Risk and bank service output

J Christina Wang[1]

1. Introduction

Defining and measuring bank output has long been a difficult and, hence, somewhat contentious issue. This has been made even more challenging by rapid and massive changes over the past two decades in both the form of organisations and the range and features of financial instruments offered by banks. Nowadays, banks engage in a wide range of non-traditional activities, such as underwriting firms' public offerings of debt and equity securities, standby letters of credit, and a variety of derivatives contracts (eg swaps and options). The main reason for the difficulty in measuring the output is that much bank service output is not explicitly priced. Instead, the implicit charges for financial services are bundled with interest flows between banks and their customers; on net, banks earn a positive spread between interest rates received and interest rates paid.

In this paper, we review a new measure of bank output implied by dynamic optimising models of bank operations in Wang (2003a) and Wang et al (2004). Rigorous theoretical foundations enable this new measure to resolve many extant conceptual issues. The key model implication for output measurement concerns the role of risk – in order to impute the nominal value of each implicitly priced bank service, its reference rate of interest (equivalent to "the user cost of funds") should be adjusted for the risk inherent in the associated financial instruments.[2] Otherwise, nominal bank output will be overstated.

The principle underlying these models' implications for output measurement is in fact intuitive: banks should not be counted as producing more services *merely* by taking on greater risk. This logic should be more compelling today, following the recent global financial crisis, the root of which arguably lies largely in the flawed incentive scheme that encouraged excessive risk-taking by equating more risky returns with better performance. So, in the case of bank services furnished without explicit charges, since what we observe is the sum of the implicit service revenue and net returns on bank assets associated with the services, we have to impute the former by removing the latter from the total income. A large modern literature on asset pricing under uncertainty informs the estimates of asset returns, which depend on the market context in which banks operate, particularly the way in which financial markets set the rates of return on (and thus prices of) risky securities. In the real world with risk-averse investors, securities with risks that cannot be costlessly diversified away always command returns in excess of the risk-free rate. These risky rates, as determined by markets, form the "reference rates" – the risk-adjusted opportunity cost of funds with no services attached – for estimating the returns purely due to banks' asset ownership. The

---

[1] Federal Reserve Bank of Boston, Research Department (e-mail: christina.wang@bos.frb.org).

The author wishes to thank Susanto Basu, Erwin Diewert, John Fernald, Robert Inklaar, Alice Nakamura, Marshall Reinsdorf, Paul Schreyer and Jack Triplett for helpful discussions over the years. All errors remain the sole responsibility of the author, and the views expressed herein do not necessarily represent those of the Federal Reserve System.

[2] Which components of risk – systematic vs idiosyncratic – should enter the reference rates also depends on whether there is information asymmetry between banks and holders of non-insured bank liabilities; further details below.

nominal value of bank output then equals the total income net of the pure asset returns.[3] Lastly, we discuss briefly how bank risk management activities, which have become increasingly important, affect the measurement of reference rates and, in turn, bank output.

This paper then summarises the results from empirical studies (Wang (2003b), and Basu et al (forthcoming)) that implement the new model-implied measure of bank output. These studies demonstrate that, given the available data, it is practical to correct for the systematic risk in reference rates at the level of both individual banks and the banking sector as a whole. Their numerical results indicate that, if returns due to the systematic risk are not removed from the reference rates, the upward bias in measured nominal bank output can be large – near 25% based on both micro and macro data.

Knowing the reference rates for risky assets suffices for measuring current price output. But to measure constant price output, one must also understand what exactly it is that banks actually produce. Theories of financial intermediation indicate that banks' raison d'être is to reduce transaction costs by performing tasks that mitigate the asymmetric information problems and facilitate transactions (see, for example, Diamond (1984)). So, in our framework, only bank activities – services – that fulfil these functions are considered bank output. By nature, a "service" produced by a bank is like any other professional service, eg accounting and 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 banks routinely receive compensation for these services implicitly in the form of positive interest margins. This practice of bundling implicit service revenue can pose conceptual difficulties for measuring constant price bank output, by confounding financial services and the associated financial instruments. It is therefore doubly important to emphasise the distinction: only the financial services are bank output (see also Schreyer and Stauffer (2003) on this point).

The theories imply that, to measure constant price output of implicitly charged bank services, it is best to use a quantity index of quality-adjusted transaction counts directly, instead of deflated financial balances. In particular, Basu and Wang (2008) demonstrate formally that the real quantity of bank output is unlikely to be of fixed proportion to the volume of bank assets or liabilities, whether they are book or market value, on- or off-balance sheet. The corollary is that the reference rates or the related rate spreads are not the right implicit price deflators for deriving the real output of banks. The quantity index and the relevant interest margin then imply an implicit price index for bank services provided without explicit charge.

The paper then summarises the empirical estimates in Inklaar and Wang (2007) of constant price bank output using transaction data from the US Bureau of Labor Statistics (BLS). The main result is that the output series constructed using the transaction counts exhibits rather different growth patterns from the series estimated using asset balances, translating into different patterns of productivity growth. In addition, we estimate the degree by which the existing BLS data underestimate bank output because they under-represent non-traditional banking activities such as commitment lending and securities underwriting.

The last section will analyse the implications of the theories on data priorities going forward – what additional data should be collected in order to improve the accuracy of output measurement, especially in light of the developments in the latest financial crisis?

---

[3] This imputed net income can also be viewed as corresponding to the so-called gross margin concept in the national accounts (if banking is regarded as a "margin industry" analogous to wholesale and retail trade, as well as insurance), except that, here, the margin is adjusted for the risk differential between bank assets and liabilities. See Chapter 7 of Triplett and Bosworth (2004) for an extended discussion on the relationship between the gross output vs gross margin view of banking.

2. Optimal bank operation and current price bank output

2.1 The relationship between risk and current price bank output

One must first define a concept before attempting to measure it accurately. So, what is the output of banks? Wang (2003a) and Wang et al (2004) answer this question with simple dynamic optimising models of bank operation under uncertainty, to capture the intertemporal nature of banks' operations and the integral role of risk.[4] These models focus, in particular, on the measurement of financial services provided without explicit charge. The premise underlying the measurement of implicit bank output is that a bank output measure should be invariant to how a service is compensated for – via explicit revenue, a barter for cost saving on certain inputs, or a higher-than-otherwise return on funds. The logic is straightforward. When a firm expends inputs to create a commodity that is valued by certain parties, that commodity should be recognised as an output – it is conceptually irrelevant via what medium (eg fiat money or other commodities) the firm exchanges for the output's value. Applied to banks, this principle means that, for example, it makes no difference whether a bank charges depositors for its services and at the same time pays the market rate for the depositors' funds, or pays for the funds in part with the services directly.

At the same time, the models in Wang (2003a) and Wang et al (2004) emphasise the fundamental distinction between the flow of services and the per period returns accrued to the stock of financial instruments. The returns are solely to compensate suppliers of funds for forgoing current consumption in exchange for future consumption; the returns would still be demanded even if there were frictionless financial markets and thus no need for banks to exist. Any rate of such returns corresponds to the so-called reference rate, defined as the opportunity cost of funds without any services attached. In a world with risk-averse agents and non-diversifiable risk, this opportunity cost of funds is comparable across securities only after adjusting for risk. This can be viewed as an extension of the concept of "user cost of money" (Barnett (1978)) to take account of the fact that the reward for most investments is uncertain. Consequently, the return on a risk-free security is not the appropriate opportunity cost for risky securities.

The key message from these models for estimating current price output of bank services provided without explicit charge is that, in cases where the implicit revenue is combined with the pure financial returns, the latter must be netted out of a bank's total income to impute the former. In other words, only the portion of income other than risky asset returns should be counted as bank output. To estimate the pure returns, the reference rate should be adjusted for the risk of the financial instruments associated with the service. In practice, the rate of return on a market debt security subject to the same risk but without any services attached serves as the proxy for the pure cost of funds on a loan. The corollary is that there is not a unique reference rate.

What determines this risk-adjusted opportunity cost of funds that investors expect on a financial security? Applying standard theories of asset pricing, Wang (2003a) and Wang et al (2004) show that, as long as banks themselves face no agency problems when they raise outside funds, the risk-adjusted reference rate is determined only by the correlation between a security's return and the systematic risk factors (such as the representative consumer's marginal rate of intertemporal substitution in the consumption CAPM).[5] The implication for

---

[4] Wang et al (2004) extend Wang's partial equilibrium model to a general equilibrium setting and demonstrate that all the qualitative results in Wang (2003a) continue to hold.

[5] One sufficient condition is if a bank holds a perfectly diversified portfolio. Otherwise, the reference rate for each additional financial claim depends on its covariance with the bank's existing portfolio. For details, see Wang (2009).

fixed-income securities such as loans is that the expected rate of return demanded by a lender depends on how a debt's (maturity-dependent) interest rate risk, default risk, liquidity risk, and prepayment risk covary with risk factors priced in the debt market. The expected return should, on average, equal ex post realised returns, as it is the rate required by a lender given his or her ex ante expectation of the payoff on a debt. Note that this is distinct from the interest rate that the borrower must promise in the contract and is obligated to pay conditional on his or her remaining solvent ex post. The latter is higher, because it needs to offset the positive odds of the borrower becoming insolvent ex post. In fact, this contractual rate, also referred to as yield (on market debt), rises with the borrower's probability of default and expected loss-given-default, which depends both on his or her systematic and idiosyncratic risk, even with complete markets.

Let $R_{t,n}^M$ denote the yield to maturity promised at time *t* on a defaultable market debt security with maturity *n* and a certain risk profile, but no services attached. Then, it can generally be expressed as:[6]

$$R_{t,n}^M = r_{t,n}^M + d_{t,n}^e = \left(r_{t,n}^F + r_{t,n}^P\right) + d_{t,n}^e \,. \tag{1}$$

r_t^M is the expected rate of return required on this defaultable debt. It can be further expressed as the sum of a risk-free rate $r_{t,n}^F$ and a risk premium $r_{t,n}^P$. $r_{t,n}^F$ is the yield on a debt of the same maturity but not subject to default risk, or with any embedded options (ie not callable or putable).[7] Yields on US Treasuries are arguably the best example; $r_{t,n}^F$ is the guaranteed return if the debt is held until maturity.[8] In addition, a defaultable debt must also offer a (most likely positive) return premium $r_{t,n}^P$, determined by the correlation between the probability of default or loss upon default and the risk factors priced in the market. Then, the difference between the yield $R_{t,n}^M$ and the expected return r_t^M is the default premium $d_{t,n}^e$, that is, the extra return that must be promised to investors because they are paid in full only when there is no default ex post.

Now consider an optimising bank's interest rate decision when making a loan with the same (systematic) risk profile as the above market debt. In addition to the cost of funds, the bank needs to charge implicitly for the services performed (eg screening and monitoring). As shown in Wang (2003a), the required rate of return on a loan (r_t^A), inclusive of the implicit service charge, should be optimally set as[9]

$$r_t^A = r_t^M + r_t^S = r_t^F + r_t^P + r_t^S \text{, and } r_t^S A_t = c_t \mu_t \,. \tag{2}$$

r_t^S is what we shall call the service spread, that is, the extra interest rate charged to compensate the bank for processing the loan. The optimal r_t^S satisfies the condition that the extra interest receipt (ie r_t^S times the loan balance A_t) equals the marginal processing cost of a loan c_t times the optimal markup μ_t (determined by competition in the loan market).

---

6 To be precise, this is an approximation – exact only for instantaneous returns under continuous compounding.

7 Yields on callable and putable bonds must be adjusted for the embedded option to be comparable with those on option-free debt instruments. Bonds with prepayment risk, such as mortgage-based securities, are essentially callable bonds, as borrowers are granted the option to pay off the debt (ie call the bond) before maturity.

8 Note that even for this type of debt there is still interest rate risk, ie the return is almost surely uncertain if one sells it prior to the maturity date.

9 For brevity of exposition, from now on we omit the subscript denoting maturity, unless confusion is likely.

Then, for a portfolio consisting of *N* loans of different maturities and made in different periods, its reference rate is a weighted average of individual loans' reference rates:

$$r_t^M = \sum_{i=1}^{N} \omega_i \, r_{\tau_i, n_i}^{M_i}, 0 \le \tau_i \le t, n_i \ge t - \tau_i \,. \tag{3}$$

ω_i is the portfolio weight of loan *i*, τ_i is the period in which loan *i*'s interest rate was set, and n_i is its maturity. Empirically, equation (3) is especially relevant for loans that typically have long maturities, such as real estate loans.

The optimising bank then sets the contractual interest rate (R_t^A) accordingly:

$$R_t^A = R_t^M + r_t^S + (d_t^{A,e} - d_t^{M,e}) \,. \tag{4}$$

This highlights the fact that the reference market securities are only required to have the same systematic risk as the loans; the securities can have a different expected probability of default or prepayment from the loans. Therefore, the accuracy of empirical estimates (to be discussed below) relies positively on the degree to which each category of loans and the reference securities used have the same systematic risk despite the securities' generally lower default probability.

2.2 Imputing current price output of implicit bank services

Derivations in the previous section imply that, on average, a bank's nominal output of implicit lending services to borrowers can be imputed as:[10]

$$Y_t^A = r_t^S A_t = \left(r_t^A - r_t^M \right) A_t \,. \tag{5}$$

According to the nomenclature of the 1993 System of National Accounts (SNA 93), r_t^M is the reference rate. Previous discussions make it clear why the reference rate should be risk adjusted. Consequently, the rate varies depending on the financial security or portfolio of securities associated with the services considered.

By comparison, the US National Income and Product Accounts (NIPA) currently impute bank services to borrowers using a nearly risk-free rate as the reference rate:

$$\left(r_t^A - r_t^F \right) A_t = \left(r_t^P + r_t^S \right) A_t = Y_t^A + r_t^P A_t \,. \tag{6}$$

The output value imputed according to equation (6) will be an overstatement of the actual service output. The informal justification for equation (6) is to regard $r_t^P A_t$ as compensation for a so-called "risk-bearing" service. Wang and Basu ((2006), Section 3.4) discuss at length why risk-bearing is not a productive service according to the conceptual framework of SNA 93 and, more importantly, why the NIPAs' imputation results in inconsistent accounting of the fund-using firms' output, by making it dependent on their source of funding (ie the public debt market vs banks) even given identical underlying true output. Moreover, this formula can create a perverse incentive for banks, and financial institutions in general, to seek higher risk exposure in order to appear more productive. In fact, this may partly explain what is referred to as the "reach for yield" phenomenon characterising financial intermediaries' reaction to the environment of prevailing low interest rates from 2003 to 2005. Such risk-seeking behaviour is now widely blamed for sowing the seeds of the most severe post-war financial crisis that started in the summer of 2007. So, our proposed modification of the accounting method for bank output can have an added benefit of providing financial

[10] See Wang et al (2004) for a detailed discussion of how the actual value in each period deviates from this average.

intermediaries with better incentives so that they are less likely to take excessive amounts of risk.

The value of implicit depositor services can be similarly imputed. Let D_t denote the deposit balance, r_t^D the interest rate paid, and $r_t^{M'}$ the corresponding reference rate (ie the return on a market security with the same risk). Then, the nominal output of depositor services is

$$Y_t^D = \left(r_t^{M'} - r_t^D\right) D_t . \tag{7}$$

For insured deposits in the United States (ie up to $100,000 per individual), the relevant market rate is the risk-free (Treasury) rate, ie $r_t^{M'} = r_t^F$. For the remaining, uninsured, deposits it is $r_t^{M'} > r_t^F$ because the holders are exposed to some risk in bank asset portfolios.

Note that equation (5) implies zero implicit services (to asset issuers, ie Y^A = 0) if a bank passively holds market securities in its investment portfolio, since $r^A = r^M$. Likewise, equation (7) implies zero implicit services (ie Y^D = 0) to holders of bank term liabilities (that is, commercial papers, market and privately placed bonds), since the interest rate paid equals the reference rate (ie $r^D = r^{M'}$). Also note that under virtually all circumstances (that is, whenever there are equity holders), r^M in equation (5) is greater than $r^{M'}$ in equation (7), because bank assets are typically more risky than their liabilities. In other words, the reference rates for imputing lending and depositor services almost always differ by a positive margin.

Exhibit 1 illustrates the imputation of nominal output of implicit bank services. Note that only part of a bank's net interest income constitutes nominal output of bank services; the remainder – corresponding to the risk premium $\left(r^M - r^F\right)A$ – is excluded.[11] This is precisely because the reference rate for lending services generally exceeds that for depositor services. The risk premium, along with actual interest expenses on bank liabilities, constitutes pure transfer of capital income. It is part of the factor income generated by the capital used in the borrowing firms' production or consumers' consumption. This income is then transferred from the end users of funds to the ultimate suppliers of funds – the bank shareholders. Only when all investors are risk neutral or all risk is idiosyncratic will this risk premium disappear. Exhibit 1 illuminates how our model-based output measure differs from the NIPAs' current measure, which uses a (nearly) risk-free rate as the single reference rate (see Fixler et al (2003)). As we have argued, this overstates bank output by the amount of the risk premium.

2.3 Implications for current price output of fee-generating bank services

The theories have an important measurement implication for bank services that generate explicit fees: contrary to the common opinion, not all fee income is compensation for services and so should not be automatically considered bank output. This is the only logical conclusion as long as pure asset returns, risk-free or risky, are classified as transfers; the natural corollary is that the capitalised present value of future flows of asset returns should also be regarded as transfers, no matter what it is called – fees or otherwise. A clear example is loans: if a portion of the interest income on loans is counted as transfers of pure property income, so should its corresponding present value. This means that, when a bank receives a lump sum payment from securitising a pool of loans, the only consistent solution is to partition the total receipt and count as bank output only the part that represents the loan

[11] As shown, it is ignored for simplicity that the balance of loans exceeds that of deposits, with bank equity making up the difference.

buyer's implicit payment for the bank's screening services. The rest of the receipt, corresponding to the value of the loans themselves, is but a transfer of property income.[12]

In fact, this measurement implication is particularly relevant for the accounting of non-traditional bank activities, most of which generate explicit fees. Inklaar and Wang (2007) show that the same logic applies to any exotic securities underwritten by banks, by proving that it holds for options. As first argued by Black and Scholes (1973), options can be used to express virtually all contingent claims. For example, loan commitments or lines of credit can be modelled as put options written by banks to their borrowers. Credit derivatives such as credit default swaps (CDS) can also be expressed as options.

2.4 Empirical estimates of current price output of implicit bank services

Several studies have implemented the bank output measure implied by the above theoretical framework using both bank-level and industry-level data. Collectively, they impart the key message that it is feasible to implement the new output measure using publicly available data, even though the estimates can be noisy because of data limitations. To improve the accuracy, the theoretical framework summarised above provides guidance on what additional data to collect and how to prioritise the effort.

Wang (2003b) uses balance sheets and income statements filed quarterly by all US bank holding companies (BHCs) to their regulator (the Federal Reserve), combined with interest rates on market debt securities most comparable to bank loans, to estimate the current price output of individual BHCs according to the new measure.[13] The new output series, which is later updated to Q1 2008, is then compared with estimates for the two existing measures – one based on the real book value of assets (and liabilities in some cases) and used in virtually all micro studies of bank productivity and the other used in the US NIPA.

The reference interest rate for the loan portfolio associated with each type of borrower service is approximated using publicly available data on the expected rate of return (r^M) on market debt with the most similar risk – depending on attributes such as term, default and prepayment probability, and borrower type (eg firms vs consumers). The regulatory filings provide data only on a few broad categories of loans, which we match with market securities as follows: residential mortgages with mortgage-backed securities (MBS), consumer loans with certain asset-backed securities (ABS, eg based on credit card receivables), and commercial and industrial (C&I) loans with commercial paper.[14] As such, each r^M we use can be a noisy proxy for the true reference rate on the matched category of bank loans, since we do not observe the maturity and risk composition of loans and thus cannot assess how accurate the match is. On the other hand, using market debt returns almost certainly underestimates the true risk premia on loans, since the market securities have, on average, much lower realised default rates and are thus likely to command a lower risk premium as well.

The main findings of BHC-level output estimates can be summarised as follows. First, output according to the new measure is more than 25% smaller than that based on the US NIPA measure (ie using the risk-free reference rate). More importantly, output estimates according to different measures exhibit different time series properties. The new series is more volatile

---

[12] To see the logic clearly, consider the alternative: if the bank receives an explicit servicing fee along with the market value of just the loans, then only the servicing fee should be considered the bank's output.

[13] Accounting data from regulatory filings both by banks and by BHCs are publicly available from the Federal Reserve Bank of Chicago at: www.chicagofed.org/economic_research_and_data/bhc_data.cfm.

[14] Since the payoff on whole loans can be synthesised using MBS of different ratings, the loan reference rate should ideally be computed as the weighted average of differentially rated constituent MBS yields.

over time, and it appears to be affected more by the recessions during the sample period (1990 to 1991 and 2001). This can be largely explained by the countercyclical share of market securities on banks' balance sheets and thus their returns (in excess of the risk-free return) in overall income. Since the NIPA measure counts such "excess" returns on securities as part of bank output, whereas the new measure does not, the latter is more volatile over the business cycle. Second, the new series shows a higher growth rate than the two existing measures, especially in the mid- to late 1990s. This seems consistent with the observation that investment grew steadily during those years. Lastly, 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 term. This may be viewed as informal support for the new measure of bank output if one assumes that the free entry condition holds in banking at least in the long term.

Basu et al (forthcoming) estimate the current price output for the US banking industry. At the aggregate level, more data are available on the risk characteristic of bank loan portfolios to enable a more accurate estimate of the overall (credit) risk premium. In particular, the Survey of Terms of Business Lending (STBL), conducted by the Federal Reserve, provides summary information on the risk rating and interest rate collected at the individual C&I loan level. Since the survey oversamples large banks, it offers a fairly representative mapping of the relationship between the credit risk and loan interest rate for the banking industry as a whole, enabling us to estimate the risk premia on C&I loans more accurately. In short, it is likely that industry output is estimated with less measurement error using aggregate data.

The industry-level estimates confirm the qualitative results from the BHC-level analysis in Wang (2003b). Even based on conservative estimates of the risk premium (that is, the difference between the risk-adjusted reference rate and the risk-free rate), incorporating risk in reference rates (see Exhibit 2) is shown to be quantitatively important (see Table 1). Compared to bank output calculated using the risk-adjusted reference rate, output computed using the risk-free reference rate according to the current NIPA method overstates imputed bank output by 45% on average (see Exhibit 3). It, in turn, overstates total bank output, which also includes explicit fees, by 21%. Only bank services to households contribute to GDP, whereas the services to firms (non-financial as well as financial) are counted as intermediate (service) input. In sum, the impact on the US GDP amounts to an overstatement of 0.3 percentage points. Furthermore, the new measure of bank output lowers the estimate of the share of capital in banks' value added from 59% (higher than that of the petroleum refining industry) to 42% (the level for all private industries), and lowers the internal rate of return on fixed capital of the banking industry from 17.8% (9 percentage points above the average for all private industries) to 6.8%. These new estimates seem more plausible, providing additional support to the theory that the new measure of bank output improves upon the current NIPA measure.

3. Constant price bank output

3.1 Flow of bank services not proportional to the stock of financial instruments

As difficult as the measurement of current price bank output is, it is even more challenging to measure the constant price (ie real) output of banks. Apart from the major difficulty that most other service industries are confronted by (adequate quality adjustments), one must first answer the fundamental conceptual question when it comes to measuring the real output of banks: what exactly is it that banks produce? Only in accordance with the definition can the right measure of the real value of bank products be derived.

As emphasised in the theoretical studies mentioned above, the value added of banks is not the funds they borrow and lend, but the array of services that they provide to facilitate the exchange of funds.[15] Services are generally intangible and many of these bank services are not even explicitly charged for, while financial instruments such as loans are often the most tangible manifestation of those services. It is understandable, therefore, that real financial balance makes for an appealing proxy for the quantity of bank service output. But this requires the implicit assumption that the former is in constant proportion to the latter.

There is, however, little theoretical basis for this restrictive assumption. Both Wang (2003a) and Wang et al (2004) argue that, generally, there exists no invariant value mapping between flows of bank services and stocks of the associated holding of financial instruments, regardless whether the stock is measured by book or market value. To understand why financial balances are a poor proxy for financial services, consider a simple example. Suppose loan A has a smaller balance but is more risky than loan B, then monitoring loan A may well require more bank services, resulting in a bigger (implicit) income. An output measure based on loan balance will, however, give the opposite result. This example illustrates intuitively the basic problem with using financial balance to measure bank output: any single attribute of financial instruments is inevitably a poor proxy for the quantity of the associated services. These instruments are fundamentally contracts of contingent claims and thus almost certainly have multi-dimensional attributes, all of which can affect the amount of bank services produced in creating the contracts. For the purpose of bank output measurement, any attribute matters only to the extent that it affects the value and quantity of the services produced.

Basu and Wang (2008) make this point more formally through a general equilibrium Baumol-Tobin model where households need bank services to purchase consumption goods and bank deposits are the single medium of exchange in the economy. Their model shows that financial services are proportional to the stock of assets only under restrictive conditions, especially troubling among which is the requirement of constant relative technologies in the financial sector. In contrast, measuring real financial output by directly counting the flow of actual transaction services is a method robust to technological changes.

In fact, these models have further argued that the supply of banking services and the holding of financial balances can be carried out separately. Developments in the banking industry in recent decades provide strong support for this argument; it has become increasingly prevalent where financial firms provide financial services without holding the associated securities, and vice versa. More generally, Wang el al (2004) reason that the (often implicit) services produced by banks in making a loan are qualitatively the same as the services produced in underwriting a credit derivatives contract. A loan subject to default is shown to be equivalent to a default-free bond combined with a short position in a put option (Merton (1974)). Since all the credit risk in a loan subject to default risk lies in the embedded put option, issuing a loan involves similar tasks (eg screening and monitoring) as writing a put option to the borrower.[16] In short, the new output measure is invariant to the balance sheet status of financial instruments resulting from a bank service.

The fundamental distinction between service flows and asset balances may be less obvious in traditional bank lending, where the intermediary both performs the services and holds the assets. But it is no less intuitive once we focus on the underlying services. For instance, originating a $1 million residential mortgage almost certainly involves much less than

[15] The funds themselves can be thought of as a special type of "intermediate input" – special in that they are not the output of any other producing entities.

[16] In recent years, such implicit options have in fact been made explicit and traded in the rapidly growing credit derivatives market (eg credit default swaps).

10 times the services in originating a $100,000 mortgage, given the ready availability of credit scores for individuals. Fixed proportionality is even harder to defend for services to depositors vs account balances. Studies on payment services suggest that the amount of work in clearing cheques is independent of the dollar figure of the cheques or the account balance (see, for example, Radecki (1999)). The same is true for many other depositor services (eg issuing money orders, transferring funds, etc). These empirical facts are all consistent with the theoretical argument in Basu and Wang (2008).

3.2 Empirical estimates of constant price output of bank services

The theoretical models imply that it is preferable to measure the flow of services directly instead of using proxies such as balance sheet stock values just because they are more readily observed. Since traditional bank activities often generate no explicit fees for services, the usual method – deflating revenue using price indices to estimate indices of real output – is seldom applicable. The alternative is thus to construct quantity indices of tasks performed by banks; we will call this the "activity counts" method. If count data of distinct bank activities exist, such as the number of a particular category of loans originated within a quarter, then constant price output can be measured as indices of the activity counts. Ideally, each type of activity should be defined sufficiently narrowly so that there are no quality differentials and the quantity of output equals the simple sum. To aggregate across different services, the nominal share of each activity in total service revenue (ie nominal output) serves as the weight. The combination of current and constant price output then implies the implicit price deflator.

Inklaar and Wang (2007) use this approach to estimate the real output of implicit bank services of the industry as a whole. They make use of the activity counts that the BLS has in fact been tabulating since at least 1987. Specifically, for lending services, the BLS tallies the number of four types of (new and outstanding) loans – real estate, credit card, consumer instalment and C&I loans; for depositor services, the BLS collects the number of cheques cleared and electronic fund transfers (to proxy for transactions associated with demand deposits), deposits and withdrawals related to time and savings deposits, and ATM transactions; and trust services are proxied by the number of accounts managed. The growth rates of these quantity indices are then aggregated using their respective (implicit) revenue weights, which is better grounded in theory than the employment weights used by the BLS. For example, growth in the numbers of C&I, residential real estate and consumer loans is weighted respectively by the imputed implicit revenue attributed to the three corresponding types of loans. The authors show that, over the sample period, simply deflated balances tend to overstate the growth of implicit bank output relative to the measure according to the activity index. The real quantity index combined with the imputed implicit income then implies a price deflator. The authors also illustrate that, if one must use loan balances to estimate real implicit output, less biased proxies can be constructed by deflating using the price indices of the physical assets funded by the loans. Residential mortgage loans are a good example: deflating the loan balances using the general consumption or GDP deflator biases up the growth of residential lending in the years when house prices appreciated much faster than the general price level; instead, it is better to use a house price index.[17]

In addition to the traditional banking activities covered by the BLS data, Inklaar and Wang (2007) also account for the novel banking activities using data from call reports (financial reports filed by banks to their respective regulatory agencies) as well as periodic surveys by

[17] We show that the proxy of residential mortgage lending services derived by deflating the balances with a house price index can be reasonably accurate to the extent that the fraction of home purchases financed by borrowing is stable.

the Census Bureau. These encompass securitisation and servicing, investment banking (including advisory, brokerage and underwriting), loan commitments and sales of insurance contracts. As this study shows, fees and commissions from these activities as a fraction of total bank income (both explicit charges and implicit revenue) has been trending up rapidly over the past two decades. Each category of explicit income is then deflated using the best matched implicit price index derived earlier. The authors show that ignoring these non-traditional banking activities results in a downward bias in the estimate of total bank output. At the same time, they readily acknowledge that data limitations, especially in terms of detailed data for quality adjustments and the construction of accurate price deflators, introduce possibly non-trivial measurement errors into the estimates.

Wang (2003b) estimates real output for individual BHCs. Since no activity data are available at this micro level, quantity indexes are derived by deflating revenue using the aggregate price deflators for bank services. The deflators for implicitly priced bank services (whose revenue is imputed as in the previous section) are derived as described above. The estimates of bank output at the BHC level (updated to Q1 2008, most recently) are qualitatively similar to those at the industry level: the new output series exhibits faster (owing to the non-traditional activities), albeit more volatile, growth, compared with the book value-based and US NIPA output measures. Wang (2003c) then applies the real output series to estimate productivity and returns to scale at the BHC level (also updated to Q1 2008). Average productivity growth across all the BHCs (weighted by the sum of explicit and implicit service revenue) estimated using the new output series is faster but more volatile, and there appears to be at least a moderate degree of increasing returns to scale, as opposed to the constant returns to scale generally found in the micro banking efficiency studies.[18]

Not surprisingly, the BLS activity count data illustrate the primary empirical difficulty with the direct approach to bank service accounting – activities are classified so broadly that there is often much quality heterogeneity even within the same category. For instance, C&I loan portfolios as well as their evolution over time may differ substantially across banks in terms of *characteristics that matter for the amount of services performed*, such as the distribution of loan size and credit rating, industry and geographic mixes, etc. Consequently, using the simple sum of C&I loan numbers to measure the output of total C&I lending may introduce serious measurement errors, probably even biases. Moreover, screening and monitoring may be two rather different activities. So, ideally, we should instead tally the number of, along with the interest rate on, new vs outstanding loans for each sufficiently narrowly defined category within which the services entailed are similar enough to allow a simple count of transactions to measure bank output.

An alternative to the direct approach is feasible for services that are charged for explicitly in some cases but not in others. Nowadays, for example, originations of residential mortgages often carry explicit fees in the United States, but sometimes the fee is amortised and subsumed into interest over the loan's maturity. Since essentially the same services are performed in both cases, explicit price indices derived from former cases can be used to deflate the imputed implicit service revenue in the latter. This approach may be more useful going forward, as banks have been charging explicitly for an increasing share of services (Stiroh (2000)).

[18] Because of the greater volatility of the new output series and hence a more severe errors-in-variable problem, it leads to a bigger upward (downward) bias in the estimate of the degree of returns to scale from the cost (production) function. Since the former is above 2 and the latter slightly below 1, the correct value is most likely to be moderately above 1.

4. Concluding remarks

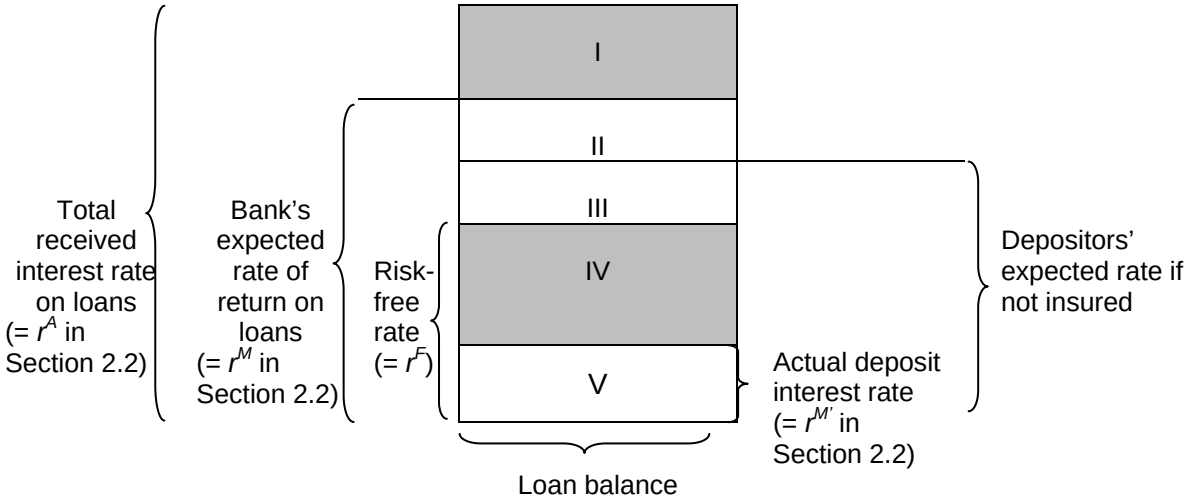
The statistical properties of an output estimate cannot per se establish its validity. Instead, it should be justified on theoretical grounds – consistency with basic economic theories that can rationalise the operation of the firms concerned. This is the principle underlying some recent efforts (Wang (2003a), Wang et al (2004)) to model bank operations. Consistent with widely received theories of financial intermediation, production and asset pricing, these models yield a coherent framework for measuring the output of bank services. The key is to recognise that banks perform qualitatively the same services – processing financial information and transactions – and so their output should and can be measured in the same way whether or not the services are priced explicitly or associated with any on- or off-balance-sheet financial claims. Implementing the output measure implied by these models generally entails constructing the quantity index of each type of service based on quality-adjusted transaction counts. The true revenue from each type of service serves as the weight for aggregating across different service types. In cases where implicit charges for services are bundled with asset returns, the true service revenue needs to be imputed from the total bank income by removing the risk-dependent returns on the associated assets.

Follow-up empirical studies (Basu et al (forthcoming), Inklaar and Wang (2007)) demonstrate that the measure of bank output implied by the above theoretical framework can, in fact, be implemented using publicly available data. The sources of data range from yields on market debt most comparable to bank loans in terms of risk characteristics, financial statements filed by banks to their regulators, bank transaction counts and revenue by activity, as tabulated by statistical agencies. The primary finding is that both current and constant price output based on the new measure have rather different levels and time paths of growth than those based on the extant output measures over the past 20 years. Different patterns of output growth will translate into different patterns of productivity growth. These differences suggest the need to examine whether bank output has been overstated during this period and, more generally, to what extent the post-1995 productivity acceleration according to official data was due to the mismeasurement of financial service output.

This conceptually sound measure, however, can only be imprecisely implemented at present because of data limitations. We therefore advocate that an effective approach to improving empirical estimates is to collect additional data that are called for by the theory. Arguably, the most important among such data needs is the transaction counts of different types of bank activities, starting with those accounting for the largest shares in (explicit plus implicit) bank service revenue. Bank activities should be classified according to characteristics that matter for the amount of services provided, so that each type corresponds to homogeneous services. For example, originating conforming residential mortgages in the United States requires an essentially uniform set of tasks and can be considered a single type of activity. In addition, in order to impute implicit service revenue more precisely, we need more detailed data on the interest rates charged along with risk ratings of the associated financial instruments. Such data may become more readily and widely available as they will be required by new regulatory rules to enable better risk management within each intermediary and more effective supervision. Meanwhile, when one has to use approximations for practical purpose, one must be clear about the conditions under which the proxies are appropriate.

Exhibit 1

Decomposition of a bank's total interest receipt



Notes:

1. The content of each area:

Area I: implicit fees for intermediation services in lending (eg origination and monitoring)
Area II: loan risk premium
Area III: deposit insurance premium
Area IV: implicit fees for transaction and payment services (eg 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 x deposit balance

NB: when loans are funded by deposits plus equity, the loan balance exceeds the deposit balance. An equivalent calculation is to adjust the two deposit-related rates on the right of the block and the risk-free rate 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 (eg having AAA-rated bonds outstanding).

Exhibit 2

Average interest rates on loans and deposits, and corresponding reference rates, Q2 1997–Q4 2007

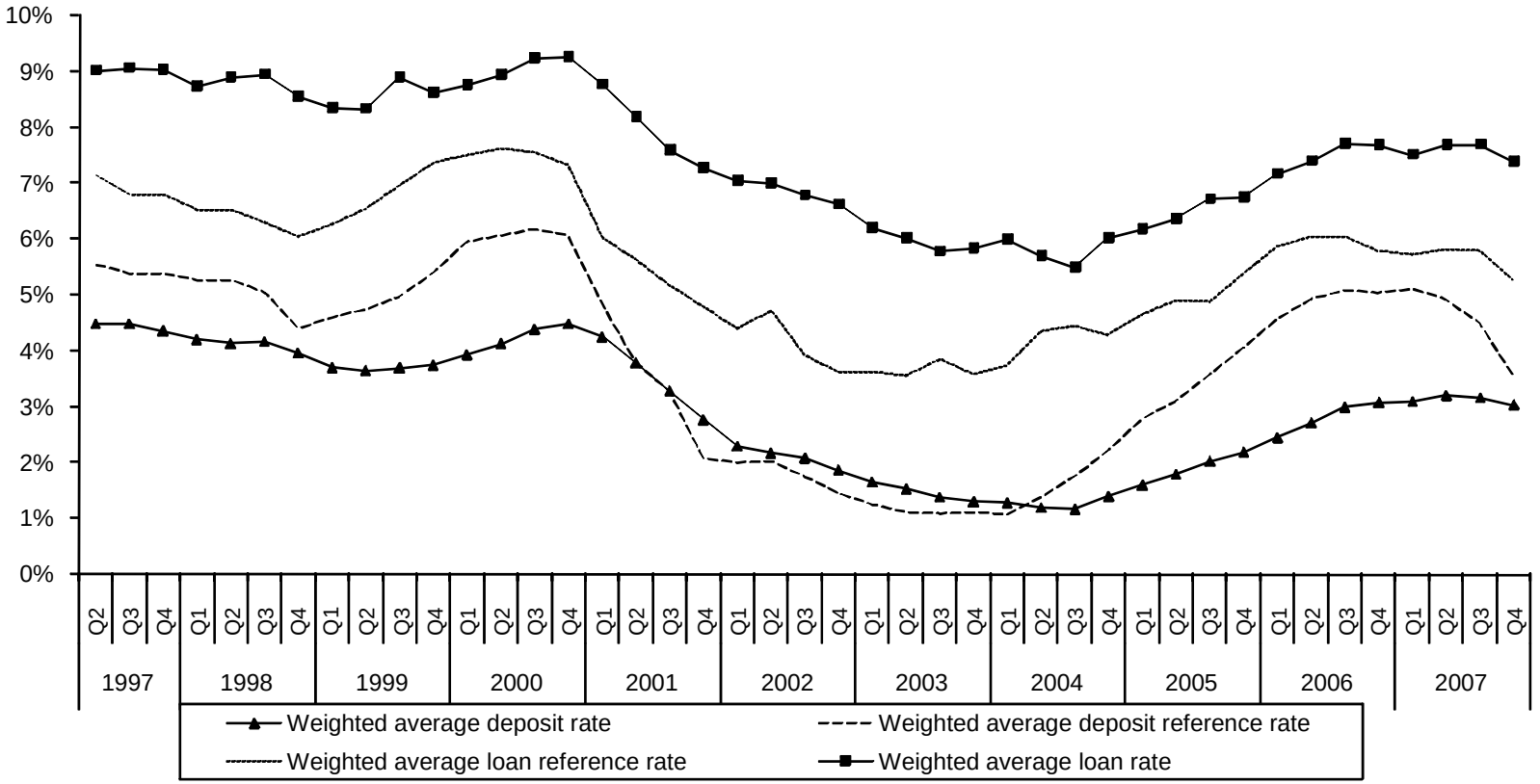


Table 1

The effect of risk adjustment: imputed output of US commercial banks at current prices in Q4 2007

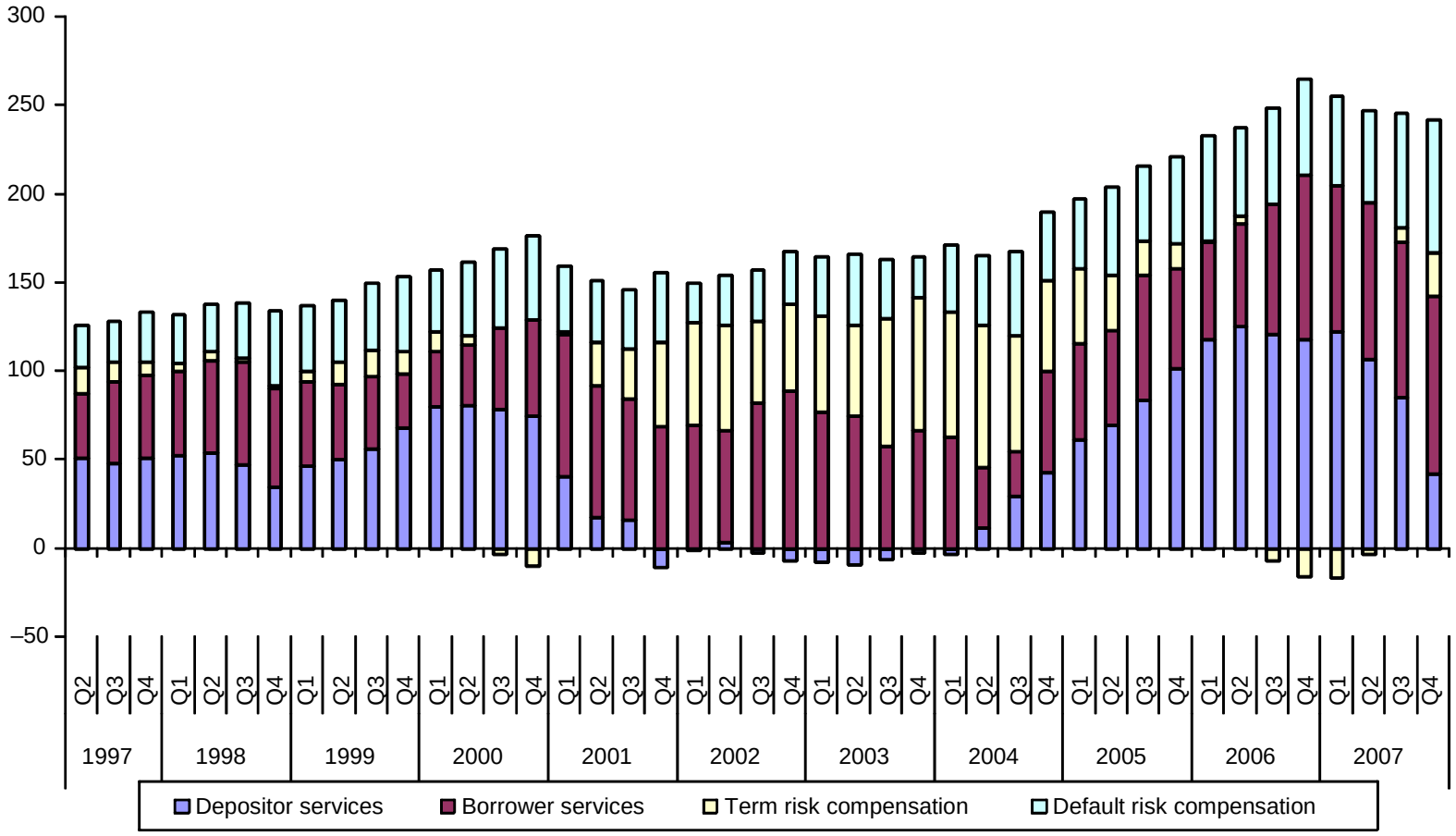
USD billions

	Average balance	Interest flow	Annualised interest rate	Reference rate			Imputed output		
				risk-free	term	default & term	risk-free	term	default & term
Deposits in domestic offices	5,504	152	2.8%	3.5%	3.5%	3.5%	39.0	42.5	42.5
Demand deposits	486	0.0	0.0%	3.5%	3.5%	3.5%	16.9	16.9	16.9
Time and savings deposits	5,018	152.2	3.0%	3.5%	3.5%	3.5%	22.1	25.6	25.6
Loans in domestic offices	5,471	395	7.2%	3.5%	4.0%	5.7%	205.0	177.0	83.7
Real estate loans	3,545	235.3	6.6%	3.5%	4.3%	5.8%	112.2	84.3	31.5
Consumer loans	804	80.9	10.1%	3.5%	3.5%	4.9%	53.0	52.9	41.1
Commercial & industrial loans	1,123	78.8	7.0%	3.5%	3.5%	6.0%	39.8	39.8	11.1
Total	10,975	547	6.6%	3.5%	3.8%	4.6%	244.0	219.5	126.2

Exhibit 3

Imputed output of US commercial banks and risk compensation at current prices, Q2 1997–Q4 2007

USD billions



References

Barnett, W A (1978): "The user cost of money", *Economic Letters*, vol 1, no 2, pp 145–9.

Basu, S, R Inklaar and J C Wang, "The value of risk: measuring the services of U.S. commercial banks", *Economic Inquiry*, forthcoming. Also available at: www.bos.frb.org/economic/wp/wp2008/wp0804.pdf.

Basu, S and J C Wang (2008): "Technological progress, 'money' in the utility function, and the 'user cost of money'", paper presented at the American Economics Association meeting. Available at: www.vanderbilt.edu/AEA/AEAStat/.../Basu_and_Wang_1-2008.pdf.

Black, F and M S Scholes (1973): "The pricing of options and corporate liabilities", *Journal of Political Economy*, vol 81, no 3, pp 637–54.

Diamond, D W (1984): "Financial intermediation and delegated monitoring", *Review of Economic Studies*, vol 51, no 3, pp 393–414.

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*, vol 83, no 9, pp 33–44.

Inklaar, R and J C Wang (2007): "Not your grandfather's bank anymore? Consistent measurement of non-traditional bank output", paper presented at the NBER/CRIW Summer Institute 2007. Available at: www.nber.org/~confer/2007/si2007/PRCR/wang.pdf.

Merton, R C (1974): "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, vol 19, no 2, pp 449–70.

Radecki, L (1999): "Banks' payments-driven revenues", *Federal Reserve Bank of New York Economic Policy Review*, vol 4, no 2, pp 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, Zurich, 28–29 August.

Stiroh, K (2000): "How did bank holding companies prosper in the 1990s?", *Journal of Banking and Finance*, vol 24, no 11, pp 1703–45.

Triplett, J E and B P Bosworth (2004): *Productivity in the U.S. services sector: new sources of economic growth*, Brookings Institution Press, Washington DC.

Wang, J C (2003a): "Loanable funds, risk, and bank service output", *Federal Reserve Bank of Boston Working Paper Series*, no 03-4. Available at: www.bos.frb.org/economic/wp/wp2003/wp034.htm

———(2003b): "Service output of bank holding companies in the 1990s and the role of risk", *Federal Reserve Bank of Boston Working Paper Series*, no 03-6. Available at: www.bos.frb.org/economic/wp/wp2003/wp036.htm

———(2003c): "Productivity and economies of scale in the production of bank service value added", *Federal Reserve Bank of Boston Working Paper Series*, no 03-7. Available at: www.bos.frb.org/economic/wp/wp2003/wp037.htm

——— and S Basu (2006): "Risk bearing, implicit financial services, and specialization in the financial industry", *Federal Reserve Bank of Boston Working Paper Series*, no 06-3. Available at http://www.bos.frb.org/economic/ppdp/2006/ppdp063.pdf

———S Basu and J G Fernald (2004): "A general-equilibrium asset-pricing approach to the measurement of nominal and real bank output", Conference volume on Price Index Concepts and Measurement, forthcoming, Conference on Research on Income and Wealth (CRIW). Available at: www.bos.frb.org/economic/wp/wp2004/wp047.htm.