIs—processing a wide range of information and transactions—and the activities of other professional services firms, such as providing consulting and accounting services. For instance, in the case of banks, one of their important services to borrowers is the processing of credit information through screening and monitoring.[4] All of these services are generated through a production process that uses primary inputs of labor and capital, as well as intermediate inputs.

More importantly, Wang (2003a) demonstrates the separability between the flow of financial services and the stock of financial products. By nature, financial services are typically intangible; financial products such as loans and derivatives are often the most tangible manifestation of these services. However, Wang (2003a) and WBF (2004) show that flows of FIs' services are not inherently tied to stocks of their securities holdings. So, there is *no* theoretical

---

[3] WBF (2004) extend Wang's partial equilibrium model to a general equilibrium setting and demonstrate that all the qualitative results in Wang (2003a), summarized in this paper, continue to hold.

[4] Banks' role in resolving information asymmetry is well recognized. For example, Fixler (2004, p. 223) observes that "these [financial] services have to do with overcoming the problems of asymmetric information, a view that is widely accepted in the financial economics literature.…The role of banks as a lender arises out of the private nature of the lending transaction—the bank…reduces the problem of adverse selection…[and] the problem of moral hazard. These are the financial services that Fama (1985) and others identified when they argued that banks were special."

basis for fixed proportionality between the flow of services and the stock—be it measured by book or market value—of financial securities.

The separability between the flow of services and the stock of securities has become quite evident, as more and more firms are providing services without holding the associated securities or *vice versa*. For instance, in securitized mortgage lending, now the dominant means of supplying credit to home buyers, there are banks that specialize in providing bookkeeping and payment services to both the borrowers and the holders of mortgage-backed securities without holding any of the underlying mortgage loans on their balance sheets. On the other hand, a growing number of small and medium-size banks hire specialized firms to originate and service their mortgage loans (Bergquist 2002).

A significant implication of the separability between financial services and the stock of financial products is that the importance of an individual bank or even the banking sector as a whole is not necessarily related, let alone proportional, to the size of its asset holdings.

Theories, in general, offer no reason to assume that the quantity of services produced is in a fixed ratio of the balance of financial products, and the onus should be on those who assume so to justify that assumption. Fixed proportionality is hard to defend even for the case of traditional services, especially for transaction services such as banks' payment services to depositors. The trade literature in banking suggests that banks see check clearing as a basically homogeneous service, regardless of the dollar figure of the check.[5] The same is true of many other depositor services (for example, issuing money orders, transferring funds, etc.). By comparison, the fundamental separation between service flows and asset balances may seem less evident in cases where the financial firm both performs the services and holds the assets, as in traditional bank lending. However, once we identify and focus on the true underlying *services*, the detachment of service flows from stocks becomes equally intuitive. For instance, originating a $1 million residential mortgage probably entails a cost much lower than 10 times the origination service needed for a $100,000 residential mortgage. (The two originations likely

---

[5] See, for example, Kimball and Gregor (1995) and Radecki (1999); both present industry data on the average and marginal cost of processing a payment via various means that make no reference to the size of the payment.

involve about the same amount of work nowadays, given the ready availability of credit scores for individuals.)

2.2 The Corresponding Definition of Pure Returns—The Role of Risk

One crucial implication of the service-centric definition of FIs' output in Wang (2003a) and WBF (2004) is that the flow of services is qualitatively distinct from the per-period returns accrued to the stock of financial assets. The returns are solely to compensate suppliers of funds for forgoing current consumption in exchange for future consumption. Free of any elements of services, such returns can be called "pure returns." In cases where pure returns are joint with implicit service charges, the former must be netted out of an FI's total income to impute the latter.

It must be noted that the concept of "pure returns" is related yet distinct from that of the usual opportunity cost. In a world with risk, it is wrong to use the rate of return on a security with no credit risk as the opportunity cost of funds for every risky security simply because one could invest in the riskless security instead. By that argument, it would be equally valid to use a junk bond rate as the opportunity cost, since one also forgoes the opportunity to invest in high-risk—thus high-yielding—junk bonds! Furthermore, even within Treasury securities, a 10-year bond yields a higher average rate of return than a 90-day bill if held over the same 90-day period, because the former is subject to the risk of time-varying (long-term) interest rates. Since risk matters for asset prices, the opportunity-cost argument alone—without reference to risk— provides little theoretical guidance in setting the "pure return" (that is, opportunity cost) of funds, let alone justifying an arbitrary definition of the opportunity cost in terms of a risk-free rate. (See Wang 2003a and WBF 2004 for more detailed exposition of this point.)

Then, what determines the rate of "pure return" investors expect on a financial security? Applying standard theories of asset pricing, Wang (2003a) and WBF (2004) show that the pure return depends positively on the correlation between a security's risk and systematic factors (such as the business cycle). In fact, if a security's risk does not vary with systematic factors, then its expected rate of return is merely the risk-free rate, no matter how volatile its return; for example, if a loan's default risk does not vary with macroeconomic conditions, then any lender

should expect a return equal, on average, to the risk-free rate, even though the interest rate the borrower must promise to pay *conditional on his solvency* rises with his probability of default. (See Wang 2003a for detailed explanation of the distinction between the contractual interest rate and the expected rate of return, or equivalently, the distinction between *default* and *risk* premium.) In practice, for publicly traded securities, one can simply take the expected rate of return as determined in the relevant markets. For private securities, the rates of return on comparable public securities—those with the same or the closest risk characteristics—serve as good proxies.

Think again about the example of mortgage origination; suppose, conditional on the borrowers' respective income and credit scores, that default is more likely for a $1 million loan than for a $100,000 loan, and that the default risk covaries with business cycles. Then, investors will demand a higher rate of return on the former even in expectation.[6] Either expected rate can be well approximated by a suitably weighted average of the rates of return on mortgage-backed-securities issued by private FIs. Since realized returns, on average, equal the expected return, the $1 million loans should, on average, generate a higher flow of interest income. This would be true even if, as suggested above, the two kinds of loans entailed the same origination services.

Yet another reason why risk is intrinsic to the "pure return" is that no security is entirely risk free, unless risk is defined very narrowly. In the standard parlance of finance, "risk" encompasses any factors that render a security's future payoff uncertain: inflation risk, credit (that is, default) risk, liquidity risk, time-varying risk premium, etc. Thus, even most Treasurys, though free of default risk, are not free from other risks: Most pay nominal coupon rates and so different maturities are subject to different inflation risk and time-varying risk premia. Furthermore, all categories of risk are equally relevant qualitatively; that is, they are all taken into account in the market-determined rate of return, although their quantitative effects may differ. In particular, credit risk is no more important than any other type of risk. So, there is no

---

[6] Combined with the former's higher default probability, this will lead to an even higher interest rate promised in the contract for the $1 million loan.

theoretical basis for excluding from a security's pure return any portion that corresponds to one or more specific types of risk.

2.3 Comparison with the Existing Implementation of the User-Cost Approach

Underlying the current NIPA measure of financial output is the notion that implicit financial services can be thought of as *a flow of services from a stock*. For instance, a flow of implicit services to borrowers (depositors) is derived from a stock of loans (deposits). As mentioned earlier, this is the theoretical view taken in the models pioneered by Donovan (1978) and Barnett (1980), and later adapted in Hancock (1985), in which a user cost of money is conceptualized as analogous to the user cost of capital in Diewert (1974). The user cost of money is defined to be the interest rate on an asset held solely for financial return and not for any other financial services (Barnett 1995). The NIPA view U.S. Treasurys to be such assets, and so take an average Treasury rate to be the single reference rate to divide net interest into implicit borrower services versus depositor services.

This notion of financial services and the definition of user cost of money exemplify the two primary conceptual problems with the existing application of the user cost approach to the measurement of financial output. They also epitomize the essential difference between our new model and the existing approach. First of all, at its core, the latter still does not recognize the fundamental distinction between the flow of true services (according to the essence of the definition in national income accounts) and the flow of risk-related returns *per se*—those accruing purely to the associated financial securities. This confusion is manifested in the use of interest *rate* differentials as the implicit price of services; this practice amounts to assuming fixed proportionality between the flow of services and the stock of financial securities.

Secondly, this confusion corresponds to a notion of "pure returns" based on an opportunity-cost argument that omits risk and thus is unable to pin down the reference rate (the so-called user cost of money). The telling sign is that, with the opportunity cost tenet alone, the NIPA are unable to choose a definite reference rate within the family of Treasurys: There are roughly two classes and many candidates—either the rate on a single Treasury security of a given maturity or a weighted average of rates for securities with different maturities (see, for

12

example, Fixler 2005). This indeterminacy arises precisely because the theory underlying the NIPA measure disagrees with modern finance theory's consensus that every security's "user cost" depends on all the risk factors—often more than one—that affect its payoff. In addition, as stated above, few securities, and under very limited conditions, are truly risk free.[7]

In contrast, as shown earlier, Wang (2003a) and WBF (2004) resolve both confusions by accounting formally for the risk in the financial instruments FIs offer. Their stochastic models of FIs can be thought of as a risk-augmented user cost approach; that is, the user cost of funds subject to uncertainty, which is almost always present in reality, is shown to depend on the risk. The new method thus implies that there are almost surely multiple reference rates, because risk differs both across FIs and across categories of financial instruments.[8] Of course, the multitude of reference rates implies an aggregate rate, but the latter will be such a complex weighted average that it is hardly likely to correspond to the rate of any single security.

The key conceptual difference between the new stochastic and the existing, non-stochastic user-cost approach is best illustrated in the limit case where transactions are all costless. In that ideal situation, according to the existing approach, there would be no differentials between loan interest rates, the reference rate, and deposit interest rates. The logic is simple: Without uncertainty, all interest rate differentials are entirely attributable to implicit charges for transaction services; so, if transactions were costless, there would be no more rate differentials. In contrast, even in an ideal world with costless transactions, the stochastic approach would still feature different rates for financial products with different risk characteristics. Clearly, only the latter result is consistent with asset pricing theories in modern finance.[9]

---

[7] One rare example of returns with certainty is inflation-indexed Treasury bonds (TIPS) held to maturity. If sold prior to maturity, however, even TIPS become risky, as they become subject to a time-varying risk premium.

[8] In fact, not only do different categories of loans and securities within each bank likely require disparate reference rates because of different risk characteristics, but the bank-wide reference rate will also differ across banks to the extent that banks have financial portfolios with different overall risk profiles.

[9] In fact, almost all standard asset-pricing theories, such as the CAPM and Black-Scholes option-pricing formula, assume no transaction costs and study exclusively how returns vary according to risk.

It must also be noted that the new stochastic model's risk-based criterion for pure returns dovetails with its definition of financial services to form a coherent system that fully accords with the essence of the user-cost concept. Given the model's definition of what constitutes true financial services, virtually all market securities are not associated with any services and are held solely for the returns they provide. That is, risk-related market returns fit precisely the concept of pure returns. Conversely, once the flow of returns due purely to securities' risk-related future payoff schedule are netted out of an FI's total revenue, the residual income will contain only compensation for its implicit services. This corresponds exactly to the model's definition of FIs' (nominal) output.

3. The Model-Implied Measure of Financial Institutions' Output

3.1 Nominal Output

We use banks as an example to illustrate briefly how to implement the output measure proposed in Wang (2003a) and WBF (2004). We term the financial output measured according to the new models "pure services," to distinguish it from the current measure in the NIPA.

First, consider the nominal value of implicit services that are compensated through "interest" income (for example, bank lending services), based on a balance of assets (for example, loans). Let A denote that balance, r^A denote the interest rate charged net of the expected default rate,[10] and r^M the rate of return investors expect on a market security with the same (or most comparable) risk profile (for example, the market rate on comparable commercial paper or asset-backed securities). The nominal service output—denoted Y^A—is then imputed as:

$$\mathcal{Y}^A = (r^A - r^M)A. \tag{1}$$

So r^M is the risk-adjusted reference rate for imputing implicit lending services. Note that (1) applies to both on- and off-balance-sheet securities. That is, A can represent not only traditional loans but also the relevant balance of loan commitments or stand-by letters of credit. (1) also implies that, if a bank just holds some market securities passively, then $r^A = r^M$ and its Y^A is zero.

---

[10] It equals the rate of return the bank receives on average. See Wang (2003a) for the derivation.

Likewise, consider the nominal value of implicit services compensated by paying a lower interest rate (for example, deposit services) on a balance of liabilities (on, for example, deposits, or off the balance sheet). Let D denote that balance, r^D the actual interest rate paid, and $r^{M'}$ the rate of return on a market security with the same risk. This makes $r^{M'}$ the reference rate for implicit services to holders of bank debt (for example, depositors). Then, the nominal output—denoted Y^D—is imputed as:

$$\mathcal{Y}^D = (r^{M'} - r^D)D. \qquad (2)$$

Most of banks' services to holders of their liabilities are to depositors. For insured deposits in the U.S. (that is, up to \$100,000 per individual), $r^{M'} = r^F$ (the short-term Treasury rate). For the remaining, uninsured deposits, $r^{M'} > r^F$, because the holders are exposed to some of the risk in banks' asset portfolios. When the liabilities are commercial paper or bonds banks issue in the markets, Y^D will be zero, since the interest rate paid is exactly the market reference rate; that is, $r^{M'} = r^D$. This seems a rather intuitive result, since no services—according to the new, pure-service definition of output—are produced.

Note that, under virtually all circumstances (for example, whenever there are equityholders), r^M in (1) is greater than $r^{M'}$ in (2), because bank assets are typically riskier than their liabilities. In other words, the reference rates for imputing lending and depositor services almost always differ.

Figure 1 illustrates the imputation of nominal implicit bank service output, depicting the polar case in which all bank service revenue is implicit (that is, bundled with interest), and bank liabilities are composed entirely of deposits.

Above all, note that only *part* (areas I and IV) of a bank's net interest income (gross interest income minus interest payment, equal to the sum of areas I through IV) constitutes the true value added of bank services.[11] This mirrors the above result that the reference rates for lending and depositor services differ. Area I, the part of interest income exceeding the (risk-

---

[11] As shown, the balance of loans exceeds that of deposits, with bank equity making up the difference. Also, area V'—the deposit insurance premium—is part of the bank's overall explicit payment for deposit funds.

adjusted) return expected on market securities with risks comparable to the given set of loans, corresponds to Y^A above. Area IV, the interest depositors forego, corresponds to Y^D above.

In contrast, the other two areas—II and III—are both merely pure interest transfers. Area II, equal to $r^M A - r^{M'} D$, is the return bank shareholders demand for bearing the (systematic) risk of the loan portfolio. This part of bank shareholders' income, however, is, in fact, factor income originally generated by the capital used in the borrowing firm's production. We explain in the next section why, in a consistent accounting system, this income should be counted as part of the borrowing firm's value added instead of as part of the lending bank's. Area III, equal to $[r^{M'} - (r^F + \alpha)]D$ (where α is the insurance premium charged by FDIC), is the value of (the put option embedded in) deposit insurance accruing to bank shareholders.[12] So, area II is the factor income transferred from end users of funds to ultimate suppliers of funds (the bank shareholders), while area III is a transfer from taxpayers to bank shareholders.[13] Only when all investors are risk neutral will these two elements of interest transfer disappear.

Note that while Figure 1 depicts a situation in which the expected rate of return on loans is greater than that on deposits (or liabilities in general), the opposite can and does happen, usually at the end of a boom, when long-term (loan) rates fall below short-term (deposit) rates. In this case, the logic for imputing bank value added remains the same, but banks now suffer a loss on the funds because of the inverted term structure—a loss that is analogous to a capital loss on inventories. In actual calculations, the imputed value of bank output implicitly priced is added to explicit fee income to yield total bank output.

In practice, the reference rate on a financial product (or a portfolio) should simply be the market interest rate on market securities with the same or the most similar risk features. Wang (2003b) applies the pure-service output measure to bank holding companies (BHCs), in particular detailing which market interest rates to use as reference rates for imputing the value of depositor services and multiple categories of lending services.

---

[12] Area III exists for an individual bank if the deposit insurance premium it pays to the FDIC is lower than what would be required by a private insurer, given the risk of the bank's asset portfolio.

[13] For the banking sector as a whole, area III will disappear if the total deposit insurance premium is actuarially fair. But there may still be transfers within the banking sector—from low-risk to high-risk banks.

3.2 Real Output and Its Price

As the new models make clear, FIs' services are qualitatively analogous to consulting or accounting services. Hence, their real value should be derived from the nominal value in the same way: deflated using quality-adjusted price indices. To continue the above example, real bank output is

$$Y^i = \mathcal{Y}^i / P^i \text{ , } i = A, D. \tag{3}$$

$\mathcal{Y}^i$ represents the nominal service output mentioned above, and P^i is the corresponding deflator. Conversely, FIs' real output can be constructed as an index of various transactions processed, such as (in the case of banks) the number of specific categories of loans originated or deposit accounts serviced. The real output index then implies the price deflator. In fact, this is basically how the BLS currently constructs the index of real bank output (Technical Note, 1998).[14] WBF (2004) delineate how the BLS's method should be modified to accord with the pure-service output measure. The BLS aggregates the services associated with lending and deposit-taking, using as weights the share of employment devoted to each activity. Economic theory suggests, however, that one should use the respective nominal output share of each activity as the weight instead. Of course, one needs to use the risk-adjusted interest-rate spreads to derive the nominal output measures. Thus, the aggregate measure of real financial- sector output also changes with the adjustment for risk, but indirectly, through the shares needed for aggregation. [15]

More generally, the new models' most crucial implication for the measurement of FIs' real output is that, in principle, the quantity of services is not necessarily proportional to the

---

[14] The BLS aggregates several categories of loans weighted by their respective interest rates to get lending output; it then sums up lending and depositor services weighted by their respective employment shares to obtain the index of total bank output.

[15] To be precise, the BLS actually follows a two-step procedure: It aggregates several categories of loans weighted by their respective interest rates to get lending output, and then sums up lending and depositor services weighted by their respective employment shares to obtain the index of total bank output. In our proposed scheme, everything would be symmetric: the different categories of loans would also be weighted by their nominal output shares. These revenue shares, however, are unlikely to correspond to relative interest rates of the loans, because, as we have discussed extensively, each loan interest rate must be adjusted for the risk associated with the interest flow.

balance of associated financial instruments, and so interest rates or rate differentials in theory are not the price of services. In the case of banks, this means that Y^A is not a fixed proportion of the balance of financial assets (A in (1)), nor is Y^D of the balance of deposits (D in (2)); P^i bears no fixed relation to any gross interest rate (for example, $1+r^M$ or $1+r^F$). This implication is the corollary of the pure service definition of financial output: The flows of returns—including the risk premia—accrued to the funds invested in non-bank productive capital are not part of FIs' output.

We recognize the practical difficulty in collecting all the data needed to construct broad indices of financial services that adjust adequately for quality differences. In particular, the number of many types of transactions (for example, new loans originated) is often not observed, although the balance of the associated financial products is (for example, the book value of new loans). Nonetheless, we believe it is imperative to clarify first the theoretically-ideal output measure. The theory can then serve as the guide for improving the measure in practice. Wang (2003b) discusses in detail what data (for example, numbers and marginal operating costs of various types of transactions) one needs to collect and what weights one should use to construct the index of an individual FI's real output of pure services.

3.3 Comparison with the Existing Measures of FIs' Output

At the firm level, it is the norm to define output as goods or services that generate revenue. For many FIs, especially traditional banks, most of their income is "interest" from financial assets. Applying the usual micro definition of output to FIs, however, contradicts the convention in national income accounting: that interest is a transfer and not a true factor income. So, for years, banks' output was essentially counted as negative in national accounts.

Over time, we have recognized that FIs' net interest income is implicit compensation for their services, as reflected in the SNA93 guidelines. In accordance with SNA93, the U.S. NIPA equate the *entire* interest margin to the nominal value of implicit financial services. Furthermore, since the 2003 comprehensive revision, the NIPA also apply a single reference rate to divide the entire spread into two parts: The part of interest income that is above the risk-free return (that is, $(r^A - r^F)A$ in Figure 1) is counted as the nominal value of implicit borrower

services, while the return forgone by depositors (that is, $(r^F - r^D)D$ in Figure 1) is counted as depositor services.

The resulting differences thus lie primarily in the value of (implicit) services to borrowers. The NIPA include the risk-related returns on the funds lent by bank shareholders (that is, areas II and III in Figure 1), whereas the new measure does not. For the same reason, the two measures differ on nominal output of services to holders of uninsured deposits; it is less in the NIPA than in the new measure by $(r^{M'} - r^F)D$ (that is, by the risk-related returns depositors require; see also Figure 1). Because of deposit insurance in the U.S., both measures agree on nominal output of services to holders of insured deposits.

The distinction between the two measures is best illustrated with a special case in which a bank funds its loans entirely with uninsured deposits (for example, via jumbo CDs). Then, since investors in the bank's bond will demand $r^{M'}$, which is greater than r^F, the NIPA will impute a value of borrower services that exceeds the whole of net interest, and a value of depositor services that is negative! This example highlights the conceptual problem in the NIPA's choice of r^F as the reference rate. We elaborate in the next section why counting risk-bearing as an output is not a good argument for using r^F.

The more crucial difference between the two measures lies in the way nominal output is decomposed into quantity and price. The two measures imply different concepts of real output and, in turn, different price deflators. The NIPA now explicitly identify the two interest *rate differentials* as respective implicit *prices* for the implicit financial services. That is, the discounted rate differential $(r^A - r^F)/(1 + r^F)$ is viewed as the supply price of implicit services to borrowers, while the rate differential $(r^F - r^D)/(1 + r^F)$ is the price of implicit depositor services (Fixler et al., 2003; Fixler, 2005). These definitions of output prices imply that balances—in *current* dollars—of financial assets (largely loans) and deposits are the quantities of real output of borrower and depositor services, respectively.

In contrast, the pure service measure of real output bears no definitive relation to balances of financial assets or deposits, so the implied price deflators are not necessarily proportional to interest-rate differentials. Instead, real output is an index of the relevant bank services; for example, real services to borrowers equal an index of, say, numbers of different

categories of loans screened and reviewed, with each type of activity weighted by its marginal operating cost. In turn, the implied price deflators are thus analogous to common deflators such as the CPI and the GDP deflators.

Our proposed measure of real financial output is similar conceptually to what Humphrey (2004, p. 215) calls a direct measure of output quantities, although we propose to deflate imputed nominal output by a constructed price deflator, avoiding the need to add output components across all financial firms. Humphrey notes that "academics would prefer such a measure of banking output to what they now use" if it were available for individual banks. If an industry price deflator were constructed along the lines we propose, banking could follow the usual practice of constructing real output by deflating nominal firm revenues by the industry deflator.

The same argument distinguishes the new output measure from that used in virtually all studies of individual banks and bank holding companies (BHCs). The existing micro measure of bank output basically uses the deflated book value (BV) of financial instruments (assets and sometimes deposits) as the quantity, and interest rates or rate differentials as the price.[16]

Wang (2003c) implements the new measure using BHC data and compares the results with those from the two existing measures. The new measure of output differs noticeably from both existing measures, and several findings stand out. First of all, the new measure of output is 25 percent smaller than the SNA93 output measure, which is based on the entire gap between interest receipts and payments—that is, not accounting for the risk of bank asset portfolios. However, the growth rate of the new series is higher than those of both the SNA93 and the BV-based measures. Second, the new measure of output growth declines the most during the 1990–

---

[16] More specifically, there are three variants of the micro measure, each set apart by the treatment of deposits as an output or an input. The asset approach views deposits as an input for making loans, which, together with market securities, constitute the output whose quantity is measured using the deflated book value. The value-added approach views every financial product whose creation requires labor and capital as inputs, and it thus records deposits as an output. It measures output using the book value of financial assets and deposits. The user-cost approach, which is regarded as the foundation for the NIPA measure, classifies input and output endogenously: It specifies a reference rate and treats financial assets (liabilities) whose realized rates of return are greater (less) than the reference rate as output, and others as input. So transaction deposits are typically found to be outputs in data. For a recent survey, see Berger and Humphrey (1997).

91 recession, and its time series is more volatile than both existing output measures—most likely because of measurement errors in the new series, as data limitations necessitate the use of many proxies. Last, the nominal value of the new output on average equals total non-interest cost, implying a zero-profit condition, even though the new measure is constructed completely independently of any cost information.

Clearly, even in the simple setting of the example, neither the BV of loanable funds nor the net interest income is likely to equal bank value added. Moreover, despite the example's simplicity, its logic for measuring bank value added through a proper division of total income remains valid, even when banks cannot precisely assess the type of individual borrowers and thus have to use interest rates as part of the sorting mechanism, as in Stiglitz and Weiss (1981).

3.4 Why Bearing Risk Is Not Counted as A Financial Service

Some may argue that the financial output measure in the NIPA is valid if one regards bearing risk by FIs as a form of services. Bank shareholders' rewards for bearing the risk on their funds lent to productive firms (that is, area II in Figure 1) should thus be counted as part of banks' value added. We explain in this section why it is inconsistent to count FIs' risk bearing as a productive service.

The foremost argument against risk bearing as a service of FIs is that it will lead to inconsistent accounting of the output of the borrowing firms. Consider two completely identical firms—they produce the same widgets, at the same prices, and face the same risk of default (for example, from the fluctuating price for their output). One firm borrows from a bank and the other from the bond market. Assuming costless transactions, both firms should face the same debt interest rate of r^M (in Figure 1's notation) as dictated by the no-arbitrage condition.[17] Under an accounting system (for example, the current NIPA) that counts risk bearing as a service output of FIs, the firm that borrows from the bond market will be credited with producing nominal value added of $wL + r^K K$ (L: labor, K: capital, w: wage rate, $r^K = r^M$: rental rate of

---

[17] With costless transactions, $r^A = r^M$ in Figure 1. This assumption is solely to simplify the explanation without any loss of generality. A single rate on fixed-income securities that fund projects of a specific risk profile will arise so long as there is an integrated and competitive bond market.

capital, assuming zero depreciation and constant price for K). But the firm that borrows from a bank will be credited with a nominal output of only $wL + r^F K$, while the residual $(r^K - r^F)K$ is credited to the bank for its "service" of bearing the risk.[18] What this nominal adjustment implies about the measure of real output is not at all clear. It is, nonetheless, clear that such an accounting system will likely credit identical firms with producing different amounts of output, depending on the source of their borrowed funds.

The mirror image of the above inconsistency is that counting risk bearing as a service of FIs will distort the relative importance of the financial sector and even the size of GDP, both across countries and over time. Consider two countries (A and B) that are identical in every other respect except that most households in country A borrow from banks whereas most in B borrow from the bond market through securitization. Then, the loan-risk-as-service accounting system will show that, in each period, A's financial sector produces more output than B's, and, as a result, that A's GDP is greater than B's. Now, assume that a few FIs in country B take over some FIs in country A and gradually introduce securitization to A. During that process, both the output of A's financial sector and A's GDP will actually shrink as a smaller amount of consumers' interest payments goes to banks, even though every real economic activity in A remains the same.[19]

More importantly, it is both theoretically unintuitive and practically difficult to devise an accounting system that counts risk bearing as a service output yet avoids the above inconsistencies.[20] To make a non-financial firm's output invariant to its mode of funding, the

---

[18] This result, for simplicity of exposition, implicitly assumes that the bank is fully funded with equity. Otherwise, the residual amount credited to the bank would be $r^M A - (r^F + \alpha)D$, in Figure 1's notation.

[19] A more subtle consequence of such an accounting system is that FIs' output, and likely GDP in turn, will fluctuate as households' valuation of risk varies over time (say, due to changing risk aversion) or across segments of the financial markets. Imagine that households become more risk averse tomorrow relative to today. Then, the risk premium will rise, and so will bank output, in a system that counts risk bearing by FIs as output, even if banks perform exactly the same tasks. This means that measured bank output will fluctuate over a business cycle even if bank activities stay constant, since it has been established that risk premia on bonds covary with the business cycle (see Fama, 1984; Fama and French, 1989 and 1993).

[20] Such a system has not, to the best of our knowledge, been worked out, and the current NIPA—which count only FIs' risk-related return as a service—do not meet the consistency requirement. We will discuss the most salient features necessary for such a system, but it is beyond the scope of this paper to develop a full-fledged version and consider all the relevant consequences.

risk-related extra return on its capital (that is, $(r^K - r^F)K$) will have to be netted out of its value added, whether the premium is paid explicitly to banks or other types of FIs, or implicitly to bondholders who are households.[21] The economic intuition for such an adjustment is unclear. In addition, some form of structural asymmetry will remain inevitable in such an approach. If the implicit interest the firm pays is classified as purchases of risk-bearing services from its bondholders, then clearly the risk premium FIs' pay to their bondholders is treated asymmetrically in such a system: It is not considered as FI *bondholders'* output but as FIs' *own* output of risk-bearing services. If the implicit interest the firm pays is classified as its own output of risk-bearing service—distinct from its regular output (for example, cars for auto companies)—then total output of identical firms can again differ, even though their regular output is now made symmetric.

Such a risk-as-service system also leads to peculiar outcomes when applied to measuring the output of FIs other than banks narrowly defined. Consider, for example, mutual funds. Legally, the account holders are owners of a fund's assets—shareholders. So a risk-as-service system will count the risk premium on fund assets (actual return in excess of the risk-free return) as fund owners' income for their risk-bearing service, and so an equal amount must be counted toward the fund industry's service output. However, from an economic perspective, what does it mean to credit stock mutual funds with producing trillions of dollars of value added corresponding to the difference between average stock returns and U.S. Treasury returns? In particular, compare a stock index fund with a bond index fund. It is well known that managing an index fund account entails the same technology and virtually the same amount of work, regardless of the underlying securities. But the risk-as-service system will credit the stock fund with a much greater output and, in turn, greater productivity most of the time, since stocks typically command a higher risk premium. Of course, in periods of stock market slump, those funds would presumably produce large *negative* quantities of value added, and hence

---

[21] More precisely, what is counted as payment to FIs (such as banks) for their risk-bearing service contains only the risk premium to FIs' *shareholders* but not the risk premium to FIs' bondholders. That is, only area II in Figure 1 is counted, not the entire risk premium on FIs' lending, equal to the sum of areas II, III and V'. So, the amount that should be netted out of bondholders' interest receipts is generally less than the entire risk premium on the bonds.

would appear to have negative levels of productivity. On the flip side, firms with stocks outstanding would presumably look much better in those bad times because of their negative purchases of intermediate financial services!

Furthermore, a system that treats bearing risk as a productive service is inconsistent with the essence of the long-standing national income accounting principle that the provision of *finance* is not a productive activity, once the concept of finance is updated to accord with modern asset pricing theories. National accounting was initially designed in a non-stochastic framework, so "finance" (that is, funds *per se*) was defined as providing funds under certainty, so that all securities would carry the same rate of return, the risk-free rate. As far as the funds are concerned, however, modern theories make no qualitative distinction between Treasurys and corporate bonds—all intertemporal financial contracts are considered risky, and credit risk is but one among the many kinds of risk relevant for fixed-income securities such as loans (time-varying term structure being the most prominent). If one wants to treat bearing risk as a service, one must specify what principles determine the relevance of various kinds of risk; it seems arbitrary to consider only credit risk. It seems equally arbitrary to count only the risk-related part of the return on funds but not the rest, given that both parts are returns to compensate investors for deferring consumption. So, we believe that the correct modern definition of "finance" in income accounting should be the funds behind any intertemporal contracts, which are almost always risky.

Finally, treating risk bearing as a service output also causes serious conceptual difficulties for productivity studies. Consider two banks with identical technology for making loans of all types. One chooses to make relatively safe loans, while the other makes risky loans at higher interest rates (say, because its shareholders are less risk-averse). Suppose both banks provide the same lending services, such as screening and monitoring. A risk-as-service measure will yield a greater output and thus greater productivity for the bank that makes riskier loans. Such a measure is likely to equate prudence with lower productivity or, more generally, to be influenced by FI owners' risk attitude *per se*. This seems inconsistent with established principles in production theories.

In contrast, none of the above inconsistencies or conceptual difficulties arise under the risk-augmented user cost approach in Wang (2003a) and WBF (2004). Since bearing risk is not considered a form of service in the new framework, a non-financial firm's output is invariant to its source of funding (be it banks or bond markets), since its cost of capital always includes the risk premium. Symmetrically, FIs are credited only with the services they actually produce using labor and physical capital. For instance, mutual funds' output includes only transactions, book-keeping services, and, in certain circumstances, financial advice. So an index fund, be it for stocks or bonds, will have the same output, independent of the period return on the underlying assets.

Moreover, under the new output measure, cross-country comparisons of financial-sector output and GDP are, in principle, not unduly affected by different structures of financial systems. Likewise, the growth rates of financial output and GDP are not distorted by structural changes in the financial system or varying risk attitude over time. Relative to any measure that treats risk bearing as a service (as do the current NIPA), the new output measure is also much closer in spirit to the national-accounts principle that the provision of risk-adjusted funds *per se* is not a productive activity. Last, the new measure will yield the same value of output and productivity for FIs that carry out the same tasks, regardless of their attitude toward risk.

4. Our Theory's Implications for the Evolution of Financial Systems

Our new theory's conception of bank value added helps to settle the long-standing debate about the role of deposits—are they an input or an output? The answer is that they are both: They provide the bank with loanable *funds* for financing loans and are thus an input into the lending process; on the other hand, the set of *services* provided to the suppliers of deposit funds should be viewed as bank outputs.

The argument against assuming fixed proportionality in general between financial service output and the size of asset portfolios also applies, more intuitively in fact, to non-bank FIs. In the case of rating agencies, we would not use the dollar value of securities rated to measure their output; instead their output should be more akin to the number of various types of issues rated. The same holds for mutual funds. We understand that their value added is the

services they provide, and the quantity of these services is more closely related to the number of accounts and the complexity of the investment strategy than to the size of asset portfolios. Non-fixed proportionality explains why mutual funds offer discounts to owners of accounts with large balances; likewise, expense ratios of managed funds are far higher than those of index funds because the former provide more services, such as investment advice.[22]

Similarly, the distinction between the flow of financial services and the per-period return on assets also applies to non-bank FIs. That is, service output bears no definite, let alone fixed, relationship to the risk-related interest rate on an FI's portfolio. Again, in the case of mutual funds, a fund with a riskier portfolio generates greater output only to the extent that the extra risk requires more information processing. But we would not say that the fund creates higher value just because its assets carry higher e*x ante* (and, on average, *ex post*) returns. Without frictions in financial markets, higher returns would just be reward for bearing higher risk, and investors would be indifferent across all assets at the margin.

In addition, the emergence and growing popularity of new banking activities become natural when viewed within our new framework, given productivity growth in transaction processing. First, the separation between the processing of residential mortgages by specialized non-bank financial firms and the supplying of funds by small- and medium-size banks is inevitable, given the economies of scale in handling paper work—now mostly standardized—for both borrowers and investors. Likewise, more small business loans are being securitized, as credit screening becomes more standardized with the introduction of scoring systems.

At the same time, banks are using their branch networks to sell an ever broader array of non-bank products, such as mutual funds, life insurance, and annuities, which are then managed by insurance companies. On the other hand, a growing number of non-bank financial institutions are providing bank-like services. For example, mutual funds offer money market accounts with checking account features; finance companies underwrite mortgages; and American Express issues cards to small businesses, essentially providing business loans. All

---

[22] Since non-bank FIs such as mutual funds generally charge explicit fees, the NIPA, in fact, do not need to impute any implicit financial services to them. We are merely discussing a thought experiment: Suppose mutual funds did not charge fees, but recouped their costs using interest margins. Then, what problems would result from applying the current FISIM procedures to measuring mutual fund output?

these developments demonstrate that it is the services, not the funds *per se*, that are the value added in activities of both banks and non-bank FIs, and that the services can be supplied separately from the funds.

5. Conclusion

In previous work, we developed dynamic stochastic models of financial institutions (in particular, banks). Here, we have used that work to clarify the relationship between the value added of financial institutions and portfolio risk: Financial services are separable from the funds borrowed and lent by financial firms, and the assessment of a financial firm's performance must be adjusted for the risk of its portfolio. Our new models provide a framework for identifying various service outputs of banks and other FIs, measuring their nominal output, and constructing the corresponding price indices to measure real value added. Moreover, the models resolve some long-standing debates in the banking literature. The new model-implied measure of financial firms' output also supplies a crucial input into a wide range of research on FIs, such as estimating the degree of returns to scale and productivity growth of banks, and analyzing the impact of bank mergers and other changes in industry structure.

Our new definition of financial services—consisting purely of those services produced with real resources—is perfectly consistent with the growing recognition that services by FIs are not equivalent to the financial products they offer. In fact, by focusing on the *production* activities carried out in FIs, our definition furthers the understanding of FIs' value added, which is primarily the processing of various finance-related information and transactions to facilitate the redistribution of risk and cash flows. In particular, our definition of financial services avoids the multitude of inconsistencies resulting from treating risk bearing as a service output.

The new measure of bank output implied by our models achieves a consistent treatment of FIs' (especially banks') output over time, whether or not their services are explicitly charged for and whether the associated financial securities are reported on or off the balance sheet. The new theory's key messages are: 1) financial services are not necessarily tied to, let alone proportional to, the size of financial securities on FIs' books, and 2) financial firms' value added

and productivity should be evaluated based on returns adjusted for the relevant risk exposure. The new measure, therefore, provides a consistent means for assessing the importance of financial institutions in both cross-section comparisons and time-series analyses.
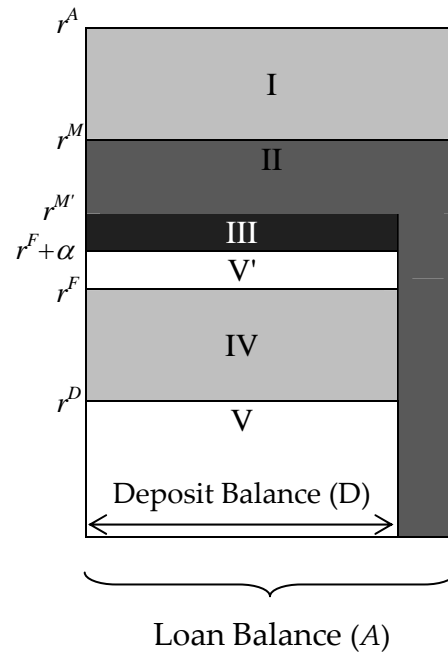
Figure 1. Decomposition of a Bank's Total Interest Receipt

Notes:

r^A: Total received interest rate on loans

r^M: Expected rate of return required on market securities with the same (or most comparable) risk characteristics as the loans

$r^{M'}$: Expected rate of return depositors would require without deposit insurance

r^F: Risk-free rate

r^D: Actual deposit interest rate

α: Insurance premium rate charged by FDIC

Area I: implicit fees for intermediation services (for example, origination and monitoring)

Area II: loan risk premium for bank shareholders

Area III: implicit deposit insurance premium (taxpayer insurance subsidy through FDIC)

Area IV: implicit fees for depositor services (for example, transaction and payment services)

Area V: interest paid to depositors

Area V': the premium banks pay (to FDIC) for deposit insurance

References

Barnett, W. A. (1980). "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory." *Journal of Econometrics* 14(1): 11–48.

Barnett, W. (1995). "Exact Aggregation Under Risk." In *Social Choice, Welfare and Ethics*, Proceedings of the Eighth International Symposium in Economic Theory and Econometrics. New York: Cambridge University Press, pp. 353–374.

Barnett, W., K. Milka, and P. Meenakshi. (1995). "Estimating Policy-Invariant Deep Parameters in the Financial Sector When Risk and Growth Matter." *Journal of Money, Credit, and Banking* 27(4): 1402–1429.

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): 175–212.

Bergquist Eric (2002). "Popularity of Outsourcing Seen Rising as Rates Climb," *American Banker*, September 16.

Bureau of Labor Statistics. (1998). "Technical Note on Commercial Banks—SIC 602: Output Components and Weights." Manuscript, December.

Cochrane, J. H. (2001). *Asset Pricing*. Princeton: Princeton University Press.

Diewert, W. E. (1974). "Intertemporal Consumer Theory and the Demand for Durables." *Econometrica* 42(3): 497–516.

Donovan, D. J. (1978). "Modeling the Demand for Liquid Assets: An Application to Canada." *International Monetary Fund Staff Papers* 25(4): 676–704.

Fama, E. F. (1984). "The Information in the Term Structure." *Journal of Financial Economics* 13(4): 509–528.

Fama, E. F., and K. R. French. (1989). "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics* 25(1): 23–49.

Fama, E. F., and K. R. French. (1993). "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33(1): 3–56.

Federal Reserve Bulletin. Various Issues.

Fixler, D. J. (2005). "Valuing Implicit Financial Services and the Impact on Commercial Bank Activity," 2004 Index Number Conference Volume, *forthcoming*.

Fixler, D. J (2004). Comments, in "Productivity in the U.S. Services Sector: New Sources of Economic Growth," by Jack E. Triplett and Barry P. Bosworth, Brookings Institution Press.

Fixler, D. J., M. B. Reinsdorf, and G. M. Smith. (2003). "Measuring the Services of Commercial Banks in the NIPA: Changes in Concepts and Methods." *Survey of Current Business* 83(9): 33–44.

Hancock, D. (1985). "The Financial Firm: Production with Monetary and Nonmonetary Goods." *Journal of Political Economy* 93(5): 859–880.

Humphrey, David (2004). Comments, in "Productivity in the U.S. Services Sector: New Sources of Economic Growth," by Jack E. Triplett and Barry P. Bosworth, Brookings Institution Press.

Kimball, R. C., and W. T. Gregor. (1995). "How Distribution Is Transforming Retail Banking: Changes Leading Banks Are Making." *Journal of Retail Banking Services* 17(3): 1–9.

Radecki, L. (1999). "Banks' Payments-Driven Revenues." Federal Reserve Bank of New York *Economic Policy Review* 5(2): 53–70.

Schreyer, P., and P. Stauffer. (2003). "Measuring the Production of Financial Corporations." Draft Final Report, OECD Task Force on Financial Services (Banking Services) in National Accounts. Presented at the OECD/BSF/SNB Workshop, August 28-29, 2003, Zurich, Switzerland.

Stiglitz, J. E., and A. Weiss. (1981). "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71(3): 393–410.

System of National Accounts. 1993. United Nations. Available at http://unstats.un.org/unsd/sna1993/introduction.asp

Triplett, Jack E. and Bosworth, Barry P. (2004). Productivity in the U.S. Services Sector: New Sources of Economic Growth. Washington DC: Brookings Institution Press.

Wang, J. C. (2003a). "Loanable Funds, Risk, and Bank Service Output." Federal Reserve Bank of Boston Working Paper 03-4. Available at http://www.bos.frb.org/economic/wp/wp2003/wp034.htm

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. Available at http://www.bos.frb.org/economic/wp/wp2003/wp036.htm

Wang, J. C. (2003c). "Productivity and Economies of Scale in the Production of Bank Service Value Added." Federal Reserve Bank of Boston Working Paper 03-7. Available at http://www.bos.frb.org/economic/wp/wp2003/wp037.htm

Wang, J. C., S. Basu, and J. Fernald. (2004). "A General-Equilibrium Asset Pricing Approach to the Measurement of Nominal and Real Bank Output." Federal Reserve Bank of Boston Working Paper 04-7. Available at http://www.bos.frb.org/economic/wp/wp2004/wp047.htm