MEASURING THE OUTPUT OF BANKS:

WHAT DO BANKS DO?

by

JACK E. TRIPLETT
U. S. Bureau of Economic Analysis

Background Paper for

VOORBURG GROUP ON SERVICES STATISTICS

Williamsburg, Virginia, U.S.A.

October 19-23, 1992

NOTE: This paper was originally presented at the Western Economic Association meetings in July 1991. It has also been presented in other seminars, including the National Bureau of Economic Research and Statistics Canada. A revision is underway, so comments are both appreciated and solicited.

Measuring the Output of Banks: What Do Banks Do?

by Jack E. Triplett

Chief Economist,
U.S. Bureau of Economic Analysis
Washington, DC 20212

Prepared for the session:
"What Do Banks Do?
Bank Output and Productivity"

Western Economic Association,
Seattle, Washington
July, 1991

BEA
June, 1991

Measuring the Output of Banks: What Do Banks Do?

By Jack E. Triplett*

Two unresolved questions dominate research on measuring the economic activity of banks:

(1) What are the outputs?

(2) What are the inputs?

In few other topics in economic measurement are the issues so primitive. In this paper, I review existing empirical measures of bank output, and present a measurement model that is intended to move the discussion beyond the first round of primitive questions in order that we may begin to engage other--still primitive--questions, such as: What are the units, and what is the price?

I. What Do Banks Do? Existing Empirical Approaches

Explicit measures of banking output appear in at least two different literatures in economics. The oldest is the measure of banking output contained in the national accounts of most countries (including the United States). More recently, several banking output measures have appeared in research on bank regulatory issues.

A. Traditional National Accounts Approach

In the national accounts literature, the topic of banking output is usually characterized as "the banking imputation problem"--which suggests that the measure incorporated into national accounts is not regarded as wholly satisfactory. The imputation "problem" arises largely as a consequence of the

national accounts treatment of interest flows.

> "...[When] an enterprise receives capital from the public and transfers it to other enterprises...it is reasonable to follow the general rule that the income and product originating in a firm shall depend on its net interest payments, that is, interest paid minus interest received. Such a rule produces negative value added for banks" (Bailey, 1971, emphasis supplied).

One can debate whether this "general rule" is in fact as reasonable for financial firms as it has historically been accepted to be. The rule excludes the major source of bank revenue (income from lending activity) from the measure of banking output. Gorman (1969) colorfully remarks that the national accounts treatment of interest flows--unless adjusted--leaves the "commercial bank...portrayed as a leech on the income stream."

To avoid the "leech on the income stream" portrayal, banks are assumed to provide unpriced or "free" services to depositors (such as checking accounts for which no explicit changes are made) whose value is equal the entire net proceeds from banks' lending operations.[1] Thus, the national accounts imputation for the value of unpriced depositor services provided by banks is:

$$\Sigma f_u S_u = - (\Sigma i_j D_j - \Sigma r_i L_i)$$

where f_u and S_u are the implicit fee and implicit quantity of unpriced service u, D_j designates a deposit with interest cost i_j, and L_i is a loan (or other interest-earning asset) with return r_i. The total output of the banking industry includes the imputed value of free services, so defined, plus the value of services for which an explicit charge is levied (certified checks and so forth--a very small part of bank revenue), or:

(A) Bank output = $\Sigma f_p S_p + [-(\Sigma i_j D_j - \Sigma r_i L_i)]$

where the subscript p designates explicitly priced services.

The national accounts approach was first introduced by Yntema (1947), and has been employed in the U.S. National Income and Product Accounts (NIPA) since 1947. It was accepted by Stone (1947), and incorporated, with a small modification that will be ignored here, into the United Nations' System of National Accounts, or SNA (United Nations, 1968).[2] See also Jaszi (1958), and Goldberg (1984).

Finding an economic rationale for the national accounts treatment of banking has always been a problem. One attempt relies on the liquidity of bank deposits, relative to alternative investments:

> "A commercial bank in effect converts illiquid assets into liquid, readily transferable ones.... In exchange for its services in connection with deposits, and for their comparative liquidity, it receives these deposits at low interest compared with what the depositer could earn on alternative, less liquid assets" (Bailey, 1971).

Similar language appears elsewhere in the national accounts literature.

Liquidity clearly explains why bank deposits earn relatively low interest; but in the national accounts model, depositors receive the <u>entire</u> proceeds of banking activity, and banking proceeds are earned from investments in relatively illiquid assets. The national accounts banking model implies that depositers, had they not opted for "free" checking accounts, <u>could</u> have earned what the banker earns on loans (rather than earning, say, the actual interest on some alternative, but still liquid, bank deposit that does not provide "free" services). Depositors own liquid assets, in this model,

without paying a liquidity premium. The liquidity of bank deposits cannot explain the national accounts banking model.

A more appealing rationale characterizes the national accounts approach to banking as an "agent" model: Banks exist to act as agents for their depositors, and for this reason pay out their entire earnings to depositors.[3] In the national accounts model of banking, all commercial banks are assumed to behave as if they were true mutual savings banks, or mutual insurance companies. Since the agent rationale is the only one that logically explains the national accounts banking imputation, it is the one incorporated into the summary table that accompanies this review (Figure 1).

Logical consistency, however, does not mean that the national accounts model of bank behavior is empirically valid. If banks really acted as depositors' agents and paid out their entire earnings to depositors, why should so much of the proceeds of banking be paid out in unpriced services? Wouldn't deposit interest be greater?[4]

Leaving aside questions about the motivations of bankers that are implied by it, there other problems with the national accounts banking output measure. Banks' production of uncharged depositor services ought indeed to be one component of banking output, but why should depositor services (and explicit fees) be the only components? What about bank services to borrowers? The current revision of the SNA sought to remedy this omission by spreading the banking imputation around among both depositors and borrowers, in proportion to their deposits and loans (Figure 1), but this proposal leads to other anomalies. Borrowers would be deemed to receive "free" services from banks (record-keeping and so forth) when they contract for loans; but the finance the bank provides them (and for which they pay interest) is not,

because of the national accounts view of interest, an economic service provided by banks. Do borrowers negotiate loans to obtain record-keeping services, rather than to obtain finance? Why should the provision of finance to borrowers not be an output of the banking business? If finance does not provide productive services, as the national accounts model of banking assumes, why should banks earn revenue from loans? Why, in other words, should banks' sources of revenue not be as good an indicator of what they produce and sell as are the revenue sources of a coal mine or a laundry?

These criticisms are quite old. The parallel between banking revenues and those of coal mines and laundries, for example, appears in Warburton (1947). He noted that "...national income estimators...treat [banks'] interest received, which constitutes the bulk of their sales receipts, as negative expenses," and argued that a better approach would recognize that "the market value of the services of the banking industry, as expressed in its sales receipts, is as clear cut as in any other industry..." (Warburton, 1958). Over thirty years ago, the National Accounts Review Committee of the National Bureau of Economic Research wrote of the national accounts banking imputation: "What is needed is a thorough review...of the treatment of financial intermediaries in the national accounts with a view to developing an alternative...procedure that would conform more closely to the realities of the activities of these enterprises" (National Bureau of Economic Research, 1958). See also the criticisms of Bowman and Easterlin (1958) and Hodgman (1969); recent statements of similar criticism are Ruggles (1983), Sunga (1984) and Zee (1981).

The long-standing economic criticisms of the national accounts banking output model have not been answered so much as simply brushed aside. In the

current draft of the proposed SNA revision, the banking treatment proposed by Yntema in 1947, as modified in United Nations (1968), is to be continued with little change.[5] The reason why banking continues to be represented by the Yntema (1947) model is that more realistic depictions of the activities of banks conflict with rules that have been adopted elsewhere in national accounts--particularly, rules for handling interest payments. The economic appropriateness of the output measures for individual industries, such as banking, have been accorded secondary importance in the design of national accounts. These priorities are summarized in Haig's (1985) complaint that "The arguments that have been put forward to justify a change [in the measurement of banking output] amount often to little more than the assertion that a change in treatment of interest is necessary to solve the banking problem."

I review, in Triplett (1991), these "other" considerations that motivate the national accounts treatment of interest, and in consequence, of banking output. I believe that most of them are outmoded. As noted already, the lending activity of banks, that is their provision of finance as a productive service to borrowers, is definitionally excluded from the national accounts measure of banking output. This exclusion logically follows from a set of assumptions and hypotheses, which include: (a) a view of the economic role of finance that, as Richard Ruggles pungently remarked, "only a very conventional national accountant or a Marxist would dare to suggest" (Ruggles, 1983); (b) a view of value added that is mostly inconsistent with the economic theory of production, as incorporated in, say, Sims (1969) or Sato (1974); and (c) a view of "property income" (interest is property income and property income is a component of value added) that seems obsessively concerned with old debates

about the functional distribution of income (i.e., between capital and labor), and that, even so, has been applied at a level--the measurement of industry output--that was of little importance to these debates even in the days when the functional distribution of income was of more central importance to economics than it is today.

B. Bank Production Function Approaches

Measures of bank output have also been developed to carry out research on bank regulation issues. If substantial economies of scale or economies of scope exist in banking, for example, large banks have competitive advantages over small banks, so that bank deregulation might increase concentration in banking. To explore such questions, researchers have estimated explicit multi-output production or cost functions, where various bank financial "outputs" and the usual capital and labor inputs are specified. A dozen studies of this kind are surveyed in Clark (1988), and Hancock (1991) provides a comprehensive review of what I will characterize as the "bank production function literature."

Alternative approaches to banking output appear in the bank production function literature. One strand might be termed the "activity" approach: One simply counts all of the activities of banks.[6] No distinction is made between inputs and outputs; or perhaps more precisely, anything a bank does--such as accept deposits--is defined as a bank output, so long as there is a cost associated with it, or it absorbs real resources. Benston, Hanweck and Humphrey (1982, p. 440) remark: "Output should be measured in terms of what banks do that cause operating expenses to be incurred." The Bureau of Labor Statistics (BLS) banking labor productivity measure (see Dean and Kunze, 1990) uses an output concept that is essentially the activity approach--bank output

includes counts of loan and deposit activities (such as loan applications processed and checks cleared), that are weighted together, following the usual BLS methodology, by the labor hour inputs associated with the various activities.

In a second production function procedure in the bank regulation literature, the researcher distinguishes between banking activities that are concerned with processing the bank's financial inputs and those banking activities that are properly considered the outputs of a bank. For example, Mester (1987) assumes, of savings and loan institutions, that "output is best measured by the dollar value of earning assets of the firm, with inputs being labor, capital, and deposits;" three outputs (two types of loans, plus other assets) and three deposit inputs (passbook, NOW accounts, and certificates) were specified. Because only bank assets, and not bank liabilities, are specified as bank outputs, this is usually termed the "asset" approach to defining bank output (though sometimes it is also referred to as the "intermediation" approach). As recorded in Figure 1, the asset approach implies that banks buy funds and sell funds, much the same as any other specialized merchant.

In both activity and asset approaches, authors frequently comment that it is difficult to determine whether particular activities (deposits are the most questionable ones) are bank inputs or outputs. Both approaches explicitly or implicitly accept that the determination of bank inputs and outputs should be derived from a theory of banks as financial firms, yet no such theory is developed very far, and for the most part readily available data are employed in the empirical work in a decidedly ad hoc manner. Authors also consider whether bank output activity is best specified by the count of

the numbers of loans (and deposits, in the activity approach) of different types, or by the dollar volume of loans (and deposits); in many cases, both measures are used on the grounds that, empirically, both contain information.[7] Whether other characteristics of loans (riskiness, for example, or compensating balance requirements) should be be included in the analysis is often discussed, but less often acted upon.

A third production function approach eschews determining, a priori, whether a particular banking activity is an input or an output. This "user cost" approach first appeared in the work of Hancock (1985), is followed by Fixler and Zieschang (1990), and is developed comprehensively in Hancock (1991). Appealing to Barnett's (1980) notion of the "user cost of money," Hancock permits a particular banking activity to be an input or an output according to the sign of its derivative in a bank profit function, which she estimates empirically. In the empirical results, loans are bank outputs (which is consistent with both activity and asset approaches--and, of course, inconsistent with the national accounts approach). Hancock reports that time deposits are inputs, but that demand deposits are outputs. This last result is also replicated in Fixler and Zieschang's work.

3. Summary: Concepts of Bank Output in Empirical Implementations

There is a striking contrast between the concept of bank output incorporated into national accounts, on the one hand, and any of the three variants found in the bank production function literature, on the other.

In the bank production function literature, all authors agree that loans constitute the primary output of banks. In the national accounts measure, bank loans are not a part of the output of banks at all. Since there is a strong dissent to the prevailing view (see above), one may say that the

function and national accounts literatures are wonderfully disjoint. What is considered in the one is ignored almost completely in the other. What is measured as the primary output in the one is excluded definitionally or by oversight in the other. What is controversial in the one is conventional knowledge in the other. And--it goes almost without saying--what is frequently cited in the one is outside the corpus of knowledge of the other.

II. What Do Banks Do? The View from the Finance Literature

The national accounts and the bank production literatures both view the operation of banks from the perspective of the theory of production. That is, they both look for analogies between the banking business and models that are typically applied to manufacturing or other nonfinancial business. This production view is more readily apparent in the bank production function literature, arising as it does from the microeconomically-oriented productivity measurement literature; but even the national accounts approach involves making up a story about imputed bank revenue, and the critics of the conventional national accounts literature (such as Warburton, 1947, 1958) ask explicitly why bank revenue is not the proper indicator of what banks produce--that is, they draw analogies with nonfinancial firms.

A much older tradition, however, is the approach to banking contained in money and banking, and in finance. These literatures contain a quite different perception of what banks do. As Fama (1980), in a well known article, remarked:

> "...the concern with banks in macroeconomics centers on their role as portfolio managers, whereby they purchase securities from individuals and firms (and a

> loan is, after all, just a purchase of securities)
> which they then offer as portfolio holdings (deposits)
> to other individuals and firms."

Gurley and Shaw (1960) discuss the operations of banks in nearly the same language--deposits are securities that banks sell to others, while their loans are securities that they buy.[8]

A reinforcing view comes from the fact that bank deposits are part of the money supply. Modelling banks as sellers of deposits is equivalent to depicting them as suppliers of money, which conforms to the major interests of money and banking and of macroeconomics. Pesek (1970), Saving (1977) and Towey (1974) construct models of banks as firms that produce and supply money. Towey also considers and rejects the idea that loans (or "credit") might be considered a bank output, and Pesek denounces it.[9]

In the language, at least, the finance approach seems to suggest modeling banks' real economic flows in the opposite direction from that implied in the literature reviewed in section I. Banks produce and sell deposits, and loans simply represent the securities they buy.

Obviously there is little economic difference in the finance and production models of bank behavior. The words "selling finance" and "buying securities" are merely alternative descriptions of the same transaction.[10] Language, however, influences thinking. There is no question that the securities-market orientation of finance, money and banking, and macroeconomics has influenced the attempt to construct production models of banking, and, one suspects also, introduced semantic ambiguity that has made it more difficult to answer the fundamental question: "What are the inputs and what are the outputs?"

In the remainder of this paper, I proceed from a production theoretic approach to defining what a bank does. One reason is that the production-theoretic view underlies most of the relevant parts of economic measurement, which suggests that it is indeed what we want for measuring the output of any industry. Though the national accounts themselves are macroeconomic, the theory that applies to their components (and to much of the aggregates) is based on the theory of production or the theory of the consumer. It seems appropriate, therefore, to extend into services the framework that has in the past proved useful for other applications.

The finance-macro approach, in the other hand, because it concentrates on portfolio management by bank customers, neglects the real side of the economy. Baltensperger, in an insightful review, noted that in a complete theory of the financial firm, costs and the banks' use of real resources must be included:

> "This requires that somehow the nature of the services produced by the firm makes an appearance in the model, in one form or another. This is not the case in models which restrict themselves to a direct application of traditional portfolio theory to the financial firm, and it is hard to achieve in such a framework. Of course, it is true in a formal sense that a financial firm is nothing but a collection of assets and liabilities. But so is General Motors.... One of the major tasks of a theory of the firm must be to explain how the firm combines resources of various kinds in order to generate...net yield and profit streams. This requires going beyond a pure traditional portfolio approach. These comments, however, are not meant to imply that portfolio

theory and risk aversion cannot play a useful role as an element of a more complete model" (Baltensperger, 1980, pp. 27-28).

Neglect of the real side of banks' activities might be justified if finance were wanted primarily for the acquisition of capital goods (which seems often implicitly assumed in the production, macroeconomics, and national accounts literatures). In this case, measures of the capital goods may be sufficient for explaining production, and finance contributes nothing beyond what is contributed by the capital goods themselves; if so, adding an independent measure of finance to the borrower's production function becomes redundant (this seems more or less the finding from empirical exercises along that line), and the analysis of finance involves only relatively uninteresting questions of ownership and financial claims.[11]

The idea of productive capital, however, also encompasses what Hicks (1974) referred to as a "fund"--that is, it includes working capital. If bank finance is largely or partly, directly or indirectly, the provision of working capital, then it is independently a productive input: It is a resource that is necessary for productive activity to take place (beyond the capital measured by capital goods). In this case, finance is appropriately analyzed as the provision to the borrower of a productive service by the bank.

To summarize the argument of the last several paragraphs, the real side of the economy is better modelled by treating banks as sellers of finance to borrowers than by viewing them as buyers of the borrowers' securities.

A second reason for preferring the production-theoretic approach is pragmatic. It may well be useful for some purposes to think of the bank's provision of finance to a borrower as buying his securities, and to think of banks as nothing more than specialized securities dealers. Indeed, if banks'

continue to shift into other financial business, it may become essential to think about them this way. But that way of looking at banks does not presently seem useful for measurement purposes because determining what securities dealers produce is, if anything, less well developed than is the measurement of banking output (see, on measurement of securities markets, Bresnahan, Milgom, and Paul, 1990). Some sort of start has been made in measuring banking output by using analogies with nonfinancial production, and unsatisfactory as it may be, the work that has already been done at least points the way. Drawing analogies with securities dealers shows, so far, considerably less potential.

Niehans and Hewson (1976, pp. 1-2) remark that "the traditional paradigm for banks [in money and banking] is the intermediary that uses monetary deposits to buy non-monetary assets...," which focusses attention on money creation and on liquidity. They compare the traditional with "an antithetical paradigm, at once older and newer, of a bank borrowing and lending funds...," which shifts the focus to "an important function of financial intermediaries which the dominant model tends to suppress, namely, the function of efficient distributors of funds." This "antithetical paradigm" underlies the measurement model for banking output that is pursued in the following section.

III. A Measurement Model for Bank Outputs and Inputs

A modern bank is involved in a complex network of financial activities. While retaining their traditional functions--accepting deposits, making loans, and executing payments--banks now sell securities and insurance, and provide other financial services. Concurrently, "non-bank" banks have invaded the bankers' traditional markets; securities brokers, for example, sometimes

provide what are essentially bank-type checking accounts. The growth of "commercial paper" (notes issued by large commercial and industrial enterprises, in lieu of borrowing from banks) permits some categories of borrowers and lenders increasingly to bypass altogether the traditional bank intermediation process. Banks have become conglomerate enterprises whose functions do not all fit into the definition of "commercial banks" (SIC 602 in the U.S. Standard Industrial Classification Manual), and conglomerate enterprises from other SIC classifications have invaded banking.[12] However one might have resolved measurement issues in the banking industry in, say, 1970, resolving them in 1990 must take account of many more economic activities.[13]

Measuring bank output is made yet more difficult if we recognize that banks provide information to their clients (the records of payments and receipts in depositors' monthly statements, for example), or if we think we must (as we would in a completely adequate measure) allow for banks' provision of amenities such as convenience (location of branches or automatic teller machines, and hours of operation), ambiance of banking offices, and so forth. Such attributes of banking are analogous to quality characteristics for manufactured goods, which one would certainly not want to ignore if the subject were the output of, say, automobiles.

Complex questions must at some stage be dealt with. However, the essence of the banking measurement problem--what are the outputs and what are the inputs?--arises in full force for the simplest of banks, those that only accept deposits and make loans. The classic, unresolved issues in measuring banking output (and perforce, measuring banking prices) do not derive from the more complex banking activities that have arisen in the 1980's, they adhere to

the whole history of banking. Thus, it is useful to step back from the more complex structure of modern banking, and to ask (to parody an old song): "What do the simple banks do?"

For present purposes I consider the simple bank that only makes commercial and consumer loans, and whose sources of funds consist only of demand deposits and saving deposits. This simple bank pays time depositors only in interest. There are no services provided to time depositors (except safe-keeping). Demand depositors receive only "free" check cashing and use of automatic teller machines (ATM); they earn no interest. The bank carries deposit insurance, the cost of which is, at least in part, born by depositors in the form of lower returns on deposits. In addition, the bank rents safe deposit boxes, for which an explicit fee is charged.

A. Banking Services from the Viewpoint of the Depositor

Assuming for simplicity that depositors hold their wealth only in bank deposits or in safe deposit boxes, the individual depositor faces a maximization problem in which--in addition to the usual choices among consumption goods and saving--the individual needs personal business services. Personal business services comprise safeguarding of wealth, as well as cash management/transactions services (check cashing and ATM-usage), which arise from consumption purchases and from wealth management.

This maximization problem may be set up in the following form:

(1) Max: $U = U(G, \Delta W)$

s.t $Y (= wL + rW) \geq GP_G + [BP_B + SP_s] + \Delta W$

where U is utility, G designates the vector of consumption goods (with prices P_G), W is a vector of wealth holdings (here, assumed to be bank deposits), with returns r, so that ΔW is the change in total wealth (in the period equal

to saving). I write the change in wealth (ΔW) in the utility function either because it is desired for its own sake, or as a variable encompassing the future lifetime consumption path; the latter is obviously properly set up as a multi-period problem involving future prices, but this is ignored here for simplicity.

In the constraint, Y is total income, composed of labor earnings (wL) plus earnings from wealth, which is allocated to consumption, purchase of business services, and saving (change in wealth). The bracketted term [B P_B + S P_S] consists of bank transactions services, B, with <u>implicit</u> prices P_B, and safe-keeping services, S, with implicit or explicit prices P_S.

Of particular interest are the <u>demands</u> for B and S, which are:

(2) $B = B(G, W, P_B)$

$S = S(W, P_S)$

Since the services are not utility-generating for their own sake, the demand for bank transaction services is derived from the costs of transactions for consumption goods, so that the level of consumption activity enters the derived demand equation, and (trivially in this case) for wealth administration. The demand for safe-keeping of wealth is also a derived demand.

The wealth vector is assumed here to consist only of deposits in banks--safe-box deposits, time deposits, and checking deposits (W_0, W_1, and W_2, respectively). The returns on those deposits are, respectively, 0, r_1 (interest), and r_2 (paid in the form of in-kind services). Thus, r_2W_2 (the income from wealth held in demand deposits in the income term) is identically equal to BP_B, the vector of depositor services, valued at their implicit prices, in the expenditure side of the equation. The reason for this is that

the consumer adjusts wealth holdings W_1 and W_2 (ignoring W_0 for the present) so that W_2 is the minimum balance in his checking account that will provide the optimal quantity of bank services (B), determined from the demand function in (2).

Suppose the bank paid interest on checking account deposits and instituted explicit charges for services (that is, suppose the vector P_B were explicit instead of implicit prices), but that the explicit vector does not differ from implicit P_B. No changes would result in the model. If income taxes were included, however, individuals might prefer implicit charges because the quantity r_2W_2 would not show up as income for tax purposes.

There are analogous expressions to equation sets (1) and (2) for business deposit holders, in which B and S are treated as business costs. They need not be written out here.

B. Bank Outputs.

Turn now to our major concern, the output of the bank. Using Warburton's (1958) criterion that an output yields revenue to the bank, the simple bank has three explicitly-priced outputs: The number of commercial loans, the number of consumer loans, and the number of safe deposit boxes.

The size of loans--and of the safe deposit boxes--may be viewed as analogous to the conventional price index "quality problem"; in the case of loans, size matters not just in the sense of a simple package size problem (such as the number of ounces in a box of corn flakes), but also because it introduces an element of risk. Another aspect of that "quality problem" is the fact that loans are the provision of finance to the borrower: The characteristics of finance that make it productive to the borrower may not be captured at all well with simple counts of loans and their sizes (just as the

provision of health services is not measured adequately in terms of numbers of doctor visits, hospital rooms, and so forth). The term of the loan, conditions of repayment, whether "compensating" balances are required of the borrower, and so forth, are aspects of the "quality problem" in measuring banking output that have unique problems, but also have parallels in the measurement of "quality change" in the goods-producing industries (for discussions of the quality problem and solutions to it, see Griliches, 1971, and also contributions by Feldstein, Griliches, Lipsey, Triplett, and Diewert, in Berndt and Triplett, 1990).

Another aspect of the "quality" of bank loans is risk of default.[14] When a car is sold, the seller usually does not care who the buyer is or what the buyer does with it. When banks provide finance to borrowers, it is more like renting: An apartment owner cares about the characteristics of the renter, and the monthly rental may reflect renter characteristics as well as those of the apartment. In employment contracts, also, characteristics of both buyer and seller of labor matter to the other party to the transaction. In labor economics such contracts are usually modelled by assuming that both parties exchange bundles of characteristics, with the observed wage payment representing only one element of the transaction (Antos and Rosen, 1975). In principle, bank loans should be handled equivalently though there are formidable data (and conceptual) problems in the implementation. I leave all these essential problems to one side, for the moment.

The simple bank also has <u>unpriced</u> outputs, including "free" checks cashed and automatic teller machine (ATM) transactions. These are clearly outputs of the bank, since they provide essential transactions services to the depositors (see equations (1), and (2), above). Safe-keeping services may

also be an unpriced output (where safe-keeping takes the form of time or demand deposits). Again, the simple counts of the number of checks and the number of ATM transactions may not fully capture the services provided: ATM availability after normal bank closing hours is a different and probably more valuable service to the customer than its use in other times. As before, these "quality" issues are vital points for an effective measurement of bank output, but they are set aside for the present to focus on matters that appear in the literature so far.

The quantities of "free" checks and ATM transactions can in principle be observed. Their prices, however, must be imputed, both for valuing bank output and to provide the measures P_g in equations (1) and (2). The imputed values of these unpriced outputs must be combined with the outputs for which explicit charges are imposed: This combination will present some unique problems, not normally encountered in measuring, say, the output of shoes.

C. Reserves

The simple bank, as with all banks, must maintain reserves. Reserves are normally placed on the asset side of a bank's balance sheet, along with its income-earning loan portfolio. Baltensperger (1980, p.3) remarked that "the first analytical models of bank behavior were models of bank reserve (liquidity) management...[in which] the problem to be solved [is] the optimal allocation...among various assets, with particular attention being paid to the choice between earning assets and reserve (liquid) assets." He shows that the optimal level of bank reserves is given by

(3) $R = R\,(r/p,\ \sigma_x)$

where r is the interest rate earned on loans, p the cost per dollar of a reserve deficiency (examples of such costs are emergency borrowing or

emergency sale of assets to cover a deficiency), and σ_x is the standard deviation of the distribution (assumed normal, for simplicity) of the outflow of deposits (withdrawals). In this model, a bank determines the level of reserves by equating, at the margin, the revenue foregone or opportunity cost of holding reserves with the expected cost of a reserve deficiency, where σ_x summarizes the probability of the bank's encountering a reserve deficiency.

This formulation makes clear the common-sense statement that reserves are a cost of the banking business, a cost that is incurred not only by the fact that deposits are of shorter durations than earning assets, but also because at least some of them are payable on demand, and hence their durations are unknown, except as probabilities. If it were not for the latter fact, withdrawals would be deterministic, not probabilistic, and rather than holding reserves, unregulated banks could operate by careful matching of maturities. In any event, for our purposes equation (3) makes clear that we should treat reserves as a cost of the bank, specifically as a cost incurred when financial inputs are obtained in the form of deposits payable on demand.

The fact that σ_x appears in (3) suggests another point of significance for measuring banking output. "Free" checks usually accrue only to depositors who maintain some <u>minimum</u> (not average) balance. When depositors commit to a checking account minimum balance, one might expect this to reduce σ_x, the variability of deposit withdrawls. If so, it accordingly will reduce the size of optimal reserves, and reduce bank costs. Thus, even unregulated banks may continue to compensate depositors with unpriced services, rather than paying interest on deposits and charging explicit fees for checks and ATM's.[15]

D. <u>Bank Inputs</u>.

The simple bank has as inputs the usual capital (that is, capital goods,

buildings and equipment), and labor. In addition, it is a financial firm, so it must obtain financial inputs, which one may count as the numbers of demand deposits and of saving deposits. What was said, above, about "quality problems" in measuring output by the simple count of loans applies as well to measuring financial inputs by counting the number and size of deposits.

E. A Measurement Model: What Banks Do

It is convenient to begin from the model in Baltensperger (1980), who specifies that a bank such as our simple one maximizes profit according to:

(4) Max: $\Pi = \Sigma\, r_i L_i + \Sigma\, f_g S_g - \Sigma\, i_j D_j - \Sigma\, w_h v_h$

Subject to: $H\,(L, S, D, v) = 0$

and $\Sigma\, L_i = \Sigma\, D_j\,(1-k_j)$

where the elements in the vectors:

L are loans (or other earning assets) with returns r (and i = 1, 2),

S are directly-priced bank services, with fees f,[16]

D are deposits, with interest costs i (and j = 1, 2),

v are inputs (labor, materials, services of capital goods), with unit input costs w,

and k is the vector of reserve ratios on different types of deposits.

Note that our simple bank has only one directly-priced service (safe deposit boxes), so the vector S has but one element, $S = S_1$, so $\Sigma\, f_g S_g = f_1 S_1$. Similarly, only time deposits pay interests, so (designating time deposits as j = 1), $\Sigma\, i_j D_j = i_1 D_1$ (because $i_2 D_2 = 0$). To avoid the "quality" problems discussed above, I assume that within each class (commercial, consumer loans), earning assets are all the same size with the same degree of riskiness. The

same is assumed of deposits. In short, all variables (E, S, D, and v) are assumed to represent homogenous groupings.

In (4), the bank production function, H (·), admits the possibility that deposits must be "produced"--that is, that generating them or processing them absorbs some quantity of real resources, which we may index v_{hD}. The relevant sign in the profit function, however, states that deposits represent costs to the bank, they are not revenue-generating net outputs. In the balance sheet constraint (the final equation in (4)), the vector k_j is determined by:

$$k_j = R_j/D_j, \text{ if } k_j > k_j^*$$

$$k_j = k_j^* \text{ otherwise}$$

where R_j is the optimal reserve determined by equation (3), above, and k_j^* is the legal minimum reserve ratio for deposit type j.

The inclusion of a balance sheet constraint requires additional comment. All firms have balance sheets, but they do not normally appear in constrained maximization problems.

The bank production function, H (·), contains the real inputs necessary for making loans--the resources required for processing forms, evaluating credit standings, and so forth. Loans can be "produced" in this sense without deposits. The function H (·) does not convey the economic process of transforming deposits into loans in the way that a production function for steel depicts the transformation of iron ore into steel. Instead, H (·) includes the real inputs that are employed in the banking business, and D appears there as a use of real inputs, and not necessarily as an input to the financial transformation process itself.[17]

The balance sheet appears in (4) because this is a financial firm: Loans cannot legally be made or assets acquired without deposits or other

sources of funds. The relation between financial inputs (deposits, or purchased funds) and the bank's acquisition of earning assets is a financial relation, not a physical production process. It is therefore expressed in (4) as a financial constraint on the profit maximization process that is parallel to, and in addition to, the physical transformation function: Both constraints are necessary to delimit the financial firm's "production" set. In most empirical work on bank production, the balance sheet constraint has been ignored.

1. Treatment of deposits in the basic model.

Our simple bank pays interest on time deposits, so the interest cost of acquiring funds from that source appears explicitly, as the per-unit cost of acquiring deposits, in the profit function of (4). Recall that in the empirical work on banks (section I.B, above) time deposits were found to have the negative sign indicated in the profit function of (4). Direct interest cost may understate the true return to the depositor because the deposit yields safe-keeping service as well as earnings.

Note that even in the case where interest cost accounts for the entire cost paid out by the bank to owners of funds, resources (v_{h0}) are still required to process those deposits. That some bank activity has a cost or absorbs labor units does not in itself indicate that it is an output, contrary to some assertions that have appeared in the bank production function literature. The observation that processing deposits entails costs is no different from observing that workers are required on the loading dock to unload rolls of steel sheet that are inputs to the production of gadgets.

Checking accounts pay no interest, so their cost does not appear explicitly in the profit function. The bank does incur costs when it acquires

funds through checking deposits, but those costs consist of the "free" checks, ATM usage and like services provided to depositors. The costs of generating these "free" services account for a portion of the input costs associated with deposits, $w_h v_{hD}$.

If one is interested only in accounting for the profit of the bank, this indirect accounting for checking deposit acquisition costs causes no problems. Measuring the output of the bank, however, or its productivity, is another story, which is addressed next.

2. An alternative model for deposits: Explicit charges and interest.

Suppose our (unregulated) simple bank were to pay interest on balances in checking accounts, but institute a system of direct charges for previously "free" services. In place of the vectors originally appearing in (4), we now have:

$$(5) \qquad S = (S_1, S_2, S_3)$$

$$\Sigma f_g S_g = f_1 S_1 + f_2 S_2 + f_3 S_3$$

$$\Sigma i_j D_j = i_1 D_1 + i_2 D_2$$

where the elements in the vector S consist of safe-deposit box rentals, check-cashings and ATM services, respectively; their associated explicit fees are f_1, f_2, and f_3, respectively; (D_1, D_2) indicate time and checking-account deposits, whose respective per unit interest payments are (i_1, i_2). This compares with the observed quantities in the initial situation:

$$S = S_1$$

$$\Sigma f_g S_g = f_1 S_1$$

$$\Sigma i_j D_j = i_1 D_1$$

Substituting the elements of (5), representing the new pricing regime, into (4) raises the quantity $\Sigma f_g S_g$, bank revenue from explicitly-priced

services. But it also increases $\Sigma\, i_j D_j$, the interest cost of deposits.[18] The vector S is, of course, a part of bank output, so observed bank output in (4) is also greater under explicit pricing.

If the set of explicit service charges and explicit returns in (5) were equal to the set of charges and returns that were implicitly in (4) before, no changes appear in the depositors' demand model (equation 2). The total bank provision of depositor services, accordingly, is the same as before (though the explicitly priced portion rises). For the same reasons, the total payments by the bank to depositors are the same as before (though the explicit payments rise). Thus, it seems reasonable to assume no change in the bank's real costs, $\Sigma\, w_h v_h$ (that is, the resources required to cash checks and service ATM's are the same as before).

It must accordingly be true that the depositor services output vector, the bank's revenue from sale of depositor services, and the depositor payments in (5) are the correct ones in both situations. The two situations do not differ in what was produced. They do not differ in the value of the payments the banks made to depositors. They do not differ in the quantities that appear in the consumer's business services utilization equation (equation 2) and they do not differ in real household income. The only difference between the two situations is that in the initial one the bank and its depositors engage in a barter transaction that is not recorded explicitly in either of their monetary accounts; in the second, the barter transaction has been monetized, rendered into the form of a conventional transaction, and recorded in the conventional way both in the bank's records and in those of the depositor. It must accordingly be true that both bank output and bank payments to depositors were understated in the initial situation where

checking deposits earned in-kind service, not explicit interest.

What is required to make the two situations comparable is to add to the bank output measure in (4) an imputation for the value of unpriced bank services. An identical imputation in (4) must be made to bank payments to depositors. In summary, the banking model of (4) is modified by substituting the following vectors in place of those explicitly-measured quantities that were originally observed in the simple bank's records:

(5a) $$S = (S_1, [S_2], [S_3])$$

$$\Sigma f_g S_g = f_1 S_1 + [f_2 S_2] + [f_3 S_3]$$

$$\Sigma i_j D_j = i_1 D_1 + [i_2 D_2]$$

where the square brackets designate imputations. How these imputations may be made is discussed in the following section.

To anticipate a criticism of these imputations, no double-counting occurs. The imputation increases measured bank output and measured bank outlays to depositors. The imputations also raise measured bank productivity. For the reasons already noted, the output and productivity measures that include the imputations are the correct ones. The imputation does not affect bank profit (which was correctly measured before), and it does not change bank value added, because the resources employed in generating "free" checks (a portion of v_{h0}) were counted all along. Parenthetically (for present purposes), because the imputations measure the correct costs and revenues from banking operations, they also yield the correct marginal conditions for profit maximization for the maximization problem stated in (4), and thus they are the correct measures for analyzing bank behavior.

In Figure 1, the measurement model developed in this section is termed the "Services" model because it includes all services of banks--finance

provided to borrowers and deposit services to depositors. No other banking model includes both these types of services.

F. Imputations for Unpriced Depositor Services

We require imputations for the volume of unpriced services provided to checking account depositors ($[S_2]$ and $[S_3]$ in (5a)), for their implicit fees or prices ($[f_2]$ and $[f_3]$), and for the implicit income payments made by the bank to depositors ($[i_2D_2]$). For both time and checking accounts, we also require an imputation for the value of the safe-keeping service provided by bank accounts.

1. An imputation for the simple bank.

A basis for imputing prices for checking account services suggest itself if we observe that our simple bank adjusts the schedule of "free" checks it makes available to depositors to the balances retained in the depositors' accounts, and if it also adjusts (even if with a lag) the "free" checking or other service schedules to changes in the interest rates paid on time deposits. Such adjustments indicate that the bank is acting to equalize, at the margin, the cost of funds it acquires from time and checking deposit sources. In fact, if the bank does not equalize the marginal costs of funds from different sources, it is not maximizing profit.[19]

If banks equalize the marginal cost of funds obtained from time and checking deposit accounts, the value of "free" services provided with checking accounts can be estimated by applying the interest rate earned on time deposits to balances held in checking accounts.[20] The total value of unpriced checking account depositor services is imputed by computing:

(6) $\Sigma_{(g = 2 \ldots n)} f_g S_g = i_1 D_2$

where $g = 1$ is safe-deposit box services (excluded from this calculation), but

all other services (free checks, ATMs, and any other services provided with checking accounts) are included in the imputation. Total checking account deposits, rather than minimum deposits required to earn services, are the variable on the right because banks are assumed to equalize the cost of funds at the margin, and excess funds, above the minimum required to obtain unpriced services, are factored into both the banks' decisions and those of the depositors.

This imputation method is simple, and it can be applied to banks in a variety of regulated and unregulated settings. However, it provides no estimate of the implicit fees or prices (f_2 and f_3), and thus provides no way to separate the total value into price and quantity components.

One solution is to obtain data on quantities of free checks cashed, ATM usages, and other services (or a sample of them). A quantity index of these services can be computed, using as weights estimates of the relative costs of providing them. The quantity index of services can be used to "deflate" the total volume of services to obtain an implicit price index. Let Q(S) be such a quantity index of bank services.[21] Then

(7) $i_1 D_2 / Q(S) = P(f)$,

where P(f) is an implicit price index for bank imputed services.

The "deflation" procedure in (7) is the opposite of the usual national accounts practice in which a price index is used as a deflator to obtain the quantity index that is "real" GNP or some component. It is justified in this situation because the quantities in S can be gathered by survey methods, where the prices or fees (f), being implicit, cannot.

Our simple bank was assumed to pay no interest on checking accounts and to provide no transaction services on interest-bearing accounts. This

assumption describes fairly well the banking regulatory environment in the U.S. prior to the 1980's. The imputation in (6) and (7) is probably the only one that can be used in the pre-1980 period.

2. Hedonic imputation for bank services.

In the present deregulated U.S. bank environment, a large number of different types of bank accounts are offered. Depending on the type of account, the account holder's minimum balance, or other characteristics of the account, the account holder may receive varying amounts of interest, and may pay varying monthly amounts in explicit service fees or as explicit charges for check cashing and other services. Conversely, depending on the type of account and the characteristics of the account, the account holder will receive varying amounts of "free" services, such as check cashing and the like.

Thus, the account holder's nominal monthly cost of a bank account may be positive or negative, depending on whether the explicit charges for bank services are greater or lesser than the explicit interest earnings paid on the account's balances. But the true monthly cost of a bank account, and the true return from it, differs from the nominal cost and return.

One can view the full monthly cost of a bank account as consisting of charges for a "bundle" of priced and unpriced services. The imputed payments to the bank for unpriced services come from the interest the depositor forgoes, relative to an account of a similar size where explicit interest is paid. A hedonic model is a natural tool for disaggregating the bundle of priced and unpriced banking services and payments.[22]

Empirically, the banking hedonic model differs from the one encountered in most price index literature (see Griliches, 1971). In the customary

hedonic model, the total value of the transaction is known and appears on the left side of the equation (e.g., the price of an automobile, or of a computer); though the quantities in the characteristic bundles are known (the speed and memory size of a computer), none of the characteristics prices are known, and therefore they must be estimated. In the banking case, the total value of the transaction is not known (because it includes imputed services), which means that we cannot use, without modification, the normal hedonic model.

A model sharing many of the banking model's properties appears in labor economics. Smith and Ehrenberg (1983), and Thaler and Rosen (1976) estimate the "trade offs" between nonpecuniary ("benefits") portions of compensation--or between improvements in working conditions--and direct wage payments. Workers are assumed to trade compensation for nonpecuniary components (not necessarily on a dollar-for-dollar basis--see Triplett, 1983), so that the value of the nonpriced components may be inferred from variations in wages associated with quantities of the nonpecuniary elements. In these labor market cases, the full value of the transaction (the value of total compensation) or of "full income" (the worker's opportunity locus--see Atrostic, 1981) is not known and must be imputed by valuing the unpriced elements of compensation.

In labor economics, the value of total compensation, or of full income, depends on the worker's endowment of human capital. The human-capital/earnings relation provides one hedonic, or quasi-hedonic, equation. A second equation relates the allocation of total compensation to direct wage payments and individual non-wage benefits, working conditions, or other non-pecuniary returns from employment. This results in a two-equation system:

(8) Total compensation = H (human capital)

(9) Wage = B (benefits, nonpecuniary compensation, etc.; given: human capital endowment)

In Figure 2, the level of compensation, and hence the distance from the origin of the compensation "trade-off" surface, is determined by the first equation; the slope (or slopes, where there are multiple benefits) is estimated from the second. Notice that the observed "price" in this system (the explicit money wage payment) appears on one axis of Figure 2.

In banking, I assume that the total yield (designated R_2) on a checking account is determined by its size, in particular on the minimum balance (δ_m) maintained in the account:

(10) $R_2 = \delta(\delta_m)$

Equation (10) preserves the possibility that larger accounts, say, may earn higher returns.

I also assume that the net expected return to the depositor will be equal across various kinds of accounts (e.g., a "NOW" account with limited free check cashing and explicit interest, compared with a "normal" checking account with unlimited check cashing and no explicit interest payments) that maintain the same balance. This means that the (depositor's) valuation of unpriced services, when combined with explicit interest payments from the account, if any, equals the total yield. Bank account holders choose the type of account that, given their estimated usage, minimizes the full cost of the account or maximizes its total yield (rather than minimizing explicit charges or maximizing explicit interest receipts). The explicit interest paid on an account is accordingly a (decreasing) function of the unpriced services provided, given the size of the account, or (where i_2D_2 indicates explicit

interest):

(11) $i_2D_2 = S\ (S_2, S_3, \ldots, S_n; \delta_m)$

This banking model is very similar to the labor model and is likewise illustrated with the same apparatus (Figure 2). The level of the return on the deposit depends on its size, or minimum balance (equation 10), which is shown as the distance from the origin of the explicit interest-bank service trade-off surface. Larger accounts yield greater returns, so we expect that $\partial R_2/\partial \delta_m > 0$. The slopes of those surfaces, which give the implicit prices of bank services in terms of interest foregone, are determined by equation (11). These trade-off surfaces will not necessarily be linear as they are diagrammed in Figure 2, but we expect that $\partial(i_2D_2)/\partial S_f < 0$.

Both banking and labor models involve a two-equation hedonic model. In the first equation in both systems, a wealth or endowment variable (human capital, or the size of the deposit) determines the level of total payment in the transaction. In the second, the total payment is allocated between an explicit money payment (wages or explicit interest) and some imputed valuations (the value of non-pecuniary benefits, the value of unpriced services). In the latter equation, implicit prices are estimated, in both cases, for all the elements in a vector of benefits, or of depositor services.

Put another way, two equations are necessary for econometric considerations. Suppose a simple-equation hedonic model were estimated (equation 11 <u>without</u> the semi-colon). The variables in S are functions of δ_m, which violates conditions for single-equation estimation.

Banks will not only provide schedules of services for differing minimum balances, the services (and possibly the minimum balance as well) will also be a decreasing function of explicit interest, if any, paid on the account.

This "price" schedule for bank accounts and their services summarizes the terms on which the bank offers personal business services to depositors. It provides the implicit prices for the bank services in the "budget" constraint of (2). The schedule is a non-linear, two-equation system in the variables i_2D_2, the interest paid on transactions accounts, the various services (S), and the minimum balance, δ_m, attached to each account.

For a depositor who coordinates his transactions to hold his balance at one of these minimum balances, the interest foregone is equal to

$$(i_1 - i_2)\ \delta_m$$

where the "reference rate" i_1 is the rate paid on alternative accounts of the same size that do not offer services. The total amount of interest actually received on a transactions account (i_2D_2) is a function of the unpriced services it provides, and of its minimum balance requirement. The greater the quantity of services provided, for a given minimum balance, the lower the interest paid on the account.

3. Safe-keeping imputation.

Accounts with minimal services still provide safe-keeping services. Little liquid wealth is held in safe-deposit boxes.

The ideal imputation for safe-keeping would come from a hedonic function on risky investments in which both insured and uninsured bank deposits appear. Most bank deposits are insured. Recent experience with government insurance programs suggests that the insurance premia actually paid on bank deposits substantially understates its cost. Alternatives, however, are not readily conceived, and the quantities are small, so I would add the value of insurance premia to explicit deposit interest, and to the other imputed depositor services, as an estimate of the value of bank safe-keeping services. This is

a small component of bank output.

G. Prices

This paper has been directed toward analyzing the flows in banking, and to formulating the concept of bank output, because these have been the points at which the existing literature has been "stuck." Once we have distinguished bank inputs from outputs, separating that output into price and quantity components would require another paper, but a few remarks about that topic are essential.

Price measures for bank depositor services come from the hedonic function discussed previously. In this case, the banking hedonic price index is similar to other hedonic price indexes, and deflation yields the national accounts quantity measure (similar to the way other deflated quantity measures are produced in national accounts).

The price of the portion of bank output contributed by lending activity is a second problem. If all loans were of the same size, same riskiness, same terms of repayment, and so forth, then the price of loans is simply the interest rate.

Loans, however, differ in an enormous number of dimensions, as discussed in section III. B, above. Much of the variance in observed loan interest rates is an extreme case of quality differentials that are conceptually similar to those that are observed in goods-producing industries. This suggests that hedonic indexes are an essential tool in measuring banking price and output. It has long been established, for example, that the kind of "mean-variance" analysis sometimes used in portfolio analysis is simply a financial application of hedonic functions. Existing research on the term structure of interest rates also looks a lot like a hedonic model. Hedonic

price indexes for loans will require similar kinds of analyses, and similar data, to those hedonic functions that have been employed in other contexts.

One would not deny that there is an enormous complexity here. But much of the complexity in banking links very naturally to other analyzes that have become fairly commonplace elsewhere in economic measurement.

IV. Conclusions: Comparison With Other Approaches

A. Value Added or Intermediation

An alternative approach to bank output, one called "intermediation" in the attached table, has not been discussed previously. In the intermediation view, banks do not buy and sell (or rent) funds. Rather, they move funds from place to place, or move them from the "ultimate" lender to the ultimate borrower. Their function is therefore comparable to a pipeline that provides the service of moving natural gas owned by someone else from the producing field to the point of final consumption.

The output of the bank under the intermediation view amounts to the margin between banks' borrowing and lending rates--in terms of the variables in (4), but ignoring fees for directly-priced services:

(12) bank output = $(\Sigma\, r_i L_i - \Sigma\, i_j D_j)$

As noted in the table, the output of the bank under the intermediation proposal equals the value added of the bank under the proposal described in section III (ignoring fees). Somewhat coincidentally (in view of the logic involved), equation (12) is also algebraically equivalent to the national accounts method (equation A). One proposal for the SNA revision used the intermediation concept of bank output (except that the SNA proposal, following the established SNA principles noted in section I, considered that what the

banks were moving was not an economic product or service at all).

For many service industries, and some goods producing industries as well, ambiguity between concepts of value added and (gross) output exists. A pipeline may move other firms' gas, and levy an explicit charge for the transportation service; it may also buy gas at the wellhead and sell it to the final distributor. The production process is the same in either case, and determining which concept of output is "right" is not worth the argument. Similarly, does a shoe store sell shoes (in which case its output is measured in pairs of shoes sold)? Or does it sell retailing services (in which case its output consists of the distributor margin on shoes)? It is not clear that it matters, conceptually. Hill (1977) notes that a repair shop might be thought of as "producing" a repaired car out of an unrepaired one and other inputs, rather than engaging in a production process called "repair services." At one time, grain mills ground the farmers' grain for a fee, rather than buying and selling on their own accounts, and some nonferous smelting operations still function this way. In all of these examples, economic processes are the same, and it is not very fruitful to argue over which interpretation of production or output is conceptually superior.

For economic measurement, it is important instead to ask: "What is the transaction that can be observed?" The pipeline model may work very well for modelling the transmission of natural gas, if in fact this is the way the business is conducted: If a fee is charged for the service of transporting a cubic foot of gas, one can collect fees and services. The difficulty with the bank intermediation view is that one never observes a bank moving funds owned by someone else from place to place and charging a fee for doing so. Defining bank output to correspond to the intermediation approach raises the question:

"What is the price?" There is no clear answer to this question because no transactions can be found that correspond to the bank intermediation model.

B. The Traditional National Accounts Approach

The national accounts approach is the only one that encompasses an imputation for borrower services. Comparison of equations (A) and either (6) or (11) shows that the national accounts imputation of borrower services is far too large. On the other hand, its output measure is too low because it omits bank loan output.

Proposals have been made in the SNA revision to impute free services to borrowers. Services may of course be provided to borrowers in addition to the service of finance. The observed r_1 on some loan may consist of a charge for finance and some other charges for, e.g., bookkeeping and certification of the borrowers' credit worthiness. There, is, however, no particular reason to separate those elements. The observed r_1 measures the full charge that the borrower pays. The checking account depositor, on the other hand, has a full return that is not measured by the term i_2 (in 4 or in 5). Whether or not both borrowers and depositors receive "free" services, the two situations are not parallel because the borrower makes a monetary payment for those services where the depositor is engaged in a barter transaction.[23]

The major difference between the national accounts approach and the proposal in section III is that I accept the provision of finance as a productive service. Thus loans are a part (a major part) of bank output in my proposal, and are excluded from the national accounts approaches.

C. Bank Production Function Approaches

The approach proposed in section III is, of course, very close to those used in much of the bank production function literature. Loans are a major

part of bank output, and deposits are financial inputs. In contrast to the bank production function literature, however, the proposal in section III includes an imputation for unpriced depositor services, which are added to bank output, and accordingly affect bank productivity. Bank production function approaches have ignored the expansion of bank ATM machines, for example (though some have introduced "convenience" variables, such as the number of branch bank offices). By separating the production of depositors' services from the activity of taking deposits, the approach in section III resolves a good part of the ambiguity in the bank production function literature about the appropriate treatment of bank deposits. In most cases in which bank deposits have been treated as an output of the bank, I believe they have implicitly stood as a proxy for the unpriced services provided to depositors.

D. The Profit Function Approach of Hancock and Fixler-Zieschang

The Hancock (1985, 1991) and Fixler-Zieschang (1990) approaches call for more extended comment.

As Hancock (1991) points out, her approach determines empirically whether specific bank assets or liabilities are inputs or outputs; it accordingly provides a criterion for their classification. Other approaches (including Baltensperger, 1980, relied on in section III) make this determination a priori. Put another way, Hancock's approach permits hypothesis testing on the signs in Baltensperger's profit function (and she rejects the negative sign placed on checking deposit accounts).

Despite this undeniable advantage, to say that a variable may sometimes be an output and sometimes an input is disquieting. Two aspects of the empirical implementation in Hancock, as well as in Fixler-Zieschang, suggest

possible bias in the estimated sign on demand deposits.

First, the bank's cost of demand deposits in Hancock's "user cost" equation consists only of direct interest payments. For the reason elaborated in section III, direct interest payments on demand deposits substantially underestimate their cost to the bank (because of the omission of payments in kind to depositors). This point was made by Offenbacher (1980) in commenting on Barnett's (1980) original formulation of the user cost of money. The omission of in-kind services should bias downward the calculation of the user cost measure for checking accounts, which would tend, therefore, to the negative (in Hancock's formulation, a variable is an output when the user cost is negative). Thus, I expect that adding the imputation developed in section III to Hancock's user cost function would alter its value in the positive direction, and make it more likely that would one find demand deposits to be financial inputs to the bank and not financial outputs.

A second point is that Hancock's model omits the financial firm's balance sheet constraint. The effect of this is much harder to work through, since Hancock's estimating system includes her profit function and six demand equations. However, simple inspection of the balance sheet constraint shows that it is unlikely to be satisfied if any sizable proportion of total deposits enter the profit function with a positive sign.

It thus appears that one of the empirical issues that so far has not been confronted--either in the work of Hancock or in the empirical implementations in the rest of the bank production function literature--concerns whether the balance sheet constraint proposed in Baltensperger's model is a valid part of the description of the financial firm.

E. Conclusion

Much of the ambiguity in the analysis of banking stems from the traditional view that banking is "different," and so it is. But so, too, does a steel mill differ from a cabbage farm. Differences do not in themselves prevent comparable analyses; the differences that exist between financial and nonfinancial firms ought not to be permitted to obscure their similarities, and therefore to obscure as well the opportunities for designing useful economic measurements.

> "There is not much difference between financial firms and food suppliers. They buy and sell credit rather than cabbages, but they shift it from producer to consumer in much the same way, sometimes selling it as it is, other times canning it or pickling it or turning it into coleslaw; they sell it wholesale to other cabbage firms, or retail in the poshest emporiums and from the humblest market stalls... Nothing very special in that...." (Economist, April 7, 1990, page S-68).

ENDNOTES

*Chief Economist, U.S. Bureau of Economic Analysis. Views are my own and do not represent official positions of either the Bureau of Economic Analysis or of the U.S. Department of Commerce.

1. In the U.N.'s System of National Accounts (SNA), "...the excess of the property income received by the banks and similar intermediaries on loans and other investments made from the deposits they hold, over the interest paid out on those deposits" (U.N., 1968, page 97). Statements in U.S. Department of Commerce (1954), Jaszi (1958), and Marimont (1969) are similar.

2. One difference between SNA and NIPA treatments is that in the former the entire banking imputation is allocated to a "dummy" financial industry so that banking output is cancelled out of GDP. In some recent SNA materials, allocation to the dummy industry has been interpreted as an imputation to borrowers, rather than depositors, though why one would have to impute interest paid by borrowers back to borrowers is one of the perplexities surrounding this whole issue. I ignore the SNA-NIPA difference for the remainder of this paper.

3. The "agent" language appears in Yntema (1947).

4. The national accounts assumption that banks act as depositors' agents is not completely without support elsewhere in economics. Bresnahan (1986), for example, assumes that competition among banks forces them to act as if they were depositors' agents. He invokes this assumption, however, to avoid the necessity for a direct measure of banking output. The assumption is less attractive as a device for actually measuring bank output. The only banks that actually function as agents are French savings banks, and they act as agents of the borrower (the government).

5. Some proposals (not adopted) would make the provision of banking services to the borrower equal to the difference between actual bank loan interest charges and some supposed "pure" rate of interest; this "service" is not finance itself, but rather some sort of markup that the bank is conceived to add to the "pure" cost of finance. See section IV.A, below.

6. In the bank literature, this is inexplicably known as the "production" approach, a terminology I have not followed.

7. This debate, and its resolution, is reminscent of the old discussion in labor economics concerning whether the number of employees or the number of hours best represented the measure of labor input; ultimately, both measures were entered independently by most researchers to give weight to the arguments on both sides. A number of parallels between labor economics issues and banking measurement issues exist and are noted below.

8. "The product of intermediation is the indirect financial asset [e.g., a bank deposit] coined from the underlying primary security [e.g., a bank loan] ..." (Gurley and Shaw, 1960, page 193).

9. "Without diminishing the importance of credit for other analytical problems in banking, it can be relegated to a supporting role when the focus is on money creation..." (Towey, 1974, pp. 58-59, emphasis supplied). Taking money creation as what banks "produce" does, however, run into the difficulty that banks do not earn revenue from deposits: "Deposits are...claims [from which] the issuing firm derives little or no direct revenue.....revenue stems primarily from the earning assets which the issue of claims enables them to obtain" (Towey, 1974, p.59). One can ask, with Warburton (1958), whether a firm's source of revenue does not indicate what it produces and sells. Textbooks emphasize that money creation is an unintended result of banking under a fractional reserve system, so holding the "focus...on money creation" is a questionable one for measuring banking output.

Pesek presents an analogy: "Out of total deposits, [a] proportion is rented: The customer [i.e., the borrower] receives demand deposit money and the bank receives, in exchange, the customer's promise to pay interest and return the loan. To view these two [i.e., deposits and loans] as separate products makes just as little sense as to say that Xerox corporation is a multi-product firm which (a) produces and maintain copying machines ("liabilities"), and (b) produces customers' promises to pay rent and return the machines ("earning assets")." Pesek, 1970, p.370. The question for Pesek's analogy is whether Xerox was engaged in producing machines or was in the rental business (or whether it had two parts, a production arm and a leasing arm). Renting equipment is, in Pesek's words, analogous to renting out deposit funds (bank loans); if the appropriate model for banking is to view its output as analogous to a renting or a leasing firm, Pesek's analogy illustrates the opposite view (that is, bank output includes loans) to the one he advocates.

10. In use of language, also, there are parallels with labor economics, where "buying a job," is sometimes used instead of selling labor, or where the supply of labor is the inverse of the demand for leisure, and for this reason is often written as if it were a negative quantity.

11. The notion that finance (in the form of loans, or in fact, equity capital) relates only to ownership claims, and that it is accordingly only a part of the legal sustem, motivates much of the national accounts treatment of interest, property income, and what is known as the "operating surplus." The implication of this view is that finance plays no real role in the productive economy.

12. Fama (1980) notes that James Tobin conjectured in 1962 that deregulation would blur or erode the distinctions between banking and other financial intermediaries, such as securities dealers.

13. On measuring the economic contribution of securities markets, see section II, above. It has long been conventional in economic measurement to treat insurance providers as if they were agents for policyholders, which implies that insurance output is the spreading of risk among the policyholders, rather than producing "insurance policies"; see, for example, U.S. Department of Commerce (1990). The agent model describes mutual insurance companies explicitly, but also conforms to the behavior of certain for-profit companies that regularly rebate favorable claims experience. The alternative view is that the "products" of an insurance company are insurance policies, which should be treated like any other products. Thus, the debate on measuring the output of insurance parallels in some ways the issues on banking output (see the review of agent models in

banking in section I).

14. Indeed, in a series of NBER financial research studies in the 1950's and 1960's which examined the "quality" of various kinds of debt, quality--though seldom explicitly defined--usually meant the probability of default.

15. Why banks should engage in barter with depositors has been addressed in the banking literature, but not resolved. Rymes (1989) argued that banks supply unpriced services to depositors because required reserves were set too high by bank regulation; Schwartz (1989) responded that Rymes' position did not accord with monetary theory, and that empirically banks whose reserves were unregulated were still observed to provide unpriced services. The analysis in the text indicates it is not above-optimal reserves that induces banks to offer "free" checks, but rather their desire to reduce their costs of reserves.

16. The vectors S and f do not appear in Baltensperger (1980).

17. This interpretation of H (•) is open to the objection that it is somewhat analogous to writing into the gadget production function the number of rolls of steel that arrive at the loading dock, and then treating separately the workers on the loading dock as real inputs to the activity "unloading steel." Though this might not be sensible for the gadget industry, financial firms and financial processes are not completely analogous to nonfinancial ones.

18. The total effect on bank profits depends not only on whether $\Sigma f_g S_g > \Sigma i_j D_j$; it depends also on whether this new pricing strategy affects the volatility of deposits, σ_x, and hence the bank's optimal reserves (through equation (3)), and ultimately (through the balance sheet constraint) the volume of earning assets that a given (average) level of deposits will permit. I am not here primarily interested in bank profit, bank behavior, or why banks provide free services, so I set aside the matter of optimal bank behavior under the conditions specified (that is, I ignore the fact that instituting explicit service prices might not, for the reasons noted, represent bank profit-maximizing behavior).

19. Available information favors the hypothesis that U.S. banks adjust schedules to equalize the cost of funds from different sources (personal conversation with David Humphrey). Even if banks do not adjust schedules (perhaps because of regulation of deposit interest), depositor adjustments, through equation (2), can bring about equilization of returns, and approximately, equilization of costs to banks.

20. This slightly underestimates the value of checking account services because the value of unpriced safe-keeping services has already reduced the interest paid on time deposits.

21. A Laspeyres quantity index, $Q(S) = \Sigma (S_{gt}/S_{gb})s_{gb}$, where b is the base year and s_g is the estimated weight (price share or cost weights) for service g, conforms to the traditional principles of national accounting, but other index number formulas have advantages, particularly the Fisher Ideal and related formulas.

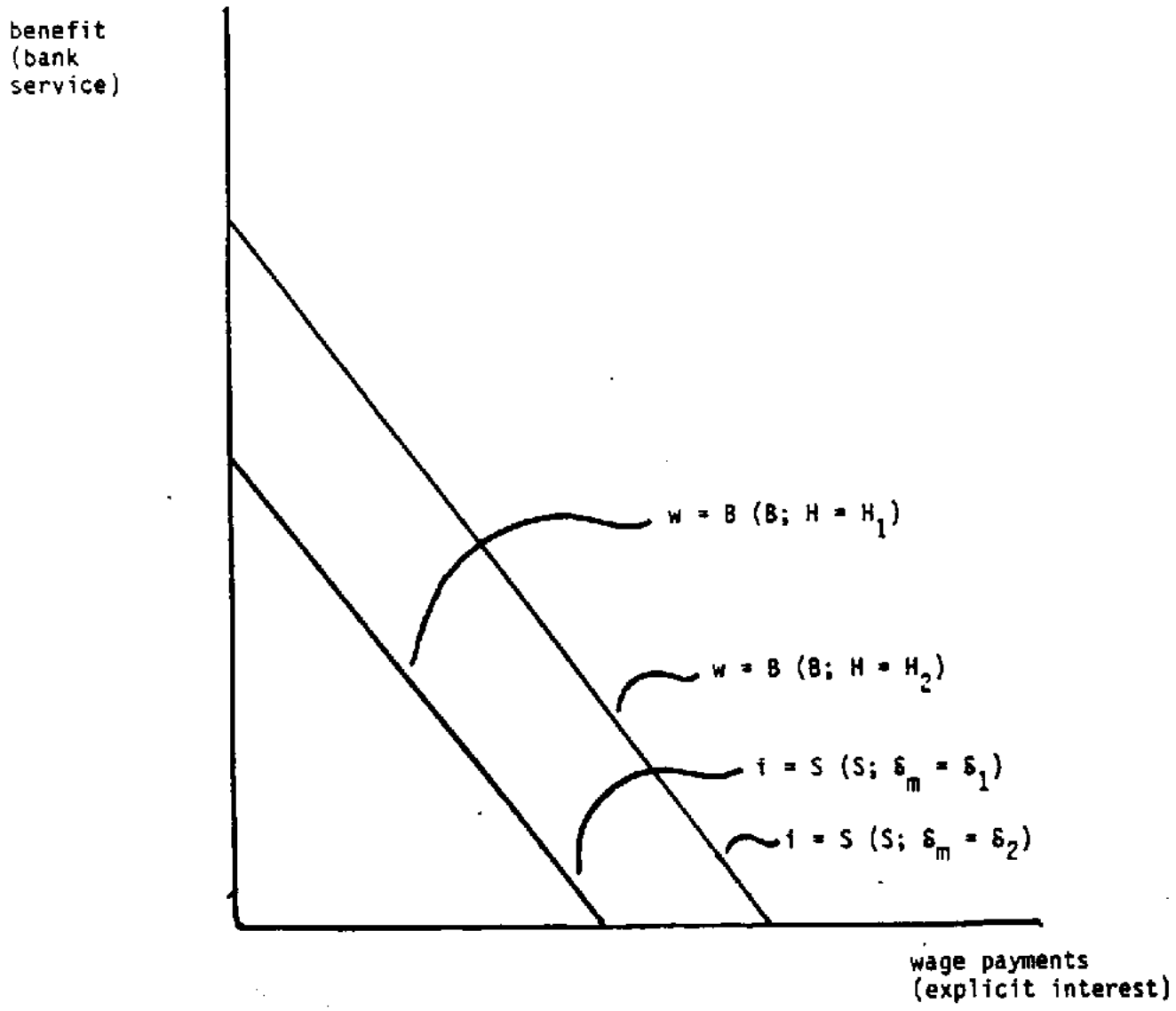
22. Another possible imputation method would employ the explicit check cashing charges some accounts contain. "Free" check cashing in accounts which have this privilege could be valued using the explicit charges where these are assessed. However, it is generally the smaller accounts where explicit check cashing charges and explicit monthly service fees appear. For a variety of reasons, the price for such services is probably not constant across all sizes of accounts. Using explicit charges from the smaller accounts would probably overstate the value of the unpriced services in the larger accounts.

23. In the SNA, separating unpriced borrower services from finance was required because the SNA insists on defining the provision of finance as nonproductive. Thus, the SNA review group thought of the "free" services provided to borrowers as economic services, but the finance itself was not, which led to a misguided (in my opinion) attempt to separate the two. Actually, the stated interest charge to the borrower often is an <u>understatement</u> (not an overstatement) of the true charge, because it is common for banks to require business borrowers to maintain "compensating" balances. I have not discussed this previously in the paper because the adjustment implied is an obvious one.

Figure 2

Wage-Benefit (Interest-Bank Service)

Trade-off Surfaces



REFERENCES

Atrostic, B.K. "The Demand for Leisure and Nonpecuniary Job Characteristics," American Economic Review, 72 (June, 1982): 428-440.

Antos, Joseph R., and Sherwin Rosen. "Discrimination in the Market for Public School Teachers. Journal of Econometrics 3 (May 1975): 123-50.

Bailey, Martin J. National Income and the Price Level: A Study in Macrotheory. New York: McGraw-Hill Book Company, Inc. (1962).

Baltensperger, Ernst. "Alternative Approaches to the Theory of the Banking Firm." Journal of Monetary Economics 6 (1980): 1-37.

Barnett, William A. "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory." Journal of Econometrics 14 (September 1980) (Annals of Applied Econometrics 1980-3: A Supplement to the Journal of Econometrics): 11-59.

Benston, George J., Gerald A. Hanweck and David B. Humphrey. "Scale Economies in Banking: A Restructuring and Reassessment." Journal of Money, Credit, and Banking 14 (November 1982, Part 1): 435-450.

Berndt, Ernst R. and Jack E. Triplett. Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth. Studies in Income and Wealth, Vol. 54. Chicago: University of Chicago Press for the National Bureau of Economic Research (1990).

Bowman, Raymond T. and Richard A. Easterlin. "The Income Side: Some Theoretical Issues." In Conference on Research in Income and Wealth. A Critique of United States Income and Product Accounts. Studies in Income and Wealth, Vol. 22. Princeton: Princeton University Press (1958).

Bresnahan, Timothy F. "Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services." American Economic Review 76 (September, 1986): 742-755.

Bresnahan, Timothy F., Paul Milgom and Jonathan Paul. "The Real Output of the Stock Exchange." Paper presented at the Conference on Output Measurement in the Services Sector, Conference on Research in Income and Wealth, Charleston, South Carolina, May, 1990.

Clark, Jeffrey A. "Economies of Scale and Scope at Depository Financial Institutions: A Review of the Literature." Economic Review (Federal Reserve Bank of Kansas City) 73 (September/October, 1988): 16-33.

Dean, Edwin R. and Kent Kunze. "Productivity Measurement in Service Industries." Paper presented at the Conference on Output Measurement in the Services Sector, Conference on Research in Income and Wealth,Charleston, South Carolina, 1990.

Fama, Eugene F. "Banking in the Theory of Finance." Journal of Monetary Economics 6 (1980): 39-57.

Fixler, Dennis J. and Kimberly D. Zieschang. "User Costs, Shadow Prices and the Real Output of Banks." Paper presented at the Conference on Output Measurement in the Services Sector, Conference on Research in Income and Wealth, Charleston, South Carolina, May, 1990.

Goldberg, S.A. "The Treatment of Interest in the National Accounts - A Review." Paper presented at the International Association for Research in Income and Wealth (August 1985).

Gorman, John A. "Alternative Measures of the Real Output and Productivity of Commercial Banks." In Victor R. Fuchs, ed. Production and Productivity in the Service Industries. Conference on Research in Income and Wealth, Studies in Income and Wealth, Volume 34. New York: National Bureau of Economic Research (1969).

Griliches, Zvi, ed. Price Indexes and Quality Change: Studies in New Methods of Measurement. Cambridge, Massachusetts: Harvard University Press (1971).

Gurley, John G. and Edward S. Shaw. Money in a Theory of Finance. Washington, D.C.: The Brookings Institution (1960).

Haig, Bryan. "The Treatment of Interest and Financial Intermediaries in the National Accounts of Australia." Paper presented at the International Association for Research in Income and Wealth (August 1985).

Hancock, Diana. "The Financial Firm: Production with Monetary and Nonmonetary Goods." Journal of Political Economy 93 (1985).

Hancock, Diana. A Theory of Production for the Financial Firm. Boston: Kluwen Academic Publishers (1991).

Hicks, John. "Capital Controversies: Ancient and Modern." American Economic Review LXIV (May, 1974): 307-316.

Hill, T.P. "On Goods and Services." Review of Income and Wealth 23 (December, 1977): 315-338.

Hodgman, Donald R. "Comment." In Victor R. Fuchs, ed. Production and Productivity in the Service Industries. Conference on Research in Income and Wealth, Studies in Income and Wealth, Volume 34. New York: National Bureau of Economic Research (1969).

Jaszi, George. "The Conceptual Basis of the Accounts: A Re-examination." In Conference on Research in Income and Wealth. A Critique of United States Income and Product Accounts. Studies in Income and Wealth, Vol. 22. Princeton: Princeton University Press (1958).

Marimont, Martin L. "Measuring Real Output for Industries Providing Services: OBE Concepts and Methods." In Victor R. Fuchs, ed. Production and Productivity in the Service Industries. Conference on Research in Income and Wealth, Studies in Income and Wealth, Volume 34. New York: National Bureau of Economic Research (1969).

Mester, Loretta J. "A Multiproduct Cost Study of Savings and Loans." The Journal of Finance XLII (June 1987): 423-445.

National Bureau of Economic Research. "The National Economic Accounts of the United States: Review, Appraisal, and Recommendations." National Accounts Review Committee, General Series, No. 64. 1958.

Niehans, Jurg, and John Hewson. "The Eurodollar Market and Monetary Theory." Journal of Money, Credit, and Banking 8 (February 1976): 1-27.

Offenbacher, Edward K. "Economic Monetary Aggregates--Comment." Journal of Econometrics 14 (September 1980) (Annals of Applied Econometrics 1980-3: A Supplement to the Journal of Econometrics): 11-59.

Organization for Economic Cooperation and Development. "Measuring the Output of Monetary Institutions in the National Account." Unpublished Paper (March, 1986).

Pesek, Boris P. "Bank's Supply Function and the Equilibrium Quantity of Money." Canadian Journal of Economics 3 (August 1970): 357-385.

Ruggles, Richard. "The United States National Income Accounts, 1947-1977: Their Conceptual Basis and Evolution." In Murray F. Foss, ed., The United States National Income Accounts: (Selected Topics. Studies in Income and Wealth, vol. 47. Chicago: University of Chicago Press (1983).

Rymes, Thomas K. "The Theory and Measurement of the Nominal Output of Banks, Sectoral Rates of Saving, and Wealth in the National Accounts." In Robert E. Lipsey and Helen Stone Tice, The Measurement of Saving, Investment, and Wealth. Studies in Income and Wealth, Volume 52. Chicago: University of Chicago Press for the National Bureau of Economic Research (1990).

Sato, Kazuo. "The Meaning and Measurement of the Real Value Added Index." Review of Economics and Statistics 58 (November 1976): 434-442.

Saving, Thomas R. "A Theory of the Money Supply with Competitive Banking." Journal of Monetary Economics 3 (1977): 289-303.

Schwartz, Anna. "Comment." In Robert E. Lipsey and Helen Stone Tice, The Measurement of Saving, Investment, and Wealth. Studies in Income and Wealth, Volume 52. Chicago: University of Chicago Press for the National Bureau of Economic Research (1990).

Sims, Christopher A. "Theoretical Basis for a Double-Deflated Index of Real Value Added." Review of Economics and Statistics 51 (November 1969): 470-471.

Smith, Robert S. and Ronald G. Ehrenberg. "Estimating Wage-Fringe Trade-Offs: Some Data Problems." In Jack E. Triplett, ed. The Measurement of Labor Cost. Conference on Research in Income and Wealth, Studies in Income and Wealth, Volume 48, pp. 347-67. Chicago: University of Chicago Press for the National Bureau of Economic Research (1983).

Stone, Richard. "Definitions and Measurement of the National Income and Related Totals." Appendix to Measurement of National Income and the Construction of Social Accounts, Report of the Subcommittee on National Income Statistics (Studies and Reports on Statistical Methods, No. 7). Geneva: United Nations, 1947.

Sunga, Preetom S. "An Alternative to the Current Treatment of Interest as Transfer in the United Nations and Canadian Systems of National Accounts." Review of Income and Wealth.

Thaler, Richard and Sherwin Rosen. "The Value of Saving a Life: Evidence From the Labor Market." In Nestor E. Terleckyj, ed. Household Production and Consumption. Conference on Research in Income and Wealth, Studies in Income and Wealth, Volume 40, pp. 265-97. New York: National Bureau of Economic Research (1976).

Towey, Richard E. "Money Creation and the Theory of the Banking Firm." The Journal of Finance 29 (March 1974): 57-72.

Triplett, Jack E. (1983). "An Essay on Labor Cost." In Jack E. Triplett, ed. The Measurement of Labor Cost. Conference on Research in Income and Wealth, Studies in Income and Wealth, Volume 48, pp.1-60. Chicago: University of Chicago Press for the National Bureau of Economic Research (1983).

Triplett, Jack E. "A Critique of the National Accounts Approaches to Banking" (in progress).

United Nations, A System of National Accounts. Studies in Methods, Series F No. 2. New York: United Nations (1968).

U.S. Department of Commerce. National Income, 1954 Edition (Supplement to the Survey of Current Business). Washington, D.C.: U.S. Government Printing Office (1954).

U.S. Department of Commerce. Bureau of Economic Analysis. Personal Consumption Expenditures. Methodology Paper Series MP-6. Washington, D.C.: U.S. Government Printing Office (June, 1990).

Warburton, Clark. "Comment." In Conference on Research in Income and Wealth, Studies in Income and Wealth 10. New York: National Bureau of Economic Research (1947).

Warburton, Clark. "Financial Intermediaries." In Conference on Research in Income and Wealth. A Critique of United States Income and Product Accounts. Studies in Income and Wealth, vol. 22. Princeton: Princeton University Press (1958).

Yntema, Dwight B. "National Income Originating in Financial Intermediaries." In Conference on Research in Income and Wealth, Studies in Income and Wealth 10. New York: National Bureau of Economic Research (1947).

Zee, Howell. "Determination of the Output of the Banking Industry in the National Accounts: A Critical Survey of Concepts." Unpublished International Monetary Fund Paper (Washington, D.C., 1981).