Preliminary
Comments welcome

PRCR
7/18/06
9:45 AM

Technological Progress, "Money" in the Utility Function, and the "User Cost of Money"

Susanto Basu and J. Christina Wang*

Abstract: Financial institutions provide their customers a variety of unpriced services and cover their costs through interest margins—the interest rates they receive on assets are generally higher than the rates they pay on liabilities. In particular, banks pay below-public-market interest rates on deposits while charging above-public-market rates on loans. Various authors have suggested that this situation allows one to measure the real quantity of financial services implicitly provided as proportional to the real stocks of financial assets held by consumers. We present a general-equilibrium Baumol-Tobin model where households need bank services to purchase consumption goods. Bank deposits are the single medium of exchange in the economy. The model shows that financial services are proportional to the stocks of assets only under restrictive conditions, including constant technologies in the financial sector. In contrast, measuring real financial output by directly counting the flow of actual services is a robust method that is unaffected by technological changes.

JEL Classifications: D24, E41, O47, D91

First draft: May 19, 2006
This version: July 5, 2006

*Basu: Department of Economics, Boston College and NBER; susanto.basu@bc.edu
Wang: Research Department, Federal Reserve Bank of Boston; christina.wang@bos.frb.org
We thank Erwin Diewert, Robert Inklaar, Marshall Reinsdorf, Jack Triplett and participants at the 2006 Ottawa Productivity Workshop for comments. The views expressed in this paper are solely those of the authors and do not necessarily reflect official positions of the Federal Reserve Bank of Boston or the Federal Reserve System.

Introduction

Accurate measurement of service output has become increasingly important for correctly measuring GDP and productivity growth. Services now account for nearly 60% of U.S. GDP, and the share continues to grow. But measuring service output, especially with adequate quality adjustment, remains challenging. Within the service sector, financial services are among the most difficult to measure, since it is not even clear how to measure *nominal* output, let alone *real* output.[1] The main reason is that financial firms often do not charge explicit fees for their services. Instead, they routinely earn substantial income in the form of a positive interest margin—the spread between interest received and interest paid. This measurement problem is made even more challenging nowadays by rapid and massive expansion in the range and features of financial instruments offered by financial institutions.

It is generally agreed that financial institutions provide their customers a variety of real services, and recoup their costs by earning a positive interest margin – generally higher interest rates received on assets than paid on liabilities. The arguably most prominent case is banks: the interest rates paid by banks on deposit balances are routinely lower than those on market securities with comparable risk (and the rates charged on loans are higher than those on comparable market securities). Depositors and borrowers are willing to accept these non-market rates because they value the services they receive. In this paper, we study the issue of measuring implicitly-priced financial services. To make the exposition intuitive, the services we model most closely resemble banks' services to depositors. However, we emphasize that much of the logic of the paper carries over to analyzing bank services to borrowers, and also applies to analyzing the services of non-bank financial institutions, such as insurance companies.

In the situation where financial institutions are compensated via an interest margin, one can measure the nominal output (akin to "gross margin"[2]) of depositor services as the interest that depositors forego by accepting a below-public-market interest rate, i.e., as the product of the interest rate spread and the current value of deposit balances. Various authors have gone further and suggested that one can measure the *real* quantity of financial services implicitly provided as linearly proportional to the real balance of deposits—which implies, of course, that the price index is linearly proportional to the interest rate spread. That interest rate spread (in general, many spreads, if there is a variety of monetary assets) is often termed "the user cost of money."

The literature that provides the theoretical foundation for this measurement method starts with Barnett (1978, 1980) and Donovan (1977). These papers, as well as those that follow, assume as a primitive that monetary assets enter consumers' utility function directly. This assumption follows the shortcut to modeling money demand pioneered in the classic paper of Sidrauski (1967). It has been clear to monetary economists from the start of that literature[3] that the presence of money in the utility function (MIUF) is a simplified description of a more complex reality, where money somehow aids consumers by making transactions easier. But without an explicit derivation giving rise to MIUF, it remained unclear whether this shortcut could ever be rigorously justified, and if so under what conditions.

Nearly 20 years later, such a derivation was finally provided in an important paper. Feenstra (1986) uses the tools of duality to show "functional equivalence" between money in the utility

---

[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.
[2] This value is analogous to the so-called "gross margin" in wholesale and retail trades, except that here the nature (i.e., risk) of what is bought and sold – funds – is altered (as discussed further below). It can be construed as the nominal value of a bank's "gross output," i.e., including compensation for intermediate inputs (e.g., stationery and utilities) used in producing depositor services—but not the actual funds borrowed and lent. Since such purchased inputs account for a tiny share in financial firms' "gross margin," we ignore them and use "output" synonymously with "value added" throughout the paper.

function and a class of general transaction cost functions in the budget constraint. He shows, furthermore, that those cost functions can be derived from a variety of models of money demand in which money reduces the transaction costs of purchasing consumption goods.

More importantly, Feenstra (1986) makes clear (though implicitly) that the existence of an equivalent MIUF formulation is a convolution of the consumer's primitive preferences (which depend *only* on his consumption) with a technology for making transactions that is assumed to be a stable function of consumption and real (money) balances. But this leads to a natural question: in this era of massive financial innovations and deregulation, is it innocuous to assume that such a transactions technology will indeed be stable over time? If the transactions technology changes, due to technological progress in the financial sector, does the MIUF representation also become unstable over time? And, to return to the question of measurement where we started, if the MIUF representation indeed turns out to be unstable, what are the consequences for the inspired shortcut of measuring real financial output as proportional to real balances?

These are the questions that we address in this paper.

To address these issues, one needs a model where (1) there are transactions costs, (2) financial institutions provide services to reduce these costs, (3) providing services is costly, and (4) that cost is recouped via an interest margin. In addition, the model needs to be a general equilibrium in order to understand how technological changes on the firm side affect the functional equivalence result on the consumer side. We thus present a general-equilibrium model of the demand for monetary assets that follows the seminal work of Baumol (1952) and Tobin (1956). We can then easily compare our results to those of Feenstra's (1986), since the Baumol-Tobin model is one of the specific cases that he analyzes.

---

[3] Not least to Sidrauski himself, who wrote of his own paper that "it is incomplete and the assumptions on which it is based are relatively crude abstractions" (1967, p. 534).

In brief, we find that if all technologies are stable and the economy is at a steady state, then Feenstra's (1986) functional equivalence result still applies—there is a stable MIUF representation. In this case, real bank service output is proportional to real deposit balances. Unfortunately, this result is not robust. We show that there is no stable MIUF representation whenever financial sector technology changes, in which case the ratio of bank service output to deposit balances becomes unstable as well. It is thus no longer valid to use real deposit balances to construct an index of real financial output. This means the approach proposed in studies such as Fixler (2006), though easy to implement and thus appealing, is valid only under conditions that are likely to be too restrictive. For example, one could never use output data from the asset-based approach to calculate TFP growth in financial services, since the output measure would be valid only if TFP growth were zero.

We suspect that, in this era marked by rapid and pervasive innovations in the financial industry, there is a multitude of reasons why the relationship between real asset balances and real financial service consumption is unlikely to be stable.. Therefore, instead of using real asset balances, one should construct quantity index for financial services using the same methods that are used to measure other service sector outputs in general. In particular, it should be recognized that the services underlying financial transactions are qualitatively similar to professional services such as consulting and accounting. Real quantity indices can then be constructed for precisely defined financial transactions. In fact, real output of various bank services to borrowers and depositors as traditionally measured by the BLS are exactly such quantity indices (see the BLS *Technical Note*, 1998).

Such a set of real quantity indices of financial services immediately imply a set of price deflators, given nominal output. These deflators are almost surely not proportional to interest rate differentials on the associated financial instruments, such as the interest rate spread between a bank's deposits and money market securities.

This paper is organized as follows. Section I presents our general-equilibrium Baumol-Tobin model of financial transactions. Section II uses the model to analyze the existence of a stable MIUF representation of households' preferences, and applies the results to measurement of financial sector output. Section III contains more general reflections on this set of issues. Section IV has concluding observations, and suggests future research.

I. A General-Equilibrium Baumol-Tobin Model

In this section we study a modified version of the well-known Baumol-Tobin model to analyze the relationship between real "money" balances and transaction services when holdings of a medium of exchange lower transaction costs. We start with this model, whose features best resemble payment services, in order to both build intuition and compare directly with related previous studies, particularly Feenstra (1986). We then show that its conclusions apply more generally to most other financial transaction services as well.

Furthermore, since our focus is to uncover the conditions under which there is a constant relationship between real balances and transactions, we start with the case without uncertainty, in order to highlight the key intuition underlying those conditions. We also keep the non-essential features of the model simple so that the effects of changes in the technology for producing financial services can be derived analytically. We begin by introducing the optimization problem of the representative member of the four respective categories of players in the economy: households, a goods-producing firm, market index mutual funds, and banks.

A. Consumers

To facilitate comparison, we formulate the consumer's problem similarly to Feenstra (1986).[4] One major difference is that Feenstra treats the demand for money – the medium of exchange in general – as a demand for currency narrowly, whereas we abstract from currency altogether and model bank deposits as the sole form of money, since currency is now used in a rather small fraction of payments and accounts for an even tinier fraction of monetary assets in a modern economy. The consumer's problem in this model thus becomes one of choosing whether to keep her assets in a mutual fund that pays a high rate of return, or hold some assets as bank deposits that can be used to purchase consumption goods but pay a lower interest rate.

Specifically, households supply labor inelastically; at the same time, they own the financial intermediaries and, indirectly, the goods-producing firm. All households are identical, and the representative household maximizes with perfect foresight the present value of discounted utility over an infinite horizon:

$$\sum_{t=0}^{\infty} \beta^t U\left(C_t\right), \tag{1}$$

subject to the constraints:

$$C_t + P_t^N N_t + E_t + \bar{D}_t = W_t + \Pi_t + \left[\left(1+r_t\right) E_{t-1} + \left(1+r_t^D\right) \bar{D}_{t-1}\right], \tag{2}$$

$$\text{and} \quad C_t = 2N_t \bar{D}_t . \tag{3}$$

In the objective function (1), $\beta < 1$ is the discount factor and C_t is consumption in period *t*. We assume conventional first- and second-derivative properties for the utility function: $U' > 0$ and $U'' < 0$. In (2), W_t is the wage (one unit of labor is supplied inelastically) and Π_τ is economic profit (if any) from the ownership of the firm and financial institutions. E_t is the consumer's

[4] Feenstra also analyzes a number of other models of the demand for money, and proves results for a generalized transactions technology. The Baumol-Tobin model satisfies the assumptions of the general technology.

holding of the market portfolio in an index mutual fund. $\bar{D}_t$ is the *average* level of bank deposits within a period.

The real rate of return on equity held in the mutual fund is r_t and the interest rate paid by banks on deposits r_t^D. We will show shortly that $r_t^D < r_t$.[5] We date the interest rate by the period when the payments are distributed, so r_t is the interest rate promised on savings at time *t-1* and paid at *t*. Thus, the right-hand side of the budget constraint, (2), is the sum of the resources the household has to spend at time *t*: wage and profit income plus gross interest income from its holdings of capital and bank deposits. The left-hand side gives the uses of funds: to purchase consumption and mutual fund services ($P_t^N N_t$, to be discussed later), or be saved as capital or deposits.

The optimization problem for financial transactions embodies the logic of the money demand models of Baumol (1952) and Tobin (1956). The consumer wishes to keep a constant flow level of consumption within a period. Each period is normalized to have length 1, so *C* is both the flow level of consumption and the total amount of consumption within a period.

Each unit of consumption must be purchased using transaction services that only banks can provide. For concreteness, think of an economy where all consumption goods are purchased with debit cards, with the bank immediately transferring the appropriate sum from the buyer's account to the seller's account. As a normalization, we assume that one unit of consumption requires one unit of payment services.

For simplicity, we abstract from currency. The only form of "money," which we define as any asset used as the medium of exchange, is bank deposits. Deposits pay a below-public-

[5] We implicitly assume that all contracts in the economy are indexed for inflation, so banks pay a real rather than nominal interest rate on deposits. This assumption simplifies the notation, but creates no substantive change in the analysis.

market interest rate (since $r_t^D < r_t$). But unlike the case with currency, the foregone interest on deposits is not a deadweight loss but rather an implicit payment for transaction services provided by banks. That is, banks use real resources to process transactions but cannot recoup their expenses directly with fees, say, because of regulatory restrictions. This setup is to highlight this study's focus – the relationship between real quantity of services and real balance of the associated assets, and we will further elaborate on this later when we discuss the optimization problem facing banks.

We further assume that each unit of consumption, which has a price of 1, must be paid for using bank deposits. Formally, as in equation (3), the total assets transferred to banks within a period must equal total consumption within that period. This can be interpreted as a "cash"-in-advance (actually, a bank balance-in-advance) constraint. Note that since bank deposits are depleted continuously at a constant rate to pay for a constant flow level of consumption, the average deposit balance ($\bar{D}$) is half the amount of money deposited in each transfer ($2\bar{D}$).

N_t is the number of balance transfers from the household's mutual fund account to its bank account.[6] The reason for a finite number of transfers is that mutual funds face costs of trading—exchanging funds in the stock index into bank deposits—and must pass those costs on to consumers. Each transfer costs the consumer P^N. Exactly as in the Baumol-Tobin model, consumers balance the flow costs of foregone interest against the fixed costs of making transfers between accounts.[7] Rather than spelling out the details of the consumer's optimization problem for securing a constant flow of consumption with transactions costs, we have imposed the

[6] For simplicity, we ignore the integer constraint on N_t, especially since it affects none of our results. Feenstra (1986) shows that the same generalized transaction cost function obtains when the integer constraint is explicitly considered, such as in Barro (1976).

[7] As is usual in this literature, we assume that the cost is independent of the size of the transaction. This is in fact another example where the asset balance bears no defnite relationship to the amount of transaction services.

conditions that come out of that optimization problem. For the steps involved, see Feenstra (1986).

We defer the derivation and discussion of the consumer's optimality conditions until subsection I.F. In the meantime, we present the other actors in the economy, the firms, and study the interactions between firms and consumers in each market.

B. Goods Production

There is a large number of competitive firms producing a homogeneous consumption good with constant-returns, Cobb-Douglas technology. The production function for the representative firm is:

$$Y_t(s) = A_t(s) K_t^G(s)^{\gamma} L_t^G(s)^{1-\gamma}, \tag{4}$$

where $\gamma < 1$ and K^G and L^G are capital and labor used in the production of goods. The production function (4) should be interpreted as giving the instantaneous *flow* of output at each instant, $Y(s)$, given the instantaneous capital and labor inputs $K(s)$ and $L(s)$. If the capital input is constant for the entire period, then Y is also equal to the *volume* of output (recall that a period is one unit of time in length). We introduce the distinction between flow and volume in order to allow banks to lend their deposits as capital to the goods-producing firms. Instantaneous production allows firms to produce goods using inputs over a fraction of a period.

The firm maximizes

$$Y_t(s) - R_t(s) K_t^G(s) - W_t(s) L_t^G(s)$$

subject to (4), where R is the rental rate of capital. Note that the price of goods, like that of consumption, is normalized to one. The first-order conditions are

$$R_t(s) = \gamma A_t(s) K_t^G(s)^{\gamma-1} L_t^G(s)^{1-\gamma}, \tag{5}$$

$$W_t(s) = (1-\gamma) A_t(s) K_t^G(s)^{\gamma} L_t^G(s)^{-\gamma} \,. \tag{6}$$

C. Mutual Funds

Mutual funds are financial intermediaries that manage all household assets other than bank deposits. They are fully equity funded and own the capital stock, which they rent to the goods-producing firms. They receive all capital income on behalf of households, and add that income to households' capital holdings (that is, automatic reinvestment of all dividends and profit). For simplicity, we assume that the receipt of income (including wages) does not incur transactions costs. At a household's direction, the fund manager also periodically transfers some of the household's mutual fund assets to its bank. These transfers *do* incur transactions costs. Mutual funds are also price-takers in both the output and factor markets. Thus, the profit-maximization problem the representative fund faces is

$$\underset{K,L}{\text{Max}} \; \left(R_t(s)-\delta\right) K_t^E(s) - r_t E_{t-1} - r_t(s) M_t(s) + P_t^N N_t - W_t L_t^Z - R_t K_t^Z$$

subject to the constraints:

$$K_t^E(s) = E_{t-1} + M_t(s) \,, \tag{7}$$

$$\dot{M}_t(s) = W_t(s) + R_t(s) K_t(s) + \Pi_t(s) - 1\left(N_t(s)\right) 2\bar{D}_t \,, \tag{8}$$

and

$$N_t = Z_t (K_t^Z)^{\gamma} (L_t^Z)^{1-\gamma} \,. \tag{9}$$

K_t^E is the capital managed by the mutual fund on behalf of households and used by the goods-producing firms. N_t is the number of times household assets are transferred to banks, and P_t^N is the price charged for each transfer. The consumer's mutual fund holdings have a component that is predetermined at the end of the previous period (E_{t-1}) and another component that is the net of income accumulated over the period and transfers of assets out of the mutual fund and to banks.

Income consists of labor and capital income. Transfers are the last term in equation (8). 1(N) is a variable that takes on the value 1 if a transfer takes place at instant *s*, and is zero otherwise. Whenever a transfer takes place, assets amounting to $2\bar{D}_t$ are transferred to the consumer's bank account.[8]

K^Z and L^Z are the capital and labor used in by the mutual fund industry in producing the fund transfer services. Note that we have assumed the same Cobb-Douglas shares for capital and labor as in the goods-producing sector. This assumption simplifies analysis – making relative prices a function only of relative technology – without altering the results qualitatively.

Note that here we have placed capital depreciation, at rate δ, directly in the objective function. Since we have made production a flow concept, we do the same for depreciation. We assume that the depreciation rate depends on the fraction of the period over which capital is used. Capital used in production the full length of the period loses δ of its value.

The first-order conditions are

$$r_t(s) = R_t(s) - \delta\,, \tag{10}$$

and

$$P_t^N = \frac{R_t^{\gamma} W_t^{1-\gamma}}{Z_t}\,. \tag{11}$$

There are also first-order conditions for efficient use of capital and labor, paralleling equations (5) and (6), above.

Equation (10) should be seen as an equilibrium condition rather than a necessary condition for an individual mutual fund to maximize profit. Competition dictates that the owners of capital—the households—receive all of the net marginal product of the capital they are supplying.

[8] Thus, the income terms in the consumer's budget constraint (2) (e.g., *W*) should be interpreted as including the

Equation (11) is important in what follows; we discuss its implications later at length. For now, we note that P^N is not necessarily constant, and in general will change if any of *R*, *W* or *Z* changes.

D. Banks

Banks receive deposits and provide payment services.[9] Banks lend out their deposits in the form of capital to the goods-producing firms, thereby earning the same rate of return on their assets—the net return on capital *r*—as mutual funds do on behalf of the households.

This is where the assumptions of continuous flow production and depreciation are necessary. If the productive capital stock usable at time *t* were pre-determined as of time *t-1*, then banks would be unable to use their deposits productively to make loans. This assumption would make bank deposits equivalent to currency (both would pay zero interest), and imply that banks earn no asset income. Since these features are inconsistent with modern banking, we use the device of continuous production to afford banks a productive use of their deposits.

Before we state the bank's problem as we actually model it, we show that the fundamental economics of the banking sector are essentially the same as those of the mutual fund sector discussed in the previous sub-section. Letting *S* represent payment services, then a bank that could charge explicit fees for its services and pay a competitive market return on its deposits would solve the following problem:[10]

$$\underset{\bar{D},L}{\text{Max}} \ \left(R_t - \delta\right)\bar{D}_t - r_t^D \bar{D}_t + P_t^S S_t - W_t L_t^B - R_t K_t^B$$

---

interest payments

[9] For brevity, we call these institutions banks, but it is equally valid to interpret them as money market mutual funds, or any other financial intermediaries that provide payment services while holding a balance of customers' asset.

subject to

$$S_t = B_t (K_t^B)^{\gamma} (L_t^B)^{1-\gamma}.$$

Rather than integrating over the flow profits and costs of a time-varying level of deposits, *D*, we take the harmless shortcut of expressing the bank's problem in terms of the average level of deposits $\bar{D}_t$. This shortcut makes no difference because we set up the model with flow output and flow depreciation. Optimizing over $\bar{D}_t$ is sensible for an individual bank since it has a choice of how many deposits to accept (and how many services to provide).

With this set of institutions, competition in the banking system would ensure that r^D equalled $R - \delta$, just as in the mutual fund sector, and fees would equal the marginal cost of producing transaction services. Note that in this case households would choose to put all their assets in the banking sector, since by so doing they would avoid the transfer charges from the mutual funds and yet obtain the same rate of return on their assets.

In this case, measurement of bank output would be easy. It is clear that nominal output is $P^S S$ and real output is *S*, while $r^D D$ is a transfer of asset income to consumers. Note that the way to measure output would be to count the number of transactions processed, *S*, since there would not necessarily be any constant relationship between *S* and *D*. In this equilibrium, we would have $S = C$, as usual, but we would also have $D = K$, and there is no reason why the consumption/capital ratio should be constant outside the steady state.

However, suppose that actual institutions do not allow the sensible outcome above. Specifically, suppose banks are not allowed to pay depositors the full return on funds (i.e., r_t), perhaps because regulations prohibit banks from paying interest on certain types of accounts

---

[10] For economy of notation, we simply use $\bar{D}_t$ – the representative household's average deposit balance – to denote the representative bank's average balance as well. It should be understood that a single bank's average balance can

(such as Regulation Q) or from investing in corporate equities, which regulators might deem too risky. Consequently, banks have to remunerate depositors in kind, by providing services (*S*) "for free." Thus, the two parties essentially strike a barter agreement.

We incorporate this real-world feature into our model, and assume that banks charge for services implicitly by offering a lower rate of return on deposits. For simplicity, we study the extreme case where banks levy no fees at all but instead get compensated for all productive services via the interest rate spread.[11] Under this assumption, the representative bank solves the following optimization problem:[12]

$$\left(R_t - \delta\right)\bar{D}_t - W_t L_t^B - R_t K_t^B - r_t^D \bar{D}_t$$

subject to

$$S_t = C_t = 2N_t \bar{D}_t$$

and

$$S_t = B_t (K_t^B)^{\gamma} (L_t^B)^{1-\gamma}$$

From the point of view of individual, atomistic banks, *N* is an exogenous variable that they take as given. Solving the two constraints together for $\bar{D}_t$ as a function of exogenous variables and maximizing with respect to $\bar{D}_t$ gives:

$$r_t^D = r_t - \frac{2N_t R_t^{\gamma} W_t^{1-\gamma}}{B_t}, \tag{12}$$

where we have inserted the equilibrium condition $r = R - \delta$. As with the mutual funds, equation (12) should be interpreted as an equilibrium condition ensuring that banks make zero profits.

be any multipleof $\bar{D}_t$; it is in fact indeterminate because of the constant-returns-to-scale banking technology.

[11] We do not model the regulation explicitly, but assume that the equilibrium level of r^D is always below the interest rate ceiling.

[12] Again, we have set up the problem using the simplifying treatment that banks have a constant level of deposits $\bar{D}_t$ throughout the period.

Note that the interest rate paid on deposits is declining in the number of transfers to the bank, *N*. The reason is that, for a given level of consumption, a larger number of transfers implies a lower average deposit balance, and thus lower bank income. But if the level of consumption is fixed, then so are bank expenses. On net, therefore, the bank has lower net income to distribute back to depositors in the form of interest payments.

E. The Optimal Choice of N

Now we solve for N^*, the optimal number of transfers between a mutual fund and a bank account. (For simplicity we drop time subscripts in the derivation, but the result holds at all points in time.) As in the standard Baumol-Tobin model, the optimal number of trips minimizes the total cost of making trips to the bank plus foregone interest on bank deposit balances:

$$\underset{N}{\text{Min}} \;\; P^N N + \frac{\left(r - r^D\right)C}{2N} \tag{13}$$

The first-order condition gives the famous Baumol (1952) square-root rule:

$$N^* = \sqrt{\frac{\left(r - r^D\right)C}{2P^N}}$$

Substituting in the equilibrium conditions for $r - r^D$ from (12) and P^N from (11) gives:

$$N^* = \sqrt{\frac{\frac{2NR_t^{\gamma}W_t^{1-\gamma}}{B}C}{2\frac{R_t^{\gamma}W_t^{1-\gamma}}{Z}}} \Rightarrow N^* = \frac{Z}{B}C\,. \tag{14}$$

Thus, the optimal number of mutual fund transfers demanded by the consumer will depend linearly on both consumption (*C*) and relative technology—the technology for making transfers relative to that for payments (i.e., *Z/B*). This last result is a deviation from the usual square-root formula because of the unusual feature of our setup: In equilibrium, $r - r^D$ depends negatively on N, via the dependence of the average deposit balance on N.

F. Equilibrium and Steady State

We begin by solving the consumer's optimization problem, using some of the results derived in the previous subsections. Substituting the optimal choices for P^N, $r - r^D$ and N from equations (11), (12) and (14) as well as the "deposit in advance" constraint (3) into the budget constraint (2) yields

$$C_t\left(1+\frac{2R_t^{\gamma}W_t^{1-\gamma}}{B}\right)+E_t=W_t+\Pi_t+\left(1+r_t\right)E_{t-1}. \tag{15}$$

This simpler budget constraint says that from the consumer's point of view, the need for financial services to purchase consumption goods is like a tax on consumption. Thus, the effective price of consumption is not just the price of output, which is normalized to 1, but also includes the prices of financial services, which are a function of the financial sectors' technology and factor prices. It is easy to confirm that in equilibrium the expenditure on mutual funds' transfer services equals the expenditure on banks' payment services, explaining why the "tax" on each unit of consumption equals *twice* the price of an unit of banking services.

Using the simplified budget constraint, we now solve for the intra- and intertemporal first-order conditions characterizing optimal consumption behavior:

$$U'\left(C_t\right)=\lambda_t\left(1+\frac{2R_t^{\gamma}W_t^{1-\gamma}}{B}\right), \text{ and} \tag{16}$$

$$\beta\left(1+r_t\right)\lambda_t=\lambda_{t-1}. \tag{17}$$

λ has the usual interpretation; it is the marginal utility of wealth. However, the marginal utility of consumption exceeds λ because consumption also entails expenditures on financial services.

Market-clearing conditions for capital and labor, respectively, are as follows:

$$L_t^G + L_t^B + L_t^Z = 1, \tag{18}$$

$$\text{and } K_t^G(s) + K_t^Z + K_t^B \equiv K_t = K_{t-1}^E + D_t(s) = E_{t-1} + D_t(s). \tag{19}$$

Note that $D_t(s)$ here is the actual level of deposits at each instant *s* in period *t*, as opposed to the average level of deposits, $\bar{D}_t$, which is constant within a period. So the capital stock used in goods production—$K_t^G(s)$—varies from moment to moment as well.[13]

Perfect competition in goods and factor markets ensures zero profit:

$$\Pi_t = 0, \quad \forall t.$$

The real interest rate is[14]

$$r_t(s) = \gamma A_t(s) K_t^G(s)^{\gamma-1} L_t^G(s)^{1-\gamma} - \delta. \tag{20}$$

Capital follows the law of motion:

$$K_t = I_{t-1} + (1-\delta) K_{t-1}. \tag{21}$$

Having described the equilibrium, we now characterize the steady state. Assume (for now) that all technologies are constant. Under this assumption (and without population growth), the

[13] We assume that K^B and K^N are constant within a period, and the time-varying capital stock is only in goods production. The constancy of bank capital is sensible, since because consumption, and hence the production of bank services, is a constant flow. We assume that mutual funds spend the time interval between actual transfers doing the paperwork necessary for the next transfer, and this leads to the continuous level of production that is consistent with a constant capital stock.
[14] Note that in general *r* will vary even within a period. All references to r_t thus pertain to the *average* level of *r* within period t.

steady state of this model implies zero growth in all aggregate and per-capita variables.[15] We now solve for the equations characterizing the steady state. In what follows, we use variables without time subscripts to denote steady-state values.

Inserting the definition of the real interest rate (20) into the Euler equation (17) gives:

$$\gamma A K^{G\ \gamma-1} L^{G\ 1-\gamma} - \delta = \frac{1}{\beta} - 1 \Rightarrow \frac{K^G}{L^G} = k\left(A, \gamma, \delta, \beta\right). \tag{22}$$

Note that this equation pins down the capital-labor ratio in goods production as a function of technology (A, γ, δ) and preferences (β).

One may object to the claim that the capital stock in goods production is constant by arguing that the level of bank deposits varies over a period, even in the steady state, and deposits are lent out to goods-producing firms as part of their capital stock. However, the level of capital in mutual funds is also time-varying, in exactly the inverse fashion. In the steady state (although not in general) goods production just equals consumption and depreciation at every instant. The total capital stock, which equals the sum of the assets of the banking system and the mutual funds, is thus constant over time.

Since all sectors have the same capital and labor shares, they must also have the same capital-labor ratio:

$$\frac{K^G}{L^G} = \frac{K^B}{L^B} = \frac{K^Z}{L^Z} = k\left(A, \gamma, \delta, \beta\right) \tag{23}$$

Goods output must be used for consumption or investment. By (21) steady-state investment just replaces depreciation, so we have that

$$C + \delta K = A K^{G\ \gamma} L^{G\ 1-\gamma}. \tag{24}$$

[15] The logic can be seen particularly easily from the fact that the marginal product of capital is diminishing – given a constant technology, with enough capital the marginal product of capital net of depreciation would become negative,

From the results in the previous sections, we also know that

$$S = C = BK^{B\ \gamma} L^{B\ 1-\gamma} \tag{25}$$

and

$$N = C\frac{Z}{B} = ZK^{Z\ \gamma} L^{Z\ 1-\gamma}. \tag{26}$$

Equations (18), (19) and (23)–(26) comprise eight equations (expression (23) is actually three equations) that determine eight endogenous variables: C, K, K^G, L^G, K^Z, L^Z, K^B, and L^B. (Equations (25) and (26) also determine S and N.)

II. "Money" in the Utility Function and the Technology for Producing Financial Services

In this section, we show first that, under one strong assumption, our model is able to justify rigorously a setup where consumers are assumed to derive direct utility benefits from bank deposits. That is, Feenstra's (1986) celebrated "functional equivalence" theorem is nested as one restricted version of our model. The restriction is best interpreted, in this environment without uncertainty, as precluding any changes in the technologies for making transactions.

A. A Rigorous Foundation for "Money" in the Utility Function

Suppose, for this sub-section only, that P_t^N is constant, i.e., $P_t^N = k$ for all t. Consider the original optimization problem for the consumer, with this assumption and with constraint (3) substituted into the budget constraint (2). The household's problem becomes one of maximizing

$$\sum\nolimits_{t=0}^{\infty} \beta^t U\left(C_t\right), \tag{27}$$

subject to:

---

which surely cannot be a choice that optimizing households would make.

$$C_t + k\frac{C_t}{2\bar{D}_t} + E_t + \bar{D}_t = W_t + \Pi_t + \left[\left(1+r_t\right)E_{t-1} + \left(1+r_t^D\right)\bar{D}_{t-1}\right]. \qquad (28)$$

Now define a variable X as

$$X_t \equiv C_t + k\frac{C_t}{2\bar{D}_t} \equiv C_t + \phi(C_t, \bar{D}_t)\,. \qquad (29)$$

In Feenstra (1986), as in earlier transaction cost models, $\phi(C_t, \bar{D}_t)$ is referred to as a "liquidity cost"—the cost consumers incur because they must first exchange their wealth into a "liquid" asset, which pays a below-public-market rate, and use it to procure consumption. Feenstra (1986) thus interprets X as "gross consumption," i.e., all expenditures related to consuming, which in that model equal the sum of expenditures on consumption proper plus those on "liquidity services."

Note the qualitative distinction between this so-called liquidity cost $\phi(.,.)$ and the actual bank services in our model. Here, $\phi(C_t, \bar{D}_t)$ is in fact a household's real expenditures on mutual fund transfer services, while real expenditures on bank services are $\left(r_t - r_t^D\right)\bar{D}_t$. Since X as defined in (29) is just the sum of actual consumption and expenditures on mutual fund services, it does not include expenditures on bank services and thus is not a complete measure of gross consumption in this model.

Now consider an alternative problem for the consumer, which is to maximize

$$\sum_{t=0}^{\infty} \beta^t V\left(X_t, \bar{D}_t\right), \qquad (30)$$

subject to:

$$X_t + E_t + \bar{D}_t = W_t + \Pi_t + \left[\left(1+r_t\right)E_{t-1} + \left(1+r_t^D\right)\bar{D}_{t-1}\right]. \qquad (31)$$

Constraints (31) and (28) are clearly equivalent. So the original problem (27)-(28) becomes equivalent to the second problem (30)-(31), that is, they differ only in their functional notations, if a function *V* exists such that

$$U\left(C_t\right) \equiv V\left(X_t, \bar{D}_t\right).$$

Note that real expenditures on bank services in this model are indirectly captured by the presence of $\bar{D}_t$ in the alternative utility function *V*.

Proposition 1: If P^N is constant, then problems (27)-(28) and (30)-(31) are equivalent.

Proof: The original problem (27)-(28) satisfies Assumption 1 of Feenstra (1986). Thus, the proof follows directly from Proposition 1 of Feenstra (1986). Furthermore, the function *V* so defined has a variety of useful and intuitive properties, which follow from Assumption 2 of Feenstra (1986).

Notice that the derivation of the reformulated problem (30)-(31) justifies rigorously a representation of consumers' problem that has real bank deposit balances ("money balances") in their utility functions. Money-in-the-utility-function is exactly the approach used in Barnett (1978, 1980). Those studies are commonly cited as establishing the theoretical foundation for the "user-cost-of-money" approach to the measurement of financial service output, whose quantity is measured using an index of the real balances of assets and whose price thus corresponds to the relevant interest rate spreads. See, for example, Hancock (1985) and Fixler and Zeischang (1992). Nominal output of the financial services then simply equals the product of the real balance and corresponding interest rate spread (e.g., foregone interest for depositors).

In the context of our model, since we have assumed a fixed proportionality between the real *value* of consumption and the use of bank services, and since the value of consumption is also in fixed proportion to deposit balances, one can construct an index of bank services proportional to the average deposit balance $\bar{D}$ over time. The price index for bank services will then be proportional to the rate spread $r - r^D$, exactly as in Fixler (2006).[16]

While one would obtain equivalent measures of bank services using either an index based on actual counting of S or an index based on real balance $\bar{D}$, it is clearly much easier to simply obtain deposit balances and interest rate spreads. Thus, if the procedure of Fixler (2006) can be applied generally, it is preferable on practical grounds to the procedures suggested by Wang, Basu and Fernald (WBF, 2004).

B. "Money" in the Utility Function with Time-Varying Financial Service Technology

Now we drop the assumption of a constant price P^N for mutual fund transfers. After all, in Section I we showed that P^N depends on the technology in the mutual fund sector and on economy-wide real wages; see equation (11). There is no economic reason why either must be constant over time. Feenstra (1986), however, does not consider possible changes in transactions technologies.

Thus, we return to the original household problem of maximizing

$$\sum_{t=0}^{\infty} \beta^t U\left(C_t\right)$$

subject to:

[16] It is easy to identify "the" interest rate spread in our non-stochastic environment, where there is only one non-bank interest rate. In a stochastic environment where there are many interest rates, Wang (2003a) and Wang et al. (2004) show that the reference rate (here, r) needs to be corrected for risk.

$$C_t + P_t^N \frac{C_t}{2\bar{D}_t} + E_t + \bar{D}_t = W_t + \Pi_t + \left[(1+r_t) E_{t-1} + \left(1+r_t^D\right) \bar{D}_{t-1} \right].$$

Now it is apparent that one can no longer write "liquidity costs," in Feenstra's language, as a function of consumption and real balances alone—those costs also depend on P^N. But then one can no longer use the elegant approach of defining "gross consumption" X as a function of C and $\bar{D}$ alone, and using that to obtain a time-invariant equivalent utility function, V.

Suppose we define a different problem:

$$\sum_{t=0}^{\infty} \beta^t V\left(X_t, \bar{D}_t, P_t^N\right), \tag{32}$$

subject to:

$$X_t + E_t + \bar{D}_t = W_t + \Pi_t + \left[(1+r_t) E_{t-1} + \left(1+r_t^D\right) \bar{D}_{t-1} \right]. \tag{33}$$

We conjecture—but have not yet proven—that the original problem (1)-(3) is equivalent to the reformulated problem, (32)-(33).

Nonetheless, we can show the consequences for measuring bank output if P^N varies over time. Suppose that the economy was initially in the steady state. Now suppose that Z, mutual funds' technology for producing asset transfer services, doubles once and for all. It is easily shown that, in the new steady state, the price for fund transfers (P^N) halves and the optimal number of transfers doubles, while optimal consumption remains the same.[17]

The proof follows from inspecting equations (24)–(26) in conjunction with (14). Suppose that total capital, K, consumption, C, and the distribution of capital and labor across sectors remain unchanged even as Z doubles. Clearly, equations (24)–(25) are satisfied, since no exogenous variables have changed in those equations. With unchanged inputs of K and L in the mutual fund sector but with technology that is twice as good, the output of that sector (N) must double as well.   But this is exactly what is implied by the optimal demand for transfers, (14). Therefore, the initial conjecture of unchanged K and C in the new steady state is verified. It then implies that, by equation (3), the average deposit balance, $\bar{D}$, will be halved.

Now consider the proposal to construct an index of bank service output by equating the growth rate of services to the growth rate of real deposit balances. Since C remains the same, bank services, S, have to remain constant as well. Measuring bank service output directly, by counting the number of transactions, would reveal this fact. But constructing an index of real output as proportional to the real balance of deposits would show—incorrectly—that real service output has been cut in half.

Our analysis also points out that, even in our deliberately simple setup, changes in the technology of the mutual fund sector are only one reason why the MIUF representation is not robust. In fact, it is clear from equation (14) that, even in partial-equilibrium, only proportional changes in technologies of both mutual funds and banks (i.e., Z_t and B_t respectively) will leave N^* and in turn $\bar{D}$ unchanged. Otherwise, changes in either Z_t or B_t alone have symmetric but opposite effects on N^* and in turn $\bar{D}$. Furthermore, it is easy to see that changes in P^N due to other exogenous factors (with regard to the model), such as changing relative factor prices, or even changing capital market regulations, can break down the equivalence as well. These are quite realistic conditions. We will elaborate on them later when we discuss the implications for measurement, and so just note here that an equivalent utility function over real balances is most likely a rare coincidence in real-world situations.

[17] In this model, the general-equilibrium analysis of the new steady state gives the same results as analyzing the

III. Implications for Measurement

We have shown a realistic case—changes in a transaction technology—where the household's optimization problem does not have a functionally equivalent alternative expression with real asset balances as an argument. (Note, incidentally, that the technological change took place not even in the banking sector itself, which is the sector that actually performs the transactions needed for households to consume, but rather in a sector that only transfers assets to banks.) Our example, however, demonstrates but one mechanism, albeit arguably the most relevant, through which functional equivalence breaks down.

One other important and likely reason is that other financial instruments become better substitutes for the current medium of exchange because of lower trading costs. The cost of trading market securities can fall either because the technologies for order execution, clearing and settlement improve, or because better capital market regulations mitigate the costs stemming from asymmetric information problems. Time-varying inflation is another reason; in fact, as noted by Feenstra (1986) himself, only by treating inflation as a tax levied at the end of each period does one obtain a stable transaction cost function from the Baumol-Tobin model. Otherwise, transaction costs will vary with the inflation rate. Finally, even the small modeling change of making the mutual fund and banking sectors have different capital shares of production would mean that the service-to-deposit ratio would change over time as factor prices change, not just if either technology changes.

The general conclusion is that there exists no fundamental theory stipulating any definite relationship between the quantity of service output and real balance of financial assets. Any mapping between service flow and asset balance depends entirely on features of the transaction technology and its relation to other technologies in each specific model.

---

change in P^N and N in partial equilibrium, holding consumption constant.

Not only can one easily write down models where transaction services are wholly separate from any financial assets, but real-world examples of such services abound. Fischer (1983) makes exactly the same argument "…[it is] possible to imagine institutions that make transactions without requiring any corresponding asset holding. The postal giro system is the most important example. A company making C.O.D. deliveries is another" (p. 6). More cases emerge nowadays, such as web portals (such as www.mvelopes.com) that make payments to any accounts (bank deposit, mutual fund, utilities, etc.) a customer designates, as well as utilities companies that offer their own online payment option.

The bottom line, therefore, is that using real balance of financial assets to measure financial service output is most likely a reduced-form approximation at best. It should be utilized as a last resort given the data availability and resource constraint a user faces. It should be regarded as the exception but not the norm.

The alternative that we have proposed, e.g., in WBF (2004), is in fact a return to traditional practice, at least in the BLS productivity group (BLS, 1998). It is to construct indexes of real service output based on counts of actual transactions—for example, checks cleared, ATM transactions processed, and mortgage applications screened. A major problem is choosing the correct weights to aggregate these index components. The BLS uses (rather dated) Functional Cost Analysis data to weight each output component by its share in total labor cost. Basu, Inklaar and Wang (2006) suggest and implement several theoretically preferred alternatives, but admit that in some cases we simply need more and better data.

One important conceptual advantage of the BLS-type method is that it yields consistent measures of both implicitly- and explicitly-priced financial services. While our focus has been to derive the theory for decomposing nominal output of implicit service output, exactly the same logic applies to decomposing explicit fees into a price and a quantity component. In fact, our output measure makes even more intuitive sense when applied to services that generate explicit revenue, which are more likely to be separate from any asset holding. With or without an associated asset balance, our method calls for measuring financial services just as we would any other service: clearly define each type of transaction, and obtain a quality-adjusted quantity index of the transactions (e.g., the number of conforming residential mortgage loans screened). The combination of nominal and real output then implies a price deflator.

By comparison, when one follows the current implementation of the user-cost approach and measures output of implicit services using real balance of the associated financial instruments, one must also measure explicit services on an equal footing, if one is to obtain consistent aggregates. However, in the case of services compensated with explicit fees and not attached to any financial assets, it makes little sense to use some interest rate differential (which one?) as the implicit price deflator, and derive a quantity index that is on par with the real balance of some imaginary financial assets.

In this era of rapid technological progress and proliferation of new financial instruments, including innovative combinations of financial and service features, the greater conceptual consistency afforded by our output measure seems particularly desirable. That is, no matter how the composition of explicitly versus implicitly charged services changes both over time and across financial institutions (or even within an institution), our measure of real output should, in theory at least, generate consistent aggregates.

Take the commercial banking industry, for example, where such compositional changes abound. Ever since interest rate ceilings were removed for most types of deposit accounts, banks have been broadening the range of retail transactions that carry an explicit fee schedule, while raising the interest rates paid on deposits. Most banks also offer depositors a choice between paying a per-transaction fee or maintaining a higher balance. Such changes may well have affected the ratio between real deposit balances and the amount of transaction services both across banks and over time. Similarly, on the lending side, many banks now charge a fee for loan applications and pre-approvals, whether or not a loan is actually granted later. In addition, large banks increasingly engage in (aptly-named) off-balance-sheet activities, such as underwriting derivatives contracts, that generate fees but produce no corresponding assets or liabilities on the balance sheet.[18]

Such developments have created difficulties for studying bank production technology because existing measures of output cannot generate a consistent aggregate for any bank engaged in these diverse activities. For example, Rogers (1998) measures the output of traditional bank lending using the balance-sheet value of loans, but measures off-balance-sheet activities using their explicit revenue. In contrast, the real-service-flow measure of output advocated here should in theory yield an output index that is comparable both cross banks and over time.

Our results point to the need to reexamine the findings of a large literature that analyzes the properties of banks' production technology. That literature features three approaches, which differ only in what each defines as bank output.[19] Once a type of activity is taken to be output, all three approaches invariably measure it using the deflated book value of the corresponding

[18] Many OBS derivatives are defined based on a so-called "notional value," which is reported in a special section of banks' regulatory filings. However, the notional value of a derivative contract (such as a swap or a forward) bears no definite relationship to even the actual financial worth of the contract, let alone the amount of financial services rendered.

[19] Specifically, the three approaches are distinguished 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. The value-added approach views every financial product whose creation requires labor and capital as inputs, and it thus records deposits as an output. The user-cost approach, which is also the foundation for the NIPAs' measure, classifies input and output endogenously: given a reference rate, financial assets (liabilities) whose realized rates of return are greater (less) than the reference rate is defined to be output, and others as input. So transaction deposits are typically found to be outputs in data. See Berger and Humphrey (1997) for a survey.

financial assets or liabilities. Given the substantial changes in the scope and mode of operation in banking organizations, such book-value-based output measures may have led to estimates of banking technology parameters that are biased and the resulting policy implications that are in turn flawed.

IV. Conclusion

This paper demonstrates that unrealistically restrictive conditions are needed to obtain a fixed relationship between the quantity of a financial service and the volume of its associated financial instrument. It implies that a quantity index proportional to the real balance of financial assets is unlikely to be a robust proxy for the true real output of actual financial services. This conclusion is general, even though features of the transaction technology in the model are, for the sake of intuition, chosen to resemble payment services provided by real-world banks to their depositors.

The focus of this study is the measurement of real output of financial services and the corresponding price deflators. It is a natural continuation of our earlier work developing a new measure of the nominal output of financial services. In those studies (particularly Wang, 2003a; Wang, Basu and Fernald, 2004), we argue that the user-cost approach can serve as the theoretical basis for measuring nominal output of implicitly priced financial services once it is extended to take account of the (systematic) risk in the associated financial instruments. That is, there is no single reference rate, and each specific rate depends on the risk of the relevant financial securities.

In this study, we emphasize that, independent of how to deal with risk in the measurement of nominal output, real financial services are unlikely to be demanded in fixed proportion to the real balance of any specific category of financial instruments. In fact, this argument can be made most forcibly in the case where there is no risk! Without risk, all financial assets must offer an identical rate of return, and so the quantity of any individual class of assets is indeterminate, unless additional constraints are imposed. In contrast, the quantity of each type of transaction services is pinned down by its production technology.[20]

Therefore, indexes that directly measure real services generated by financial institutions are robust to the types of changes in technology and institutions that we observe. Thus, even though these indexes are typically more difficult to construct—especially since the weights of the different components are hard to measure when there are no explicit prices—they are strongly preferred conceptually. Furthermore, an index based on direct measures of service output yields an aggregate measure of financial services that is conceptually meaningful, whether the services are implicitly or explicitly priced. This seems a particular advantage in an era of rapid innovation and increasing diversity in financial institutions' modes of operation.

The measurement community has dealt successfully with many challenging tasks—for example, constructing quality-adjusted price indexes for durables and for medical services. Now that the conceptual foundations for measuring real and nominal financial sector output are falling into place, we are confident that patient, persistent effort on both the theoretical and empirical fronts will soon bear fruit in this area as well.

[20] Fischer (1983) makes a similar argument in a model of transaction services with perfect foresight.

References

Barro. Robert J. (1976). “Integral constraints and aggregation in an inventory model of money demand,” *Journal of Finance* 31.77-87.

Barnett, W. A. (1978). “The User Cost of Money,” *Economic Letters* 1(2), p. 145-49.

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

Baumol. William J. (1952). “The transactions demand for cash: An inventory theoretic approach,” *Quarterly Journal of Economics* 66, 545-556.

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

Berger, A. N. and D. B. Humphrey (1997). “Efficiency of Financial Institutions: International Survey and Directions for Future Research,” *European Journal of Operational Research* 98(2), p. 175-212.

Feenstra, Robert C. (1986). “Functional Equivalence between Liquidity Costs and the Utility of Money,” *Journal of Monetary Economics*, v. 17, iss. 2, pp. 271-91.

Fischer, Stanley (1983). “A Framework for Monetary and Banking Analysis,” *Economic Journal*, Supplement, v. 0, iss. 0, pp. 1-16.

Fixler, D. J. and K. D. Zieschang (1992). “User Costs, Shadow Prices, and the Real Output of Banks,” in Z. Griliches (eds.), *Studies in Income and Wealth*, vol. 56, NBER.

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.

Fixler, D. J. (2006). “Valuing Implicit Financial Services and the Impact on Commercial Bank Activity,” 2004 Index Number Conference Volume, *forthcoming*.

Hancock, D. (1985). “The Financial Firm: Production with Monetary and Nonmonetary Goods,” *Journal of Political Economy* 93(5), p. 859-80.

Rogers, K. E. (1998). “Nontraditional Activities and the Efficiency of US Commercial Banks,” *Journal of Banking and finance* 22(4), p. 467-82.

Tobin, James, 1956, The interest elasticity of transactions demand for cash, Review of Economics and Statistics 38, 241-247.

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 Series*, # 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 Series,* # 03-6.
Available at: http://www.bos.frb.org/economic/wp/wp2003/wp036.htm

Wang, J. C., S. Basu, and J. G. Fernald (2004). "A General-Equilibrium Asset-Pricing Approach to the Measurement of Nominal and Real Bank Output." Presented at *Price Index Concepts and Measurement*, Conference on Research on Income and Wealth (CRIW), June 2004.
Also available at http://www.bos.frb.org/economic/wp/wp2004/wp047.htm