

Loanable Funds, Risk, and Bank Service Output^{*}

J. Christina Wang^{**}

Research Department
Federal Reserve Bank of Boston
July 2003

Abstract:

This paper develops a unified theory of bank operations that integrates theories of financial intermediation, asset pricing, and production. In a simple dynamic model, banks maximize the present value of future profits generated through three categories of qualitatively distinct functions: (1) resolving information asymmetry in order to make loans, (2) providing transaction services, and (3) financing loans with borrowed funds. Risk determines the rate of return on the funds and in turn the discount rate for future profits. But risk affects the quantity of bank services generated in the first two functions only to the extent that assets of different risk require different amounts of information processing. The model thus coherently accounts for portfolio risk in measuring bank service output. It then recognizes that only functions (1) and (2) create bank value added, whereas the borrowed funds are merely an intermediate input in the provision of bank services. Furthermore, the funds and the production function for value added are separable in a bank's optimization solution. This model can resolve some long-standing debates in the literature on bank production, such as distinguishing between the input and output roles of deposits. It also provides a theoretical basis for measuring banking output in the National Income Accounts. This banking model implies a new measure of bank output that imputes the implicitly priced services as the part of net interest income that is free of risk-related returns on loanable funds. The new measure differs significantly from the ones commonly used, suggesting a need to reexamine the conclusions of a large body of empirical literature.

Keywords: bank, service output, risk premium, value added

JEL Classification: G21, D24, O47

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

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

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

I. Introduction

The commercial banking industry has undergone sweeping changes in the past two decades.¹ In response to technological changes and the heightened competition that ensued, large-scale consolidation has taken place both within and across the traditional industry line. The merger wave has coincided with extensive deregulation, culminating in the Financial Services Modernization Act passed in November of 1999. This act overhauls the Depression-era law that separates commercial banking from investment banking and insurance companies. As a result, the convergence of different branches of the financial industry is anticipated to accelerate.

There has been extensive research studying how these structural, legislative, and regulatory changes affect bank operations, efficiency, and competition, in large part because banks play a central role in the transmission of monetary policy, provision of liquidity, and intermediation. However, there has been little consensus on a rather fundamental question: how to identify and measure bank output. This paper addresses exactly this conceptual question, which has become more important as banks integrate with other financial institutions and contribute a growing percentage of national income. A correct answer will also provide the right foundation for future studies of bank productivity and efficiency and will help to settle some long-standing debates, such as how to distinguish bank inputs from outputs—in particular, how to treat deposits.

This paper develops a simple, dynamic model of bank operation based on a framework of shareholder value maximization. The primary contribution of this model is to clarify the important relationship between the risk of bank financial assets and the output of bank services. This clarification is achieved by integrating insights from two banking literatures that have hitherto evolved mutually exclusively. On the one hand, the industrial organization literature on banking lacks a coherent model for the relationship between risk and bank services. It thus ignores risk and applies production theories to model banks as using the input of deposits to generate the output of loans (see Freixas and Rochet, 1997, chapter 3). On the other hand, the finance literature shows that the *raison d'être* of banks is to screen and monitor borrowers, resolving the asymmetric information problems that can potentially cause the credit market to break down. It then simply applies portfolio theories—pricing the risk-return trade-off—to banks' asset allocations, ignoring the actual processing of information as well as the production of transaction services using labor and physical capital. This paper aims to end the separation and to account for both risk and the processing of information and transactions, as integral components of a bank's operation. It follows theories of financial intermediation to argue that banks perform the important service of resolving information asymmetry and applies production theories to

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

model the output of bank services. Banks maximize the present value of future profits, which are generated both by bank services and by returns on the funds banks lend. Incorporating theories of asset pricing, the model shows that risk determines the rate of return on the funds and, in turn, the discount rate for future profits. But risk affects the quantity of bank services produced only to the extent that assets of different risk require different amounts of information processing.

This paper is the first to recognize the qualitative distinction between bank services and the funds they borrow and lend. Banks' fundamental role is as intermediaries in the provision of loanable funds, transferring funds from depositors and shareholders to borrowers, and then transferring the interest on the funds from the final users back to the suppliers. Hence, *only* bank services, i.e., facilitation of the allocation of funds, should be considered bank output, whereas the commodity transferred—funds *per se*—is merely an intermediate input. The funds are analogous to the merchandise wholesalers and retailers buy from manufacturers, bundle with their own services, and then sell to consumers. When borrowers use the funds to invest, returns on the funds are part of the service flow of capital and, in turn, are counted as part of the borrowers' value added. So, these returns cannot also be counted as banks' value added but are merely part of their *gross output*, even although banks collect and record such returns on their books. This conceptual distinction is obscured, especially by the hitherto common practice of lumping incomes from banking services together with returns on the associated assets.

This paper will address the fundamental issue of output measurement implied by the model. To build intuition, the paper first presents an example illustrating how conventional bank lending is in fact a bundling of two distinct functions—resolving asymmetric information problems and financing with loanable funds—and shows that neither of the existing measures of bank output is valid even in this simple framework. Then, the paper presents the formal model of bank operation. Its solution establishes, among other things, the separability of loanable funds from the production of bank value added. To complete the understanding of bank value added, this paper is also the first to complement the analysis of the supply of banking services with the derivation of the market demand for those services. The model leads to a new *flow* measure of bank output that imputes the implicitly priced services as the part of net interest income that is free of risk-related returns on the funds lent.

This paper is organized as follows. Section II classifies the multitude of banking activities parsimoniously into three categories, based on their distinct properties. Section III analyzes the market demand for each of the three categories of banking functions. Section IV specifies the respective production functions and presents a simple dynamic model of the optimal supply of these banking products. It clarifies the (separable) relation between the real activities and the financial instruments of a bank and thus provides the theoretical basis for identifying and, in turn, measuring, bank value added

and gross output. Section V reviews the old measures of bank output, spelling out the implausibly stringent assumptions needed to validate them and compares them with the new output measure. Section VI concludes.

II. Classification of Banking Functions

The lack of consensus on how to measure bank output stems largely from the lack of a coherent understanding of the relationship between the risk of bank asset portfolios and the production of bank services. Banks expend real resources (inputs of capital, labor, and materials) to produce a wide array of services and information. But these products are mostly intangible, and we observe merely the financial contracts that embody those services. Moreover, when the contracts are recorded as assets or liabilities on (and increasingly also off) the bank balance sheet, they are often reported at face values, despite their diverse risk profiles. Further difficulty is introduced by banks' common practice of offering most services "free." Instead, banks earn a net interest income by charging rates higher than those they pay. However, national income accounting treats interest income as a transfer but not as a payment for real services rendered, while regarding the revenue of a firm as its nominal gross output. This implies that only explicit fees are counted as bank output. Given the small size of fees, banks appear to produce hardly any output. This clearly counters even casual observations, and it is intuitive to conceive that banks charge implicitly for some services via interest spread. Hence, a coherent model of the relationship between risk and banking services is needed to uncover the implicit service revenue in total interest receipts and separate it from the purely risk-related return on funds. This is the purpose of the banking model here: to identify a bank's service output, clarify the relationship of the services to the risk of financial instruments, and devise a model-consistent output measure.

In order to define and measure bank output correctly, we must first thoroughly understand the functions that banks fulfill. To exemplify bank operation, I choose the set of activities traditionally referred to as banking—issuing loans through raising deposits. However, I will call this composite process lending, recognizing the considerable expansion in the scope of banking. There is no loss of generality by focusing on lending, as the model will show that the factors relevant for lending decisions are general enough to encompass many other banking activities, such as making loan commitments. In fact, it may be most useful to study the conventional lending process, which in many ways is the most confusing component of bank output, because its service revenues are implicit and bundled with the true interest income on the funds lent. Moreover, focusing on lending makes the model a single consistent framework for analyzing both historical banking activities—lending being traditionally the dominant

one²—and new activities of growing importance (e.g., securitization). They can appear quite different (e.g., on vs. off the balance sheet), but the model will show that they follow the same principles, and the model-based measure will generate comparable output data.³ Last, lending also best represents banks' special role as the intermediary for loanable funds—the financial resource needed to facilitate production.

Since the service nature of deposit taking is well established, the focus here is to demonstrate the bank output associated with the provision of funds. It has been shown that bank loans are special, compared with private debt placements by non-bank financial institutions (e.g., insurance companies) as well as public debt offerings.⁴ The well accepted explanation for the source of such uniqueness is banks' comparative advantage over other financial intermediaries in information processing and transmission. For instance, Black (1975) and Fama (1985) argue that banks have an advantage in making loans to their depositors, whose deposit histories offer valuable credit information. Others hypothesize that banks create value in issuing loans because they can reduce both the signaling cost needed for resolving the *ex ante* information asymmetry between entrepreneurs and investors (see e.g., Leland and Pyle, 1977) and the monitoring cost for mitigating the *ex post* moral hazard problem (see e.g., Diamond, 1984).⁵

The key idea common to these studies is that banks' contribution in making loans is resolving informational problems. I first illustrate this intuition through an example that underscores how the new model and its implied measure of bank output differ conceptually from the existing measures.

2.1 Identifying Bank Value Added: An Example⁶

I postulate an economy with a continuum of entrepreneurs and investors and a capital market. Each investor has one unit of endowment. Each entrepreneur has a project that needs X units ($X > 1$) of investment and will pay off and liquidate in one period. All of X has to be borrowed from outside investors. A project may be of either of two types: a High-risk type (H) pays Z_H with probability $(1-d_H)$, and 0 with probability d_H ; a Low-risk (L) project pays Z_L with probability $(1-d_L)$ and 0 otherwise.

² The recent data in Bassett and Carlson, 2002, confirm that loan issuance and deposit taking are still the most important activities for the majority of banks

³ In contrast, the book-value-based measure either simply ignores the output generated by off-balance-sheet assets, or measures it differently than similar assets on the balance sheet. For instance, Rogers (1998) measures balance sheet assets using book values but off-balance-sheet assets using revenue.

⁴ For example, Fama (1985) notices that CDs, which are subject to the cash reserve requirement, offer the same return as commercial paper (CP) and surmises that banks must be special since borrowers are willing to bear the reserve cost. James (1987) shows that bank loans do appear special, as implied by the positive response in stock prices only upon announcements of bank funding. Lummer and McConnell (1989) have similar findings.

⁵ Some other pertinent studies include Kane and Malkiel (1965), Bernanke (1983), and Goodhart (1989). See Bhattacharya and Thakor (1993) for a comprehensive review.

⁶ See Table 1 for a summary of the assumptions and solution of the example.

Assume $Z_H > Z_L$, $d_H > d_L$, and d_H and d_L are independent of the interest rate charged. Only the entrepreneur knows her type, whereas investors know only the distribution of projects: type H occurs with probability $\alpha \in (0, 1)$, and type L with probability $(1-\alpha)$. For simplicity, I assume that $\alpha = 0.5$.

Suppose that both types of projects have the same systematic component in their default risk so that they face the same risk premium.⁷ Denoting this risk premium as r^P , investors' expected rate of return on both types of projects equals $r^F + r^P \equiv r^e$, where r^F is the risk-free rate. But the actual interest rate to be charged should differ, because different default probabilities mean different default premia. If investors knew a project's type, then the interest rate to be charged for type H would be

$$i_H = (1 + r^e)/(1 - d_H) - 1 \approx r^e + d_H, \quad (\text{A1})$$

with the default premium approximately equal to d_H ; the rate for type L would be

$$i_L = (1 + r^e)/(1 - d_L) - 1 \approx r^e + d_L. \quad (\text{A2})$$

When the project type is unknown, the mean default probability is $d_M = (d_H + d_L)/2$, for $\alpha = 0.5$. The expected rate of return still equals r^e , so the actual rate to charge is $i_M \approx r^e + d_M$. Clearly, $i_H > i_M > i_L > r^e$.

Type H therefore will be happy to pay i_M to borrow funds. Assuming that type-L projects also have positive expected net present values (NPVs) at i_M , then there exists a pooling equilibrium where all projects are funded at the interest rate i_M . Type L are forced to “subsidize” type H in this equilibrium. However, the expected NPV of a type-L project may be negative at i_M . If so, we would have a market breakdown for type-L projects due to a standard “lemons” problem (Akerlof, 1970).

Now introduce into this economy a continuum of banks that possess a constant-returns-to-scale technology for precisely discerning project types. They operate in a perfectly competitive market for banking services. To focus on loan creation, I assume they are all equity-funded. Now type-L entrepreneurs have a new option: they can choose to be certified by a bank for a fee (f) and then pay the rate i_L , which is the fair price for their risk, and so the loan will be priced at face value— X .⁹ Clearly, type H entrepreneurs will not choose to be certified. Then, if a type-L project is not certified, it will have to pay an interest rate of i_H , but not i_M any more. To satisfy entrepreneurs' participation constraint, f must be no greater than the residual return to a type-L entrepreneur: $[Z_L - (1+i_L)X]/(1+i_L) \equiv f_a$; to be

⁷ This requires that the default risks of both types have the same factor sensitivities (i.e., coefficients on the market factors). One sufficient condition is that the gap between d_H and d_L is entirely idiosyncratic. This is an innocuous assumption to help highlight the contrast across different output measures.

⁸ Assume that both types of projects are viable investments, i.e., both have positive expected net present values (NPVs) at the corresponding interest rates, i.e., $Z_H > (1+i_H)X$ and $Z_L > (1+i_L)X$.

⁹ That is, a type L must promise to repay exactly X , in expected present value (PV) terms, in order to borrow X now. It is equivalent to setting a bond's coupon rate equal to the discount rate so that it will sell at par.

incentive compatible, f should not exceed the savings on interest, i.e., $f \leq (i_H - i_L)X/(1+i_L) \equiv f_b$. When $f \leq \min[f_a, f_b]$, type-L entrepreneurs will be willing to pay to be certified.

Suppose every bank charges a fee f (equal to marginal cost) that satisfies this condition, then there exists a separating equilibrium where all type-L projects will be certified, but type-H projects will not be. Investors will charge an entrepreneur i_L once she presents the certificate issued by a bank. Otherwise, they will charge i_H . In either case, they demand an expected return of $(r^F + r^P)$, and the risk premium (r^P) is their compensation for bearing the credit risk.¹⁰ One real example of such a lending arrangement is the common practice of securitizing mortgages and consumer loans, and increasingly also certain loans to small business, where banks screen borrowers and originate the loans first, and investors then supply funds directly. None of the equilibrium conditions should change if an investor also happens to be a bank shareholder. The investor should still demand r^e when lending directly to entrepreneurs, since she bears the same credit risk on the same projects. Hence, the risk-based return these investors receive should *not* be considered the value added of banks, which is just the certifying services. Assuming every loan requires the same amount of analysis, then the real quantity of bank value added should just be the number of loans certified, regardless of the loans' face value.

Now, suppose we alter the accounting procedure *without* changing any of the economics. Instead of lending directly to type-L projects and maintaining separate book entries for the loans, suppose the bank shareholders instruct each bank to dispense the funds once projects have been certified. Afterward, banks will collect interest payments from the borrowers and pass them on to the shareholders. This means the face value of type-L loans appears on a bank's balance sheet, and a matching flow of interest income appears on the bank's income statement, bundled with the certification fees; this is the same accounting as for the conventional lending of banks. The shareholders, however, should still demand r^e on those projects of the entrepreneurs, and Xr^e should be considered merely a return *transfer* from the final users of funds (i.e., entrepreneurs) to the ultimate suppliers (i.e., bank shareholders), but *not* the value added of banks. This lending procedure can be equivalently modeled as a vertical structure where banks serve as intermediaries. They buy funds from shareholders in the upstream market and then resell the funds, along with their services, to borrowers in the downstream market. As such, banks function just like wholesalers and retailers—the typical middlemen—who buy goods from producers, bundle them with services such as shipping, storage, and retail display, and deliver the goods to final consumers. In this case, it is clear that the goods are merely an *intermediate* input for retailers, and *not* their value added. The same logic applies to banks, whose value added

¹⁰ $(r^F + r^P)$ is the return demanded *ex ante*, thus equal to the realized return *on average*. The actual realized return will almost certainly deviate. Here, for simplicity, I use $(r^F + r^P)$ to approximate the *average* realized return.

should be the certifying services *only*, and not the funds *per se*. Not surprisingly, this is the same conclusion as in the setup above, where investors finance the projects directly.

This measurement convention is also consistent with national income accounting. Since borrowers invest the funds in capital, returns to funds constitute part of the service flow of capital and thus are counted toward the borrowers' value added. Hence, these returns cannot also be counted as banks' value added; otherwise there would be double counting. The funds should instead be counted as a purchased *intermediate* input for banks, and their rental cost must be subtracted from total revenue (i.e., nominal gross output) to obtain banks' value added. Once borrowers' private information is all resolved, banks should price their output package—funds plus services—according to the structure of the (downstream) market for loans. If a bank has market power in the loan market, it will charge a markup on its output. But this does *not* change how its value added is calculated, as long as it is a price taker in the upstream market for funds, since the purchase price of the intermediate input is determined solely in the funds market. As long as that market is competitive, and assuming no information problems exist between banks and shareholders, funds should be priced at marginal cost—the risk-adjusted expected return.

Modeled either way, dispensing funds through banks means the face value of type L loans appears on a bank's balance sheet, funded by the bank's shareholders. We will also see a matching flow of interest income on the bank's income statement, typically bundled with the certification fees. This is probably the intuition behind using the book value of loans as bank output in virtually all the empirical micro studies of bank production technology. But it has become clear that, even in this simple example, book values generally do not equal the quantity of bank output. Neither does the other measure that is used in the National Income Accounts. I will focus on nominal bank output to contrast the new measure with the two existing ones, given that they differ even in nominal values.¹¹ I particularly look at the case where the fees are bundled with the interest income, since it is the most challenging for devising the correct measure of bank output.

The above analysis makes it clear that the total nominal bank value added in this example is just total fees from all the projects certified, denoted by fN_L , where N_L is the number of projects certified—equal to the number of low-risk projects. When the fees are bundled with returns on the funds (viz., pure interest) and reported in the form of gross “interest” income as $[f/X + r^e]XN_L$, we can subtract the expected loan return (r^eXN_L) to recover the nominal bank value added.

In contrast, according to the book-value (BV)-based measure, bank output will be given by XN ,

¹¹ Their differences regarding the measure of nominal output can affect the choice of price index for computing the real output, but the resulting differences in deflators do not offset the problems of the existing measures in measuring the nominal output of banks.

as all the certified loans are recorded on the balance sheet.

The other measure is based on the System of National Accounts (SNA) 1993 and the latest revision (Moulton and Seskin, 2003). It uses the entire gap between a bank's gross interest receipts and the interest payments imputed at the risk-free rate—called the reference rate. It would then yield a bank output of $(fN + r^P XN)$ in this example. It overstates the true value added by $r^P XN$ —the premium investors earn for bearing the credit risk of non-bank projects.

Despite the example's simplicity, its logic for measuring bank value added through a proper division of total income remains valid in more complex and realistic situations, where banks cannot precisely assess the type of individual borrowers nor thus their default probabilities. Banks therefore have to use the loan interest rate (i_H and i_L above) as part of the sorting mechanism to induce borrowers to reveal their types and thus influence the risk of the pool, as in Stiglitz and Weiss (SW, 1981). This means banks know only the *statistical* relation between the interest rate charged (i) and the expected return (r) for the *pool* of loans, but they cannot set the interest rate of each *individual* loan according to equation (A1) or (A2).¹² Nevertheless, as long as there is a mapping from i to r , i.e., an expected return corresponding to the rate charged on a pool of loans, we can apply the same method as in the example above to partition total income and calculate bank value added. The only change is that this method now holds at the portfolio level, instead of the individual loan level.

In summary, the example highlights the conceptual problems with the two existing measures of bank output in even a simple framework. It is clear that neither the loanable funds *per se*, nor the risk premium, should be counted as bank value added.

2.2 Bank Value Added and Gross Output

The above analysis makes it clear that, if we also consider the deposit as a source of funding for loans and the services to depositors as partial payment for the funds, all the distinct functions performed in the conventional lending process can be classified into three groups—the most parsimonious categorization that is still able to capture the qualitative distinctions across the functions:¹³

1. The first group comprises activities that mitigate or resolve the asymmetric information problems associated with uncertain investment returns which could cause the capital market to break down. Also called “financial intermediation,” these activities include origination and monitoring in the

¹² For instance, $r = \int_{\underline{\theta}}^{\bar{\theta}} R(\theta, i) f(\theta) d\theta$, where $R(\theta, i)$ is the bank's mean return when charging rate i on loans to a type θ borrower, and $f(\theta)$ is the marginal probability distribution function of borrower type—two discrete ones (H and L) in the example. $E(R)$ may well be non-linear, or even non-monotonic, such as in SW (1981) when the pool of borrowers as well as the degree of moral hazard is endogenous with respect to the interest rate charged.

¹³ Banks' role in risk management and matching savers and borrowers are already accounted for implicitly in the new banking model. See Appendix A for an in-depth discussion.

lending process, tackling the *ex ante* adverse selection and the *ex post* moral hazard problems (see, e.g., Townsend, 1979), respectively. The fundamental attribute distinguishing these activities from regular services is that they have no direct value to borrowers but only a *derived* value by enabling the supply of funds.¹⁴ In this paper, I assume that banks can fully resolve any information asymmetry.¹⁵

2. This group contains the standard financial functions with symmetric information. In lending, it is the provision of loanable funds—financing capital investment by firms *other* than the bank or financing consumption by consumers—after the information asymmetry is resolved. The funds *per se* can be thought of as a special type of intermediate input, like the goods transported by shipping companies. All activities whose rewards depend solely on risk exposure—such as trading securities on exchanges and hedging through standard derivatives contracts—belong to this group.
3. The third group comprises regular services that do not involve information analysis. They are qualitatively similar to the more standard and familiar services such as transportation and storage. Most of the services are provided to depositors: processing transactions and payments (e.g., check clearing and ATM withdrawals), bookkeeping, safekeeping, etc.

Accordingly, the model divides a traditional bank into three hypothetical divisions: (A) information processing, (B) financing, and (C) (depositor) transaction services, performing functions (1) to (3), respectively. Divisions A and C consume real resources (labor, physical capital, and materials), and create the value added of a bank, whereas Division B creates no value beyond the returns that market investors should expect from providing capital to those non-bank projects. Such returns are part of the value added of the borrowers, but not that of the banks. In other words, B is merely a conduit, channeling loanable funds and their returns between the suppliers of funds and the final users. Nevertheless, B's revenue from funds—the intermediate input—should still be considered part of the (nominal) *gross* output of the entire bank, just as purchased goods are for retailers.

By unbundling the distinct banking functions, this classification overcomes one primary difficulty in measuring bank output in conventional lending. In fact, its underlying rationale is clearly manifested in the increasingly common phenomenon of securitization, which signals a trend toward specialization and hence the separation of the provision of funds from the production of services in lending. Moreover, partition of Division B's and C's functions provides a resolution to the continuing debate about the role of a deposit: is it an input or an output? The answer is both, depending on what

¹⁴ Another way to look at the differences is to note that banks *must* expend real resources to deliver a regular service (say, check cashing), but do *not* have to conduct any credit evaluation if *just* to extend funds. So, it may seem that banks conduct the information analyses for themselves, to stay in business, but not for their customers. However, without the survival of banks, borrowers eventually would not be able to obtain funding.

one means by “deposit.” If by “deposit” one refers to the loanable *funds* raised in accounts with certain attributes (say, the provision of liquidity and payment services), then deposits are an input, as the funds *per se* can be considered an input “purchased” to finance loans. When, by “deposits,” one means the set of *services* provided to the holders of those accounts, deposits should be viewed as bank outputs. Traditionally, both roles (i.e., funds and services) are integral components of a bank deposit account, and such integration is the main cause for the confusion about the role of deposits, especially since the BV-based measure uses the balance sheet value as the quantity indicator for both roles. But the two roles can be fulfilled separately. There are saving vehicles that offer little liquidity or payment services, and there are now non-bank bill-pay services that access a person’s bank account to make periodic payments. In principle, there would not have been a controversy in the first place if “deposit” had been defined clearly. I will propose an empirical separation of funds and services in section IV.

The partition also enables us to formulate the production functions of each division separately, then to study the optimal production decision of a bank as a whole, and finally to develop the correct output measure accordingly. But, first, I will examine the market demand for each of the three banking products, to provide a basis for further analysis of their optimal supply.

III. Market Demand for the Banking Products

3.1 The Market for Loanable Funds

Let me start with the market for loanable funds—the product of Division B, for it is the most thoroughly studied. Since my focus is the market for bank services, rather than asset pricing, I simply adopt the standard assumption that capital markets are competitive (i.e., all buyers and sellers are price takers) and there are no opportunities for arbitrage.¹⁶ Since investors can diversify fully by buying the market portfolio, the expected return on a security should depend only on its covariance with the systematic factors priced in the market (see models such as the CAPM, or the Arbitrage Pricing Theory (Ross, 1976)). Applied to fixed-income securities, their expected returns (r^e) are typically written as

$$r^e = r^F + r^P, \tag{1}$$

where r^F is the risk-free rate for the maturity concerned and r^P the (default) risk premium.

Expected returns (r^S) on B’s holding of securities are determined in the market exactly according to (1). For simplicity, I assume $r^S = r^F$, since banks hold mostly safe (government) debt securities. Expected returns on loans in period t (denoted by r_t^B) should also be determined by (1), since Division B finances loans—fixed-income securities—under symmetric information. Note, however,

¹⁵ One justification is that banks’ long-term relationships with customers enable them to assess accurately the relevant borrower risk characteristics (such as the mean and variance of a project’s payoff).

¹⁶ All the analysis of the market for A’s information services will remain valid as long as asset returns are determined by factors independent of those affecting the markets for bank services.

that r_t^B is *not* the interest rate to charge. Assume a constant expected default probability (d^B) of the loan portfolio and zero recovery from a defaulted borrower; then the actual interest rate to charge should be

$$i_t^B = (1 + r_t^B)/(1 - d^B) \approx r_t^B + d^B, \quad (2)$$

which is higher than r_t^B by the default premium of d^B .¹⁷ d^B is necessary to achieve an expected return of r_t^B : assuming the realized default rate $d_t^B = d^B + \xi_t$, where $E(\xi_t) = 0$ and ξ_t is a rational expectation error and thus uncorrelated with r_t^B , then expectation of the realized loan return (R_t^B) satisfies $E_s[R_t^B] = E_s[(1 - d_t^B)i_t^B - d_t^B] = r_t^B, \forall s < t$.¹⁸

The fair rate on deposits is also set by (1). Without the service component, deposits are simply fixed-income securities. Given FDIC insurance (up to \$100,000 per account), the expected rate of return should just be r^F . Without deposit insurance, depositors would demand a higher expected return, depending on the default risk of a bank's asset portfolio and its capital structure. Note that depositors' expected return on their funds should be what they demand in a capital market without transaction costs, rather than the actual net return investors receive by investing in the market. The actual return is net of charges that pay for transaction services necessary for accessing the capital market. Since providing (indirect) access is also part of the services banks supply, the return depositors expect on their funds, which is used as the reference rate for computing their implicit payment, should be the gross return.

Next, I present the first systematic model of the market demand for Division A's intermediation services, to generalize the results illustrated in the example. The model's goal is to identify the factors that affect the demand for and the pricing of A's services.

3.2 The Demand for Bank Informational Services

I model the borrower as a firm, so the analysis is most applicable to C&I loans. But the same logic applies to households' borrowing decisions (e.g., mortgages), with utility substituting for profit. The only material difference is that households generally do not have direct access to the capital market.

Consider a firm with an economically viable project that requires an initial investment of I and will yield a random stream of net income $Z \sim (\mu_z, \sigma_z^2)$ for T periods (finite mean μ_z and variance σ_z^2).¹⁹ Assume the firm needs to finance all of I with external funds, and it can choose between two perfect

¹⁷ For a more general formulation, see Grenadier and Hall (1995), who model the occurrence of default as a Poisson process with a variable intensity that may fluctuate with the business cycle. A constant d^B is equivalent to a Poisson process with a constant intensity, and it implies $r^d = 0$, since the probability of borrower default does not vary with the factors priced in the market.

¹⁸ $E_s[(1 - d_t^B)i_t^B - d_t^B] = E_s[(1 - d_t^B)(1 + i_t^B) - 1] = E_s[(1 + i_t^B)((1 - d^B) - \xi_t) - 1] = [(1 + r_t^B) - 1] + E_s[(1 + i_t^B)\xi_t] = r_t^B$.

¹⁹ A project is economically viable if it has a positive expected net present value when funded with internal funds, which are presumably free of the cost related to the asymmetric information problem.

substitutes of debt instruments: (1) a bank loan, or (2) a bond (or commercial paper) in the market.²⁰ If the firm chooses (1), banks can be modeled as the “middlemen,” purchasing the intermediate input of loanable funds in the upstream market from depositors and shareholders, combining the funds with their origination and monitoring services, and then selling in the downstream market of loans. The upstream market is assumed to be competitive and free of bank private information. So, a bank can rent funds at the marginal cost, i.e., the risk-adjusted rate of return (i^B) according to equation (2), assuming banks can resolve all the asymmetric information problems. Note again that i^B is set exclusively in the competitive *input* market, regardless whether banks can charge a markup (on every component of gross output) in the *output* market. i^B is also independent of whether banks charge for their services via explicit fees or implicitly bundled with the return on funds. By comparison, the pricing of banks’ informational services in the loan market depends not only on the competition amongst banks, but also on the price of the alternative means of financing—issuing bonds.

Let us now examine the firm’s cost of option (2). In order to raise I , the firm will set the coupon rate equal to bondholders’ required rate of return so that the bond will be priced at par.²¹ Before approaching investors for funds, the firm can choose whether to pay a cost $F(I)$ to fully resolve the problems of asymmetric information.²² Once the firm pays $F(I)$, it can set the coupon rate in accordance with the true risk of the project, and that rate is exactly i^B —the price at which banks purchase funds in the input market. Otherwise, the firm will have to pay a higher interest rate (i^N), or not fund the project at all. So, the firm’s reservation price for $F(I)$ (denoted by P^F) should be either the present value (PV) of the interest savings, or the project’s expected net present value (V^e), whichever is smaller, i.e.,

$$P^F = \min[\text{PV}(i^N, i^B) - \text{PV}(i^B, i^B), V^e], \quad (3)$$

where $\text{PV}(i_1, i_2)$ denotes the present value of interest on a bond with face value I , coupon rate i_1 , and discount rate i_2 , and so $\text{PV}(i^B, i^B) = I$.

Assuming $V^e \leq \text{PV}(i^N, i^B) - I$, i.e., it is too costly to borrow external funds directly, then the firm will issue a bond when the market equilibrium $F(I) \leq V^e$, but not if otherwise. Now consider the other funding source—bank loans. When $F(I) \leq V^e$, $(I + F(I))$ becomes the firm’s (PV) reservation price for

²⁰ To focus on close substitutes for bank financing, I ignore equity. This sacrifices little generality, since the “pecking order” theory shows that raising equity incurs a much higher informational cost than issuing bonds, and it offers no tax savings. For classic discussions, see Jensen and Meckling (1976) or Myers and Majluf (1984). Relative to bonds, bank loans may confer additional benefits such as avoiding public disclosure of project details. Such benefits decrease the effective cost of bank loans relative to bonds in the market.

²¹ Given the competitive capital market, investors will only pay the expected present value of all future interest payments. Any deviation of the bond price from I should only be an unpredictable rational expectation error.

²² Note the implicit assumption that *complete* (as opposed to partial) elimination of information asymmetry by a non-bank FI is also feasible, and is the dominant choice. Also, other factors, such as reputation, may affect $F(I)$, which typically takes the form of fees to hire rating agencies, or costs of credit enhancement or insurance.

bank loans.²³ $F(I)$ in turn sets the upper limit for how much, net of the cost of funds, a bank can charge for processing borrowers' credit information in order to make a loan—the value added of a bank by definition.²⁴ So, bank value added (denoted by f_b) is the counterpart to $F(I)$. When $F(I) \geq V^e$, the project will still get funded as long as $V^e \geq f_b$. Overall, $f_b \leq \min[F(I), V^e]$. This shows that bank loans enlarge the set of funding options for firms. Moreover, with non-bank financial institutions and a competitive capital market, banks' comparative advantage lies only in the cost savings of resolving asymmetric information, and not in the provision of funds *per se*.

Based on the discussion above of the financing for individual projects, I can now derive the demand schedule for A's services. Suppose the number of projects materialized within a period (N_t) follows an independent and identical discrete distribution (say, Poisson). Independently, the expected NPVs of all potential projects every period can be characterized by another i.i.d. distribution with c.d.f. $Z(V^e(r^F, \mathbf{R}^P))$, given the risk-free rate (r^F) and the premia of all systematic factors (\mathbf{R}^P).²⁵ I further assume that, first, for each project of a specific *type*, there exists a single value of $F(I)$ in the bond market, and, second, $F(I) > MC_A$ (A's marginal cost), i.e., banks have a cost advantage in evaluating projects. Then, the effective demand for A's services can be described by

$$\begin{aligned} Q_t &= N_t [1 - Z_t(V^e(r^F, \mathbf{R}^P))] = N_t [1 - Z_t(P_t)], \text{ for } P_t < F(I), \\ \text{and } P_t &= F(I), \quad \text{for } Q_t \leq N_t [1 - Z_t(F(I))], \end{aligned} \quad (4)$$

where Q_t is the number of projects for which bank funding would be sought at price P_t , and Z_t the sample c.d.f. of expected NPVs in period t . The thick kinked line in Figure 1 depicts one such demand schedule. The horizontal and the vertical axes denote the number of projects of a given *type* evaluated and financed by bank loans, and the price (and marginal cost) of A's services, respectively. The (unconstrained) downward-sloping curve is bent horizontally at $F(I)$, which in effect sets the price ceiling on f_b —the maximum Division A can charge for its services. The demand function makes it clear that the equilibrium quantity of Q_t depends also on the price of funds (i.e., r^F, \mathbf{R}^P), but this does not affect *how* A's service output should be measured, once given the realized number of projects evaluated by a bank. Since $\partial(V^e)/\partial E(r^F) < 0$ and $\partial(V^e)/\partial E(\mathbf{R}^P) < 0$, the entire *demand* schedule (except for the price ceiling) for A's services will decrease when interest rates or risk premia rise, *ceteris paribus*. In contrast, when a bank raises the price (up to the ceiling) of its services, there will only be a decrease in the *quantity demanded*.

²³ Otherwise, the reservation price for loans is the cost of borrowing directly from investors, subject to information asymmetry. $(I + F(I))$ is net of the tax savings on interest payments, but the comparison should reach identical conclusions if gross of the tax savings, assuming that the two debt instruments confer the same tax benefit.

²⁴ The value added consists of charges for the services, and possibly also a markup on funds if the bank has market power in the loan market. The composition, however, does not affect how the value added is calculated.

To illustrate how banks may set f_b , I depict an extreme case in Figure 1. To that end, I make three additional assumptions: (1) every bank is a monopoly in its market for Division A's services, (2) banks can perfectly price discriminate, since they can precisely discern the expected NPV of every potential project, and (3) the marginal cost (MC_A) is constant. As a result, banks will charge every borrower her reservation price: for all the projects with $V^e \geq F(I)$, Division A will set $f_b = F(I)$; for the projects with $F(I) > V^e \geq MC_A$, A will set $f_b = V^e$. In reality, such limit pricing is unlikely, since there is more than one bank in most markets, and banks can seldom estimate project NPVs precisely. The actual determination of f_b depends on the structure of competition among banks, and correct modeling can help estimate more accurately (bank-specific) price indices, and in turn quantities, of A's services.

3.3 The Demand for Depositor Transaction Services

Last, I briefly discuss the demand for Division C's services (see Appendix B for model details).²⁶ It should just be a standard demand schedule if C prices its transaction services (Y^C) explicitly, but banks traditionally offer most services to depositors without explicit charges. The model shows that this can be thought of as a special barter agreement. Moreover, the practice of offering "free" services creates implicit price discrimination, and my model is the first to recognize this special feature. This intuition rests upon the observation that the quantity of transaction services received is often not proportional to the (average) account balance, whereas depositors' implicit payment for the services (i.e., foregone interest) is proportional to deposit balance, for any given interest rate gap $\Delta r = r^F - r^D$. Thus, banks in effect price discriminate by charging implicitly, so far as a consumer's optimal account balance depends on individual-specific factors (e.g., income and preference for convenience) rather than just the quantity of services. (The model in Appendix B illustrates the optimal balance that depends on income and is thus a non-linear function of Y^C .) This suggests yet another reason why banks may prefer implicit pricing of services. The model's important implication for the measurement of Y^C is that deposit balance (D) is unlikely to be in fixed proportion to Y^C , and thus it is not a valid quantity indicator. Accordingly, neither is Δr the right price indicator. The model thus invalidates the use of deposit balances as the output measure in the value-added approach (one variant of the BV-based measure).

IV. A Model of Commercial Banks

²⁵ $Z(NPV(r^F, \mathbf{R}^p))$ can be thought of as a function of the distribution of project returns, and $Z_i(\cdot)$ is the sample formed by N_i independent draws from the distribution $Z(\cdot)$.

²⁶ Since Division C's role is to serve depositors, I consider only transaction-type deposits, which contain a considerable component of services, in all the ensuing analysis of C.

This model is the first to integrate the financial and the production aspects of banking, and it enables one to derive the general conditions under which the pricing of loans is (in)dependent of the pricing of depositor services and, in turn, deposit interest rates. In this section, I formulate the production function of each division and assemble them into the dynamic optimization problem of a bank as a whole, to derive output measures implied by the model. I also discuss how market equilibrium conditions, based on the market demand analyzed above, affect the solution of the optimization problem.

Suppose a bank only issues loans. For simplicity, I assume that it holds the loans on its books, funded only through FDIC-insured deposits. That is, uninsured external borrowing is too costly because of a high degree of asymmetric information between the bank and outside investors.²⁷ On the other hand, the bank faces a (risk-adjusted) perfectly elastic supply of tradable securities in the market. As structured, this setup is most applicable to business lending—C&I loans, which are rarely securitized because of their high degree of private information.²⁸

4.1 The Production of Bank Value Added, and Loanable Funds

It follows that Division A's value-added production function can be written as

$$\mathbf{V}^A = \mathbf{Y}^A = F^A(\mathbf{L}^A, \mathbf{K}^A, t). \quad (5)$$

That is, the amount of (a vector of different types of) information processed (\mathbf{Y}^A) in each period is a function of the labor (\mathbf{L}^A) and physical capital (\mathbf{K}^A) inputs, as well as the technology (t).²⁹ By definition, \mathbf{Y}^A is a flow variable measuring the real quantity of A's informational output.

Division B's sole purpose is to transfer loanable funds and returns between the suppliers and the users, and financial claims of the former must equal the matching obligations of the latter at all times. This identity between the sources and uses of loanable funds is here reinterpreted as a linear "production function." Then B's "input" is the *service flow* of loanable funds borrowed for a given period, while its "output" is the *service flow* of the same financial resources—now bank credit—dedicated to production (or consumption), adjusted for the systematic risk. The linear function is written as:

²⁷ Shareholders are assumed to be insiders, while deposit insurance eliminates banks' information problem with depositors. This is close to the typical situation in most banks (other than the 100 largest ones), which have about 12% non-deposit liabilities, including over 6% of very short-term, liquid obligations such as the Federal funds. A more realistic assumption that yet maintains all the conclusions is to allow banks to borrow at the fair rate such short-term market debt (e.g., securities repurchase agreements and large denomination CDs).

²⁸ However, it can be easily modified to represent loans for which banks have multiple funding choices (such as the securitization of residential mortgages and credit card obligations), basically by specifying the additional funding cost caused by bank private information.

²⁹ To focus on the distinction between A's and B's outputs, here I simply ignore the relatively minor real material inputs (\mathbf{M}^A) in A's production of services, but it should be straightforward to include \mathbf{M}^A . So, $F^A(\cdot)$ can be called A's production function of gross service output, not service value added. The same argument will also apply to real materials in the specification of C's production function below.

$$S + \sum_{j=1}^J Y_j^B = \sum_{m=1}^M (1 - k_m^*) D^m + E. \quad (6)$$

$(S + \sum_{j=1}^J Y_j^B)$ is the total fund output, where S is the security holdings, and Y_j^B the amount of the j^{th} category of loans for a given period. D^m is the m^{th} type of deposit fund input, k_m^* its ratio of cash reserves, and E the equity. Note that every variable in (6) represents a per period service *flow*, whose proxy is the real book value.³⁰ By construction, B lends in a capital market where risk-adjusted returns equalize across securities, so the equality between output and input in (6) is only in the *aggregate*, whereas B 's portfolio choice of individual Y_j^B and D^m ($j = 1, \dots, J$ and $m = 1, \dots, M$) is indeterminate. This also means that B faces identical prices in both the output and the input markets.

Last, I specify C 's production function. One component of the services transaction accounts offer is liquidity—instantaneous exchange between deposit and cash.³¹ Such convenience, as well as payment services, is supported by holding a stock of cash reserves, which serve a role analogous to materials inventory in manufacturing. Adopting the model in Ramey (1989), I will include the cash reserves in C 's production function, treating their contribution to C 's output as the service *flow* from the *stock* of “cash inventory.” With ordinary materials inputs omitted, C 's gross output (Y^C) is equivalent to its value added (V^C), and its production function can be written as:

$$V^C = Y^C = F^C(L^C, K^C, t, k^*D). \quad (7)$$

V^C (Y^C) are total services C offers to depositors. L^C , K^C , and t in $F^C(\cdot)$ are defined similarly to their counterparts in $F^A(\cdot)$ of (5). k^*D is the vector of cash reserves, where D are the volumes of different types of deposits, and k^* the respective reserve ratios. $k^* \geq k^0$, where k^0 are the required reserve ratios. Substitution is likely to exist between L , K and k^*D ; for example, a bank can operate with fewer tellers and branches when it adequately stocks its ATMs with cash.

4.2 Bank Gross Output Production Function and Separability

I now look at the bank aggregate production function of the entire lending process. A bank's aggregate value added is a composite product comprising distinct individual items of services. Like the value added of retailers, it is often, but not necessarily, attached to the rented intermediate input—funds. The analysis of the markets for the three banking functions establishes the separability between bank value added and loanable funds. That is, the gross output (Y) production function (G) can be written as $Y = G(V, M) = G(F(L, K), M)$, where V is value added, M is materials—loanable funds for banks, and

³⁰ The counterpart of (6) in *stock* values of funds is the balance sheet identity. Y_j^B should be proportional to the real book value of loans since it is the service flow of the financial resources contributed.

³¹ In terms of the nominal value of the liquidity, this treatment is consistent with its insurance nature as examined by Diamond and Dybvig (1983). But a measure of the insurance's real value is beyond the scope of this paper.

$F(\cdot)$ is the value-added production function. V and M are then said to be separable. (See Bruno, 1978.) In vector format, the gross output production function of the entire bank can be written as follows:

$$\begin{aligned} \mathbf{Y} &= [F_I^A(L_I^A, K_I^A, t), \dots, F_H^A(L_H^A, K_H^A, t), Y_I^B, \dots, Y_J^B, \\ &\quad F_I^C(L_I^C, K_I^C, t, k_1^* D_1), \dots, F_M^C(L_M^C, K_M^C, t, k_M^* D_M)]' \\ &= [(\mathbf{Y}^A, \mathbf{Y}^C), \mathbf{Y}^{B'}]' = (\mathbf{V}' \quad \mathbf{Y}^{B'})' \end{aligned} \quad (8)$$

This equation also summarizes the relationship between a bank's gross output (\mathbf{Y}) and value added (\mathbf{V}). \mathbf{V} equals the total new value of bank services created in a period, and thus is the variable to use to compute the output of the banking industry in the National Income Accounts. By comparison, \mathbf{Y} also contains the contribution from the intermediate input of loanable funds and thus measures the *complete* value a bank contributes to its borrowers and depositors.

4.3 The Overall Optimization Problem of the Bank

I now show that, under the standard objective of maximizing the market value of shareholder equity, the three divisions can be assembled to deliver a cash flow identical to that generated in the conventional bank lending practice. This model of optimal banking operation is also dynamic and can account for risk as an integral part of the bank's objective, since the discount rate used to calculate the expected present value of future profits depends on the risk of the bank's asset portfolio. In contrast, previous studies do not consider risk coherently.

I will explore only the case where each of two divisions, A and C, produces a single product, mainly because, without joint production or interactions among the markets for the elements in a vector \mathbf{Y}^A (\mathbf{Y}^C), the optimal choice of \mathbf{Y}^A (\mathbf{Y}^C) is qualitatively the same as that of a scalar Y^A (Y^C). Secondly, scalar measures of Y^A and Y^C best match the commonly available banking data, so that the model can be readily applied for empirical studies.

In the conceptual situation where the bank earns explicit revenues for its services, separate from the interest income and expense on its funds, the discrete-time bank objective function as implied by the three-division structure can be written as:

$$\begin{aligned} \underset{\mathbf{L}, \mathbf{K}}{\text{Max}} \mathbf{E}_{t_0} \sum_{t=t_0}^{\infty} \rho_t \{ & [P_t^A Y_t^A + ((1-d_t^B) i_t^B - d_t^B) Y_t^B + r_t^S S_t + P_t^C Y_t^C - (r_t^F + i^D) D_t \\ & - w_t L_t - r_t^K K_t](1-\tau)\}. \end{aligned} \quad (9)$$

Y_t^A and Y_t^C are the service output of Divisions A and C, respectively.³² $[S_t \ Y_t^B]$ is the output vector of Division B: Y_t^B is the total loan volume, and S_t is total market securities. P_t^A , P_t^C , i_t^B and r_t^S are the respective output prices. d_t^B is the realized default rate of the loan portfolio. D_t is the total volume of

³² All variables with time subscript t are realized at the end of period t .

deposits. r_t^F is the risk-free rate, and i^D is the deposit insurance premium. w_t is the wage rate, and L_t is total labor input; r_t^K is the rental rate of capital, and K_t is total physical capital. τ is the bank's income tax rate.³³ $\rho_t \equiv \prod_{s=t_0}^t (1 + R_s^E)^{-1}$ is the discount factor for the cash flow to shareholders, and R_s^E is their opportunity cost of capital for period s .

Since the extant mode of operation involves a high degree of implicit pricing (for most banking services), I instead solve the following variation of (9), and it represents the extreme case of implicit pricing—bundling all service fees with interest charges:³⁴

$$\begin{aligned} \underset{L, K}{\text{Max}} \mathbf{E}_{t_0} \sum_{t=t_0}^{\infty} \rho_t \{ & [(R_t^A + R_t^B)Y_t^B + r_t^S S_t - (r_t^F - (r_t^F - r_t^D) + i^D)D_t \\ & - w_t L_t - r_t^K K_t](1-\tau)\}, \end{aligned} \quad (10)$$

subject to the following constraints:

(i) the production functions for the three divisions:

$$Y_t^A = F^A(L_t^A, K_t^A), \quad F_L^A, F_K^A > 0, \quad F_{LL}^A, F_{KK}^A < 0 \quad (11)$$

$$Y_t^C = F^C(L_t^C, K_t^C, k^* D_t), \quad F_L^C, F_K^C, F_{kD}^C > 0, \quad F_{LL}^C, F_{KK}^C, F_{DD}^C < 0 \quad (12)$$

$$S_t + Y_t^B = (1-k_t^*) D_t + (E_t + d^B Y_t^B) \quad (13)$$

$$S_t \geq 0, \quad (14)$$

$$r_t^D = r^D(Y_t^C), \quad \text{and} \quad D_t = D(Y_t^C), \quad (15)$$

(ii) the cash reserve and capital adequacy requirements:

$$k_t^* \geq k^0, \quad \text{and} \quad e^B Y_t^B \leq E_t / e^0, \quad (16)$$

(iii) the loan “accumulation” equation:

$$Y_t^B = (1 - \delta_{t-1}) Y_{t-1}^B + \bar{A}_{t-1} Y_{t-1}^A, \quad (17)$$

$$\text{and} \quad L_t = L_t^A + L_t^C, \quad K_t = K_t^A + K_t^C. \quad (18)$$

In (10), $R_t^B \equiv (1 - d_t^B) i_t^B - d_t^B$ denotes the realized rate of return at the end of period t on the funds for loans. Division A's revenue raises the effective yield on loans by $R_t^A = P_t^A \frac{Y_t^A}{Y_t^B}$, and so

$$\frac{\partial R_t^A}{\partial Y_t^B} = P_t^A \left[\frac{\partial Y_t^A}{\partial Y_t^B} \frac{1}{Y_t^B} - \frac{Y_t^A}{(Y_t^B)^2} \right] + \left(\frac{\partial P_t^A}{\partial Y_t^A} \frac{\partial Y_t^A}{\partial Y_t^B} \right) \frac{Y_t^A}{Y_t^B}.$$

³³ I do not distinguish among different tax rates on various financial assets, e.g., municipal bonds are free of federal tax.

³⁴ Actual billing of the outputs of Divisions A and C is in between (9) and (10), and can be expressed as $Y^i = Y_1^i + Y_2^i$, $i = A, C$, where Y_1 denotes the part of output that generates explicit revenues and Y_2 the part that is implicitly charged for.

Assuming $\frac{\partial P_t^A}{\partial Y_t^A} = 0$, i.e., Division A is a price taker, then $\frac{\partial R_t^A}{\partial Y_t^B} < 0$ when $\frac{\partial Y_t^A}{\partial Y_t^B} < \frac{Y_t^A}{Y_t^B}$. One relationship between Y_t^A and Y_t^B that satisfies this inequality is Y_t^A rising at a slower rate than Y_t^B , i.e., the amount of information processing needed increases less than linearly with the balance of loans. This implies that, given the risk of a loan (i.e., given R_t^B), the larger its face value, the lower its interest rate. This is one plausible reason why large loans often have relatively low interest rates, apart from other possible reasons such as differences in bargaining power (e.g., smaller $\partial P_t^A / \partial Y_t^A$ for larger borrowers) and risk. Next, the nominal value of C's services, used to offset part of B's interest expenses, equals B's total interest savings $(r_t^F - r_t^D)D_t$, where r_t^D is the actual interest rate paid on deposits D . In sum, the terms inside the curly bracket of (10) yield the net cash flow at the end of period t , and realized default reduces both the principal and the interest of loans by a fraction of d_t^B . (Refer to Figure 2 for the time line depicting the timing of cash flows within a period.)

Since the premium (i^D) on deposit insurance is hardly sensitive to bank-specific default risk, the resulting interest savings has a present value of $V_{\text{ins}} = \mathbf{E}_{t_0} \sum_{t=t_0}^{\infty} \rho_t \{(r_t^O(i_t^B) - r_t^F - i^D)D_t(1-\tau)\}$ to bank shareholders, where r_t^O is the interest rate banks would have to pay without deposit insurance, and thus a function of i_t^B . Using r^F as the expected rate of return on D_t to impute Y^C ensures that V_{ins} introduces no bias into Y^C . However, V_{ins} can bias the imputed Y^A , since bank-specific i_t^B is not observed and likely to vary. Since the effect of deposit insurance is not central to understanding bank output, I assume that $r^O(i_t^B)$ is the same for all banks, and it is a reasonable approximation as long as no bank is overly risky.

Production functions (11) to (13) are the single-output version of (5) to (7) respectively. In Division B's production function (13), market securities S_t is to absorb any extra deposits beyond the optimal volume of loans. Thus S_t is non-negative, as in (14), and its expected return $r_t^S = r_t^F$. d^B is the loan loss allowance ratio.³⁵ Total loan loss allowances ($d^B Y_t^B$) are essentially a form of financial capital with a narrow purpose stipulated by accounting rules and banking regulations, so they will be treated equivalently to equity E_t .³⁶ The asymmetric format of (13), i.e., market securities on the asset side only but not the liability side, reflects the assumption that it is too costly for banks to issue uninsured debts in the market. Equation (15) indicates that both total deposit balance (D) and the interest rate paid (r^D) are functions of C's output, as derived in the model in Appendix B. I write them in a general functional form, since it is shown that they are most likely to be non-linear in Y^C .

³⁵ d^B is equivalent to the expectation for the probability of default. Here it is assumed to be constant: $E(d_t^B) = d^B$. In reality, d_t^B is likely to be time varying (say, over the business cycle) and autocorrelated.

³⁶ Loan loss allowances are counted toward tier-2 equity in calculating the capital-asset ratio, up to 1.25% of the risk-weighted assets. The industry average allowance/loan ratio is around 1.5%. E and $d^B Y^B$ are linked in that taking more provisions to replenish $d^B Y^B$ will reduce profit and hence the accumulation of E .

The two constraints in (16) state the regulatory requirements on cash reserves and risk-weighted equity capital, and k^0 is the minimum reserve ratio, e^B the risk weight of loans, and e^0 the minimum risk-weighted capital ratio.³⁷ Equation (17) represents the intertemporal nature of lending: loans stay on the books and require matching funds for more than one period. Hence, when originating a loan, the bank must estimate the funding need in all the future periods until maturity. Y_t^B in (17) is pre-determined, as the financing of new loans starts one period after they are originated. \bar{A}_t is the average size of loans originated in period t : $\bar{A}_t = \sum_{i=1}^{Y_t^A} A_t^i / Y_t^A$, where A_t^i is the size of the i^{th} loan, and Y_t^A the number of loans originated in period t . If $A_t^i \sim (\mu_A, \sigma_A^2)$, then the sample average $\bar{A}_t \sim (\mu_A, \sigma_A^2 / Y_t^A)$, and its mean value is exogenous. $\delta_t \in (0, 1)$ is the variable “depreciation” rate of loans—the fraction of loans maturing or in default, depending on the (pre-determined) maturity structure of the loan portfolio, and d_t^B .³⁸ Equation (18) indicates that the same kind of labor and capital are used in producing all services.

I now discuss properties of the discount factor ρ_t . It is tied to $E(r^B)$ and central to understanding the relation between risk and a bank’s optimal choice of Y^A , Y^B and Y^C .³⁹ By definition, shareholders’ claim is the residual cash flow: $R^E E = E(r^B)B - R^D D$, where B , D and E are the respective market values of a bank’s loans (ignoring securities here), debt, and equity, and $E(r^B)$, R^D and R^E are the corresponding expected rates of return.⁴⁰ The no-arbitrage condition implies $E = B - D$. So,

$$R^E = E(r^B) + (D/E)(E(r^B) - R^D). \quad (19)$$

This shows that the spread between R^E and $E(r^B)$ is the spread between $E(r^B)$ and R^D leveraged by the bank’s debt-to-equity ratio. *Ceteris paribus*, without deposit insurance, higher $E(r_s^B)$ ($t_0 < s \leq t$) only increases R_s^E (thus reducing ρ_t) but does not enhance shareholder value. Furthermore, issuing more risky loans can increase only the average pure interest earned on a bank’s asset portfolio but not its service output. Importantly, (19) reveals a conceptual error in the formulation of the existing user cost approach (Hancock, 1985 and 1991, see section 5 for details) to measuring bank output, which takes R^E to be the opportunity cost of a financial asset (or liability). It is obvious that R^E does *not* equal the opportunity cost of the asset portfolio $E(r^B)$ (let alone the required return of a particular type of loan in the portfolio), unless the bank has no debt (i.e., $D = 0$) or $E(r^B) = R^D$. Neither case is likely. As a result,

³⁷ To focus on bank output measurement, I simply assume the optimality of the observed reserve ratio. I also assign zero risk weight to all bank securities, a reasonable simplification given their high credit quality.

³⁸ To keep the model simple, I do not consider the case where A’s monitoring efforts affect d_t^B and, in turn, δ_t .

³⁹ Strictly speaking, the cash flow from A’s services may have a different risk profile than that from the pure loan interest, and the bank’s overall R^E should be the value-weighted average return on both cash flows. Here I ignore this distinction and equate R^E to only the return demanded on the cash flow of pure loan interest income.

⁴⁰ Considering tax leads to $R^E = (1-\tau)[(B/E)E(r^B) - (D/E)R^D]$, where τ is the bank income tax rate. There may also be deadweight losses due to agency problems related to debt. But neither causes any qualitative difference in the argument.

the outputs and inputs identified endogenously using that method are prone to errors.

The solution to the bank's problem defined by (10)–(18) is detailed in Appendix C. The choice variables are L_t^A , K_t^A , L_t^C , and K_t^C , and the value of S_t then follows from (13), given the initial value $Y_{t_0}^B$. One prominent feature of the solution is that, because loans require multi-period funding, the optimal amount of origination in a period (Y_t^A) depends partly on the expected shadow value of funds in future periods. In contrast, deposits are merely one-period contracts. As long as $D(Y_t^C)$ is sufficient to fund the current loan balance, only within-the-period constraints are relevant for deciding Y^C . Therefore, when available deposits are never binding, there exists a “dichotomy” between the optimal choice of Y_t^C and that of Y_t^A for all t . That is, the former is independent of the latter, as if Divisions A and C were two stand-alone entities. But A's and C's production choices become intertemporally linked when loans are constrained by the availability of funds in certain periods. The first-order conditions connecting Y_t^A and Y_{t+j}^C then show that it pays for C to solicit more deposits as long as its marginal loss is more than offset by A's marginal gain from issuing more loans. In equilibrium, A makes fewer loans in period t than if it were stand-alone, in anticipation of tight funding in certain future periods; later, in those periods, C will provide more services than otherwise optimal, in order to raise more deposits. The condition for the “dichotomy” does not depend on the market structure for A's and C's outputs. As long as deposit balance is not a constraint on the current loan balance, Y^A and Y^C are set independently; otherwise, the joint optimization of A and C will differ from the outcome that would result were the two stand-alone.

The full solution of the bank's problem also depends on how the prices of Y^A , Y^B and Y^C are set. Section 3 has shown that Division B is a price taker, whereas C prices at the marginal cost for the marginal depositor. If the market for Division A's services is assumed to be perfectly competitive, $F^A(\cdot)$ can only exhibit non-increasing returns to scale (in the relevant range). However, there is little empirical evidence of perfect competition, especially in lending to small firms and deposit taking (e.g., Berger and Hannan 1989 and Hannan 1991). Nor does the monopolistic competition of Dixit and Stiglitz (1977) seem a suitable model.⁴¹ Instead, the extensive literature on relationship banking (e.g., Allen, 1990, Ramakrishnan and Thakor, 1984, and Boot and Thakor, 1996) suggests oligopolistic competition for differentiated products in the market for A's services.⁴² Studies of depositor services find similar product differentiation. When banks can charge markups on Y^A and Y^C , they will also charge the same markup on Y^B , since the FOC is to equalize marginal cost across the margin of different

⁴¹ There may be too few banks in one (geographical) market to warrant such a model, and there may be regulatory entry barriers. In addition, preferences are unlikely to be convex with respect to (Division A's) information processing services across all banks, since each borrower is generally affiliated with just one bank.

⁴² Some plausible sources for differentiation include: convenient locations; established customer relationship, which can reduce the cost of risk assessment for the bank and search cost for the borrower; integrated financial services, which may generate synergies that enhance the value of each individual service to a customer.

inputs. But, as explained before, a markup on Y^B does not affect how Y^A and Y^C are measured. As for the mode (namely, price vs. quantity) of competition in the service markets, the discussion so far favors price competition, since most of the time banks can only announce the interest rate on loans and deposits (gross of service charges), while their number and size depend on exogenous factors such as the realized distribution of projects.⁴³ Models such as Hotelling (1929) and Salop (1979) are applicable. P^A and P^C are then functions of Y^A and Y^C , and the demands are affected by factors such as the number and relative positions of banks in a specific market (e.g., projects of a particular category, and possibly within a region as well). Given $P^A(Y^A)$ and $P^C(Y^C)$, Y^A , Y^C , Y^B , S and D are then solved similarly.

4.4 Measuring Nominal Bank Outputs: A Graphical Representation

To summarize, I present a diagram that describes intuitively how to impute a bank's service output according to the theory of bank operation developed above. Figure 3 depicts the “breakdown” of a bank's total receipt of loan interest, to identify the nominal quantity of bank value added in the case where all bank revenues are recorded as “interest income.” Of the entire gap (consisting of areas I–IV) between gross interest income and the interest payment to depositors, only the two shaded areas (I and IV) constitute the true bank value added. Area I, the part of interest income exceeding the (risk-adjusted) expected rate of return for a given type of loans, is the nominal value added of information processing services by Division A.⁴⁴ Area IV, the interest income depositors forego given deposit insurance, is the implicit value for the other value added—depositor services provided by Division C.

In contrast, the remaining two areas are returns due to the risk premium, and thus both represent pure interest transfers—part of Division B's income. Area II is the premium bank shareholders should demand for bearing the systematic risk of the loan portfolio and thus the transfer from final users of funds to bank shareholders. Area III is the value of (the put option embedded in) deposit insurance accrued to, and thus the transfer from taxpayers to, bank shareholders.⁴⁵ Only when all investors are risk neutral will these two pieces of interest transfer disappear. More importantly, their sizes vary both across banks and over time independently from the sizes of areas I and IV, because the risk premium also depends on factors irrelevant for bank value added (e.g., investors' risk aversion and legislative changes to deposit insurance).

⁴³ This holds unless a bank reaches its equity capacity and thus becomes constrained by the regulation on capital adequacy.

⁴⁴ This imputation can, on average, correctly account for bank output with regard to the common case where a bank charges only the approved but not the rejected borrowers, because the bank should take such cases into account when setting the spread. In contrast, the BV-based method underestimates output in such cases.

⁴⁵ Without deposit insurance, area III would be qualitatively the same as area IV, and the sum of III and IV would form total implicit value of depositor services.

By comparison, the BV-based approach measures bank lending services using the base of the rectangle I, and the base clearly has no definitive relationship with the area of I. The SNA93 output measures the nominal value of depositor services as area IV and that of lending services as the sum of areas I, II, and III. So, given deposit insurance, it generates the same value of depositor services as the new measure; it would understate the value otherwise. It overstates lending services as shown in Figure 3. Note, however, that Figure 3 depicts a situation where the expected rate of return on loans is greater than that on deposits, but the opposite can and does happen, usually during recessions when long-term (loan) rates fall below short-term (deposit) rates. The logic for imputing bank value added should remain the same in that case, as in the new measure, and the only change is that banks now suffer a loss on the funds, due to the different term structure of loans and deposits. The SNA93 measure, on the other hand, may generate a negative nominal value of lending services in that case.

Figure 3 makes it clear that, conditional on observing the correct expected rate of return, A's output so imputed is not distorted by deposit insurance. Nor is it biased upward by the default premium, given rational expectation of the default probability. Similarly, conditional on using the right opportunity cost of capital for depositors, the imputation of C's output as discussed above is also free of the distortion related to deposit insurance, which is captured by area III.

V. Comparison with the Existing Literature

The two existing measures of bank output reflect two different philosophies regarding the nature of bank output.⁴⁶ At the macro level, banks' gross interest income from providing funds used to be considered a pure transfer rather than payment for a real product. This mirrors the inclusion in national income of interest expenditures by non-financial firms as part of the "payment for capital." Today, it has been recognized that banks' net interest income is compensation for unpriced services other than the "direct renting of money" (Triplett, 1998, p. 10) and thus constitutes part of their value added, along with the explicit fees (see SNA, 1993).⁴⁷ In this regard, the new model can be seen as supplying the theoretical foundation for the right macro measure of bank output. In contrast, industrial organization studies of individual banks regard interest as part of income from supplying credit, and book values of the underlying assets are used as the proxy for the flow of bank output. Only a few studies have formally explored the stock-versus-flow issue of bank outputs. Pesek (1970) and Sealey

⁴⁶ Triplett (1998) discusses in detail these two measures of bank output and the potential of reconciling them.

⁴⁷ Effort has been focused on imputing and allocating the real value of uncharged-for banking services in the National Income Accounts. For example, Dean (1999) discusses the problems, while Moulton and Seskin (1999) discuss the latest improvement. Fixler and Zieschang (1991, 1998) explain the rationale behind each different national accounting system and propose the micro foundation for modification.

and Lindley (1977) rationalize the use of *stock* variables as bank outputs, arguing that the stock of loans on a bank's book is like a river rather than a stock of machinery. Two assumptions implicit in their argument are that the quantity of bank output—the (flow of) work performed to maintain the assets—is proportional to the amount of assets recorded on the books, and the ratio between the flow and the stock variables is stable over time. Humphrey (1992) empirically tests the validity of these two assumptions. He compares certain flow measures (e.g., number of loans screened) with the balance-sheet variables and finds that the assumptions do not hold. On the other hand, he also shows that, compared with his choice of flow variables, the stock variables have superior explanatory power for banks' operating costs. But this finding by no means validates the BV-based approach, because the flow variables used as the output measure ignore the quality differentials (especially those associated with information processing), and the cost variable used still contains interest expenses on funds. Given the dominating value of interest in total cost, it is not surprising that asset values should have greater explanatory power in the cost function.

Although virtually all empirical IO studies of banks use the book value of financial assets of one type or another to measure output, they are further classified into three approaches based on how they distinguish output from input, especially whether they view deposits as an input or an output of banks.⁴⁸ Among the three, only the user cost approach has an underlying micro theory. It resembles the new model on the surface, so I next contrast them in detail. The user cost approach determines the input-output status of a financial product endogenously, by the sign of the user cost of each balance sheet category. The user cost is defined as the difference between an opportunity cost and the holding revenue (cost) of an asset (liability).⁴⁹ Categories with positive user costs are treated as inputs, and those with negative user costs as outputs. Similar to the new model, this approach is based on the principle of profit maximization, but it assumes risk neutrality. It too recognizes the opportunity cost of funds, although its risk-return relation is not exactly consistent with portfolio theories (as shown in section 4.3). More importantly, Hancock (1991) uses the financial goods—securities, loans and certain deposits—but *not* the flow of services to measure bank output, even though banks are defined as

⁴⁸ The first, the asset (a.k.a. intermediation) approach, views deposits as an intermediate input in generating bank credit and sees only the financial *assets* as banks' final outputs. It thus ignores the production of depositor services, and the fees charged on deposit accounts are not counted as revenues, but are used to offset the costs of labor and physical capital. Treating deposit fees as a counter-cost can cause biases in certain estimations since these fees are likely to contain a markup. The second, the value-added (a.k.a. production) approach, treats all bank activities that incur labor or capital cost as outputs. So, all retail deposits besides the financial assets are viewed as outputs. In a sense, this method evades the question of identifying bank outputs by looking at the factor-income side of the equation instead of the product side. It also ignores the role of deposit funds as an intermediate input. See Berger and Humphrey (1992) or Bartel (1999) for further expositions.

⁴⁹ Hancock (1985) first applies the concept of the user cost of money, developed by Donovan (1978) and Barnett (1980), to define the user cost of a bank product as the difference between its opportunity cost and the holding revenue (of an asset, or holding cost if a liability). See Fixler and Zieschang (1992) for an application.

financial firms that supply intermediation services. Naturally, the unit of the financial goods is one dollar per period, and the price is the interest rate. This underscores the most fundamental distinction between the new model and the user-cost model.

The other critical distinction is that the new model distinguishes inputs from outputs according to theoretical understanding of banks' roles, independent of data, whereas the user-cost method relies on user costs calculated from data. Apart from the philosophical issue of whether there should be a data-free principle for identifying output, the user-cost method leads conceptually to indeterminacy in the case of perfect competition combined with constant returns to scale. In that case, revenue equals cost, and the financial goods associated with bank output will have a user cost of zero, causing their status to be indeterminate. Moreover, this method is by design sensitive to measurement errors in data and thus can easily result in unstable or incorrect identification of output. A number of variables, such as the expected return on funds and the increasing use of explicit fees for depositor services, can all affect the calculated user cost of a financial good and, in turn, its input-output status. For instance, capital gains and losses can change the status of a traded asset frequently between an input and an output, but it seems counterintuitive that the definition of bank output should be so unstable.⁵⁰

Yet another important difference is that the new model counts only bank services as output, whereas the user-cost model (Hancock, 1991) argues that a correct measure of bank output should be able to account for the creation of money, because one important function of the banking system is in the transmission of monetary policy. The user-cost model thus justifies the use of asset book values, which can reflect the change in bank asset holdings in response to monetary policy. But the new model makes it clear that the quantity of bank service output does not change one for one with the input of funds, which respond to policy-induced changes in interest rates. An analogy is the way a retailer's value added changes following a change in the price of the intermediate input: a toy maker's changing the wholesale price of toys will lead to a change in a toy retailer's value added only if the retailer sells more or fewer toys as a result. Besides, creating money *per se* is not an objective of bank operation but rather a side effect. Banks can maintain just a fraction of reserves because of the liquidity insurance nature of deposits.

Last, all three approaches require (mostly unstated) assumptions that are too restrictive for the stock variable to be a valid proxy of bank output. Above all, it is presumed that a *constant* stream of services is expended each period to maintain a financial good, as long as it remains on the balance sheet. This means, by definition, the BV-based measure cannot distinguish between bank value added and gross output, since they are constantly proportional to each other. This is inconsistent with the actual

⁵⁰ See also the critiques by Triplett (1992) and Berger and Humphrey (1992). The latter adopts the value-added approach.

timing of the lending process. For example, credit screening is conducted only in the beginning of a complete lending cycle, most likely before a loan is even recorded on the books. The BV-based measure also assumes homogeneity within each category of loans, both within and across banks. This amounts to adding apples and oranges, given that each category of loans as reported is broad enough to be a diverse portfolio whose composition and properties vary across banks and over time. Besides, the definition of loan categories is at the discretion of an individual bank. Using BVs also renders the measure inadequate and possibly biased nowadays, since it depends on the balance-sheet status of assets. But banks (especially large ones) are offering a growing array of products off the balance sheet (e.g., loan commitment and derivative contracts), while the book-value weight of traditional loans continues to diminish.⁵¹ Finally, within the context of the new model, the existing BV-based method can be viewed as using the period-end loan balance as a proxy to measure the output of Division B—the service flow of loanable funds. Thus, the BV-based method at best measures only one of the multiple outputs of even a simple bank. In contrast, the new measure of bank output can account for qualitatively different bank outputs without the need for stringent assumptions.

VI. Conclusions

This paper has presented a simple dynamic, optimizing model of bank operation that, for the first time, integrates insights of production theory, financial intermediation, and portfolio theories. This unified model clarifies the relationship between the risk of asset portfolios and a bank's output of services. Portfolio risk determines the rate of return on loans and banks' borrowed funds and, in turn, the discount rate used to derive the present value of future profits, part of which are generated by bank services. But the quantity of service output is affected by risk only to the extent that portfolios of different risk require different amounts of information processing. In addition, the model shows that loanable funds are merely an intermediate input that passes through banks, whereas true bank value added is only the services facilitating the provision of funds. The new model further establishes separability between the use of funds and the production functions of value added in a bank's overall optimization problem.

Partitioning total interest income according to the model, the paper develops a new measure of bank value added. Its principal innovation is to impute the value of the implicitly priced services as the residual net interest income after subtracting the required risk premium on loans. The new flow

⁵¹ Since the mid-1980s, non-interest income in all U.S. banks has risen from 26% to 40% (38% to 49% in the ten largest banks) of total income (net interest income plus non-interest income). Over the same period, 80% of non-interest income (85% in the ten largest banks) came from sources other than deposit-related charges.

measure of bank output is also shown to remain valid even when banks can only partially resolve the asymmetric information problems concerning borrowers. The new measure differs conceptually from the two existing output measures in its ability to distinguish the transfer of funds from the production of services and to account coherently for risk in the measurement of bank value added. It calls into question the results of previous empirical studies of bank production technology, all of which use one form or another of the BV-based measure of bank output. In particular, the new output measure should improve the estimates of bank productivity and the degree of returns to scale, as well as estimates of the impact of the latest merger wave on banks' operating efficiency. It can also help to estimate the markup of price over marginal cost in the pricing of bank products, which is crucial to antitrust analysis of bank mergers. For instance, when loan interest rates contain both the rate of return to funds and the compensation to bank services, assessments of markup must control for the risk of a bank's loan portfolio. In short, by resolving the fundamental question of how to measure bank output, this paper contributes to a large literature on bank production.

Moreover, this model can resolve some long-time conceptual debates in the bank production literature, particularly the one regarding the role of deposits. It demonstrates that deposit *funds* are "materials," inputs in the generation of new loans, but the transaction *services* associated with deposits are part of bank output. It also provides a theoretical basis for measuring bank output in the National Income Accounts by identifying the value-added components of a bank's gross output. Furthermore, it supplies one plausible explanation for the observation that large-denomination loans often carry relatively low interest rates.

Most importantly, since the lending process as studied in the paper involves the very same sets of issues—funds, risk, and information problems—that are also present in most other banking functions, the consistent framework developed in the model can be readily applied to analyze new banking activities, such as securitization and the outsourcing of mortgage underwriting. Using the framework developed here, we can not only identify the output produced in each activity but also explain why these new banking functions came into being and spread. The underlying economic force is exactly the fundamental separability between intermediation services and loanable funds, and the continual progress of information technology has made this separability increasingly desirable. The model can also be applied to study non-bank financial institutions, such as finance companies. Last but not least, the model's framework can easily be extended to study the implications of capital market imperfections—such as asymmetric information between banks and market investors—for the relationship between bank value added and the supply of loanable funds, and, in turn, the measure of bank output.

Table 1. Summary of the Example

		Low-Risk Project	High-Risk Project	
Characteristics	Risk-free rate		r^F	
	Investment size		X	
	Cost of capital for risky projects		$r^e (> r^F)$	
	Fraction	α	$1 - \alpha$	
	Payoff if successful	Z_L	$Z_H (> Z_L)$	
	Default rate	d_L	$d_H (> d_L)$	
	Interest rate if type known	$i_L \approx r^e + d_L$	$i_H \approx r^e + d_H$	
Equilibria without banks	1) Pooling		$d_M = \alpha d_L + (1 - \alpha)d_H,$ $i_M \approx r^e + d_M$	
	2) Market breakdown	Not invest	Invest, and pay i_H	
Separating Equilibrium with banks	Bank MC of certifying		f	
	Fees to banks	Pay f and certified.	Choose not to be certified by banks	
	Int. rate charged	i_L	i_H	
	Nominal Bank Output	New Measure	N_L — number of low-risk projects	$fN_L,$
		BV-based		XN_L
		SNA 93		$fN_L + (r^e - r^F)XN_L$

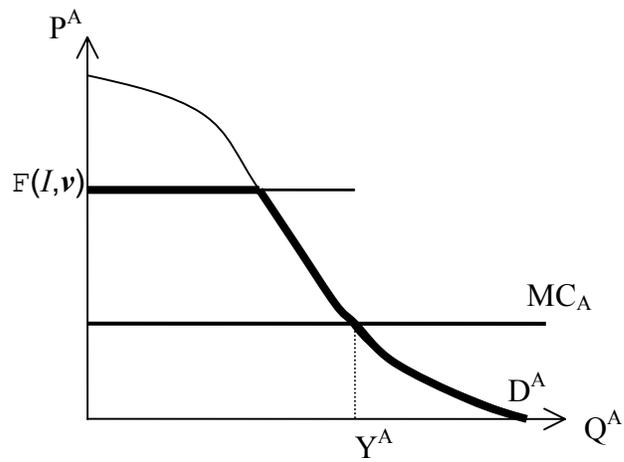


Figure 1. Demand Schedule for Bank Information-Related Services
(the Product of Division A)

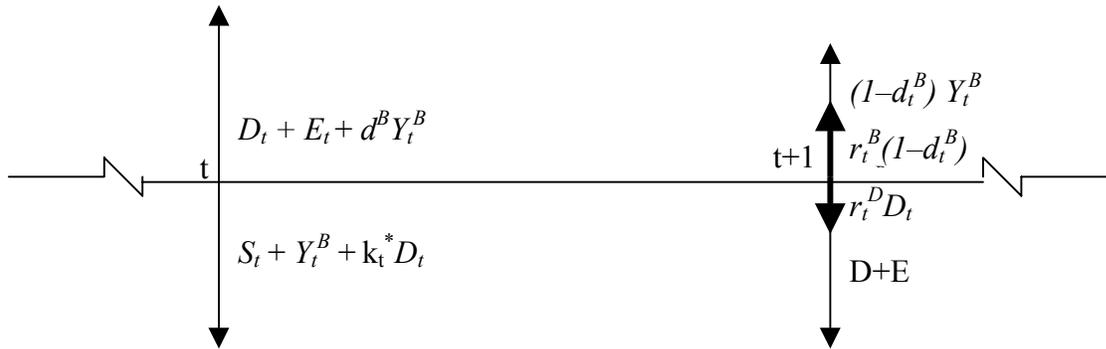


Figure 2. Time Line, and the Cash Flows (Related to Loanable Funds) within Each Period

Notes:

1. Upward arrows denote cash inflows for the bank, whereas downward arrows denote cash outflows.
2. At the beginning of each period, the prices P^A , r^B and P^C are all known, along with the expectation of default rate (d^B).
3. Realized default (d_t^B) and the related loan “depreciation” rate (δ_t) are known at the end of every period.
4. In the beginning of a period, funds flow from the depositors (and the shareholders) through the bank, to the borrowers and the market, and the inflow ($D_t + E_t + d^B Y_t^B$) exactly offsets the outflow ($S_t + Y_t^B + k_t^* D_t$). Hence, none of the terms is present in the objective function (10).
5. The amount of funds extended remains constant over the period.
6. At the end of each period, this flow of funds reverses its direction. This time, the outflow ($D_t + E_t + d^B Y_t^B$) no longer equals the inflow ($S_t + (1 - d_t^B) Y_t^B + k_t^* D_t$), and the discrepancy $d_t^B Y_t^B$ results from the loss in principal due to defaults (d_t^B). So the value of $(-d_t^B Y_t^B)$ equals the net of the receipts and payments of principal, changing the aggregate quantity of financial capital (specifically, reducing the amount of the loan loss allowances).
7. Also at the end of each period, A receives the fees for its work. Division B retrieves from only the solvent borrowers both the principal and the interest due from each loan. B forwards part of the latter to C as payment for the transaction services and then pays the promised interest and returns the appropriate principal to each of the depositors.¹

¹ A more refined model would distinguish between origination fees, received up front thus free of default loss, and A’s other incomes that were bundled into interest charges and thus subject to the same probability of borrower default. However, the simpler model here is adequate as it generates qualitatively similar results without introducing much more complexity into the solution to the optimization problem later.

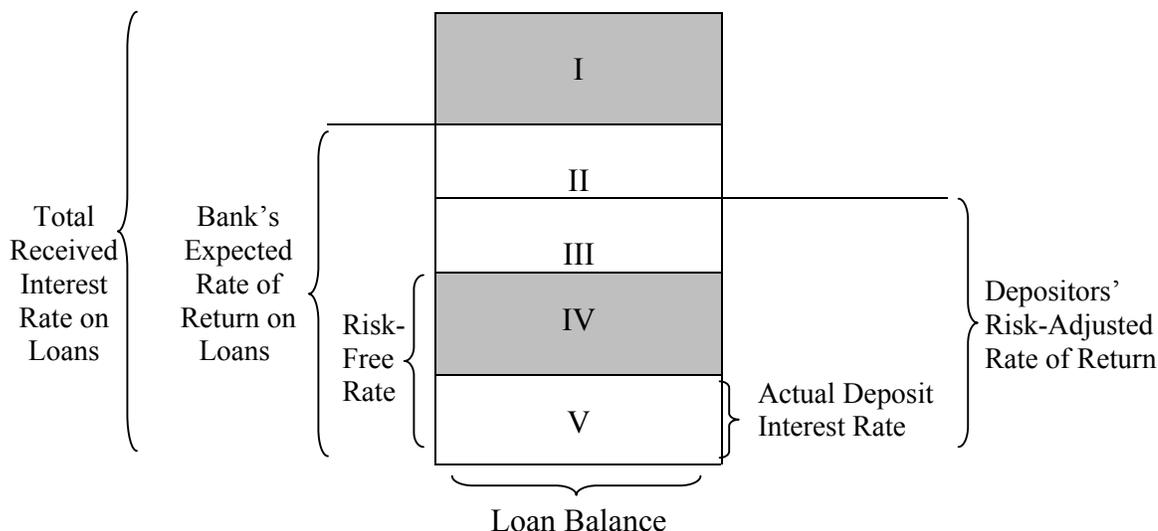


Figure 3. Decomposition of a Bank's Total Interest Receipts

Notes:

1. The content of each area:
 - Area I: implicit fees for intermediation services in lending (e.g., origination and monitoring)
 - Area II: loan risk premium
 - Area III: deposit insurance premium
 - Area IV: implicit fees for transaction and payment services (e.g., mostly to depositors)
 - Area V: deposit interest payment
 So,
 - Area (I + ... + V): a bank's total receipt of loan interest income
 - Area (II + ... + V): the bank's expected return on the funds given the loans' systematic risk
 - Area (III + IV + V): depositors' expected return on deposits given their exposure to the risk of the bank's loan portfolio (if without deposit insurance)
 - Area (IV + V): risk-free return \times deposit balance

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

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

Appendix A. Two Additional Bank Functions: Risk Management and the “Matching” Function

Banks perform two additional basic functions. One is risk management, and the other is the so-called “matching” function, which is essentially making markets. I now show that these two functions are already implicitly accounted for by the new measure of bank output, which partitions total income as described in section 4.5, even though the new model does not use the terminology explicitly.

Risk management involves activities (e.g., hedging, and holding a portfolio of market securities and cash reserves) that banks undertake to diversify their idiosyncratic risk and control their systematic risk exposure. In the financing function (i.e., Y^B), a bank’s actions do not affect the prices of systematic risk factors, as the risk-return tradeoff is determined by the market. In contrast, the actions a bank takes to manage its idiosyncratic risk can affect its return, even given the same underlying risk of operation; i.e., banks are no longer simply price takers in the capital market. This is the consequence of exactly the same set of information problems that necessitate bank intermediation. The only difference is that now the asymmetry exists between a bank and its creditors, with the bank being the party with more information, and this can lead to convex funding cost. (See, for example, Froot and Stein, 1998). This function is becoming more and more important as banks engage in increasingly sophisticated financial transactions. It is, however, not explicitly considered in the main model, primarily because it is not as central to the study of bank outputs, and this issue is moot in the model, since banks cannot issue uninsured debt. Otherwise, if modeled explicitly, managing risk can be viewed as an intermediate input produced in-house. Then, its quantity should not affect how the final output of bank services is measured. In this sense, the new measure of bank output does not preclude risk management. Its quantity is likely to be correlated with the return variance of the entire portfolio of financial assets (on and off the balance sheet). Hence, this may be the function that exhibits the most economies of scale with respect to total asset size, since better-diversified portfolios should incur less cost for risk management.

The analysis of the “matching” function was initiated by the first-generation of banking theories, which centers on banks’ cost advantage in reducing the transaction cost of bringing together savers and borrowers. But here the fundamental function banks serve is not so much matching specific buyers and sellers of loanable funds as making the market for them to meet and transact without incurring high search cost. That is, banks make markets for loanable funds. Intuitively, this function should be compensated by participants from both sides. Put in the setup of my model, Division A would receive payment for the “matching function” from the demand side, while Division C would receive payment from the supply side. So, my model encompasses this additional explanation of banking that is based on search cost, even though it is not considered explicitly. I place no emphasis on

this function in part because it is not unique to banks: qualitatively similar market-making functions are served by every type of financial institution.

Appendix B. The Market Demand for Division C’s Service Output

The traditional banking practice of offering “free” services to depositors can be modeled as a special barter agreement, where Division C sells its services through Division B to depositors indirectly. B signs a partial barter contract with the depositors, who agree to receive services as part of the remuneration on their funds, i.e., paying for the services via foregone interest.¹ B then buys the services from C in order to deliver the composite package of compensation. In contrast, banks nowadays increasingly opt for charging explicit service fees, while paying competitive interest rates for the funds. This translates into a case where Division C earns explicit revenue by selling its services directly to depositors, and Division B compensates depositors solely in monetary form—via interest payments.

The model presented below delineates how banks can achieve (close to first degree) price discrimination through the partial barter arrangement. To highlight the key features, let us look at a polar case where each account of a given type entitles the depositor to a fixed set of services.² In addition, suppose people hold all service-providing deposits solely to facilitate transactions, and each consumer holds only one account of a specific type (i.e., consumes only one unit of a given set of services). So, consumer i pays an effective price (P^i) equal to the foregone interest $P^i = (r^F - r^m) D^i \equiv \Delta r D^i$, where r^m is the actual interest rate, and r^F the safe rate. Applying logic similar to that used in the model of money demand by Baumol (1952) and Tobin (1956),³ we know that the optimal balance falls as the interest rate gap increases: $\partial D^*/\partial \Delta r < 0$, where D^* is the optimal balance. More importantly, I show that a simple variation of the model implies a positive relationship between the optimal account balance and a single person-specific factor—income: given Δr , $\partial D^*/\partial \omega > 0$, where ω is the income (or wealth).⁴ This means that people with higher incomes maintain bigger account balances (on average), and thus pay higher effective prices for the same services. It is in this sense that I say banks effect price discrimination by charging implicitly for their services. This suggests yet another reason why banks

¹ Triplett (1992) expresses a similar view.

² Account type is defined according to the terms of the services provided. Some type examples include regular checking accounts, NOW accounts, savings accounts with certain transaction features, etc.

³ Here, the transaction-type accounts are the “money.” Compared with accounts that are exclusively for savings purpose, they offer convenience, but at the same time bear an opportunity cost of Δr per currency unit.

⁴ Even if we view every withdrawal as one unit of service, we need only a minor modification of the model to obtain an outcome where the account balance increases faster than the quantity of services. The change is to make the per trip cost of going to the bank equal $b\omega$ ($b > 0$), i.e., in proportion to a depositor’s income, instead of being constant. The rationale is that if each trip takes a certain amount of time, then the opportunity cost is a linear function of one’s income (given a fixed working time). Then the ratio between average account balance and the quantity of services equals $b\omega/r$, increasing with income.

may prefer implicit pricing.

Now let us explore the relationship between the number of accounts a bank maintains (the output quantity) and total dollar balance of the accounts, as well as the actual interest payment. Let us start with the definition of a marginal depositor: for every type of deposit, if I index the account holders by a descending order of account balances, then n is a marginal depositor if $D^n \leq D^i$, for all $D^i > 0$, and $i < n$. Since $\partial D^*/\partial \omega > 0$, there will exist a non-empty subset of depositors (i 's) who maintain $D^i > D^n$, whenever $\exists i$ such that $\omega^i > \omega^n$. Given our setup, the service output vector (\mathbf{Y}^C) contains the numbers of various types of accounts; thus the number of type m accounts $Y^{Cm} = n^m$, the index of the marginal depositor. Given a distribution of ω ($f(\omega)$ and $F(\omega)$) and the function $D^{m*}(\omega, \Delta r^m)$, I can derive the distribution of individual account balances ($g(D^{m*})$ and $G(D^{m*})$ for every given Δr^m). Then, the demand function for type m services can be written as

$$Y^{Cm} = Q^m [1 - G(D^{m*}(\omega, \Delta r^m))], \quad (\text{C.1})$$

where Q^m is the total number of potential consumers, and D^{m*} is account balance of the marginal depositor, who in effect pays a price of $P^{mn} = \Delta r^m D^{m*}$. Figure B.1 depicts a simple example of a linear demand schedule, corresponding to $D^m \sim U[\bar{D}^m, \underline{D}^m]$, with \bar{D}^m finite and $\underline{D}^m = 0$. It is evident from the graph that total interest savings for the bank is the area underneath the demand curve (i.e., $\Delta r^m D^M$, where D^M is total account balance for type- m account), but not the rectangle $P^{mn} Y^{Cm}$. This means the proper service price index may well differ from bank to bank, even when they have the same interest rate differential (Δr^m). In turn, total dollar balance $D^M = D^M(Y^{Cm}) = \sum_{i=1}^n D^{mi} \neq P^{mn} Y^{Cm}$ (and $D^{mi} \geq D^{mn}, \forall i < n$) is a multiple of the area underneath the demand. It is thus a nonlinear function of Y^{Cm} too, even in this simple example.⁵ This situation resembles that of a first-degree price discrimination, except that here consumers may not pay the full reservation prices if $\Delta r^m < r^F$, i.e., $r^m > 0$.

Next, I will show that the optimal r^m banks choose endogenously will depend on $P(Y^{Cm})$, the inverse demand function. Since I have established that banks can (almost perfectly) price discriminate through implicit pricing, the optimal output quantity Y^{Cm*} should satisfy $P(Y^{Cm*}) = MC(Y^{Cm*})$, the marginal cost for supplying type- m services. So, $(r^F - r^m)D^n(r^m) = MC(Y^{Cm*})$, where $D^n(r^m)$ denotes that the marginal account balance is a function of the actual rate paid. If D^m can take on any non-negative

⁵ Specifically, given $D^m \sim U[\underline{D}^m, \bar{D}^m]$, with $\bar{D}^m > \underline{D}^m \geq 0$, then $g(D^m) = 1/(\bar{D}^m - \underline{D}^m) \equiv 1/\Delta D^m$, and $G(D^m) = (D^m - \underline{D}^m)/\Delta D^m$. Also, $\partial \bar{D}^m / \partial (\Delta r^m) < 0$, and for every given Δr^m , $\partial D^m / \partial \omega > 0$. With this uniform distribution, the demand schedule for type m accounts is $Y^{Cm} = Q^m(-D^{mn}/\Delta D^m + \bar{D}^m/\Delta D^m) = Q^m[-P^{mn}/(\Delta D^m \Delta r^m) + \bar{D}^m/\Delta D^m]$, D^{mn} being the marginal balance, and P^{mn} the marginal price. The total account balance $D^M = Y^{Cm}(\bar{D}^m + D^{mn})/2 = J(Y^{Cm})$, where $J(\cdot)$ is a non-linear function. A bank's total interest savings in turn is $\Delta r^m D^M$, obviously a non-linear function of the output quantity Y^{Cm} as well.

values, then banks have to install a minimum balance D^{MIN} , along with a zero interest rate such that $r^F D^{\text{MIN}} = MC(Y^{Cm*})$, whenever $MC(q) > 0$ for any $q > 0$. This is because banks charge implicitly for their services by lowering the interest rate paid ($r^m \leq r^F$), thus facing a natural upper bound for the equivalent price of services given that $r^m \geq 0$. This restricts banks' ability to ration supply, because they cannot raise prices without limit, as in the case where they charge for the services explicitly. More realistically, there is likely to be a positive common minimum balance D^{min} (on average) for any account to be useful in facilitating transactions. However, as long as $D^{\text{min}} < D^{\text{MIN}}$, we will still observe $r^m = 0$, and a bank-installed minimum balance.⁶ This seems to be the case for most checking accounts. The line labeled (I) in Figure B.1 depicts such a situation. When $D^{\text{min}} > D^{\text{MIN}}$, banks need to pay $r^{m*} > 0$ such that $(r^F - r^{m*}) D^{\text{min}} = MC(Y^{Cm*})$, and the marginal depositor in this case holds a balance of D^{min} . The line (II) in Figure 1.C.1 depicts this case, which roughly agrees with the structure of NOW accounts and most savings accounts.

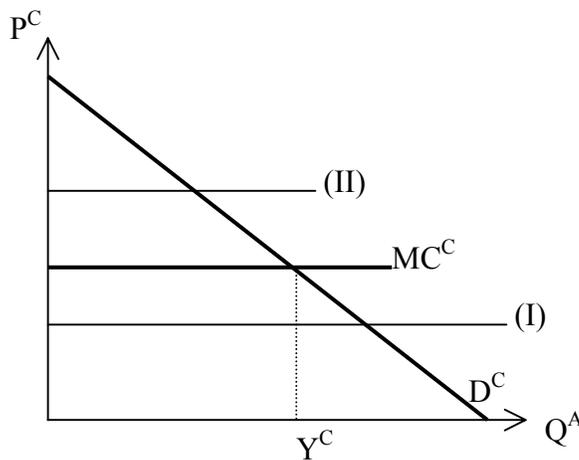


Figure B.1 Demand Schedule for Depositor (Retail Banking) Services (the product of Division C)

Appendix C. Solution of the Bank's Optimization Problem

Denote the respective Lagrangian multipliers for constraints (13), (14) and (17) as λ_t , η_t , and ϕ_t , all in present values, and substitute (11), (12), (15), and (18) into (10); then the first order conditions

⁶ All the consumers with $P^i < r^F D^{\text{MIN}}$ may still receive services by paying explicit fees.

(FOCs) for the bank's optimization problem are:⁷

$$\partial L / \partial L_t^A: \quad \mathbf{E}_{t_0} [\rho_t (P_t^A F_L^A(L_t^A, K_t^A) - w_t)(1-\tau)] + \phi_{t+1} \bar{A}_t F_L^A(.,.) = 0, \quad (\text{C.1})$$

$$\partial L / \partial K_t^A: \quad \mathbf{E}_{t_0} [\rho_t (P_t^A F_K^A(L_t^A, K_t^A) - r_t^K)(1-\tau)] + \phi_{t+1} \bar{A}_t F_K^A = 0, \quad (\text{C.2})$$

$$\partial L / \partial Y_t^B: (\phi_t) \quad \mathbf{E}_{t_0} [\rho_t R_t^B(1-\tau)] - \lambda_t + \phi_{t+1}(1-\delta_t) - \phi_t = 0, \quad (\text{C.3})$$

$$\partial L / \partial S_t: (\lambda_t) \quad \mathbf{E}_{t_0} [\rho_t r_t^F(1-\tau)] - \lambda_t + \eta_t = 0, \quad (\text{C.4})$$

$$\partial L / \partial L_t^C: \quad \mathbf{E}_{t_0} [\rho_t (-(r_t^D D_t)' F_L^C(L_t^C, K_t^C) - w_t)(1-\tau)] + \lambda_t (1-k_t^*) D_t'(Y_t^C) F_L^C = 0, \quad (\text{C.5})$$

$$\partial L / \partial K_t^C: \quad \mathbf{E}_{t_0} [\rho_t (-(r_t^D D_t)' F_K^C(L_t^C, K_t^C) - r_t^K)(1-\tau)] + \lambda_t (1-k_t^*) D_t'(Y_t^C) F_K^C = 0, \quad (\text{C.6})$$

$$\text{The complementary slackness condition is:} \quad \eta_t S_t = 0, \quad \eta_t \geq 0, \quad S_t \geq 0. \quad (\text{C.7})$$

There are two possible solutions to (C.7). The first is $S_t > 0$, so $\eta_t = 0$, i.e., the deposit balance is a non-binding constraint on the volume of loans. In this case, $\lambda_t = \mathbf{E}_{t_0} [\rho_t r_t^F(1-\tau)]$ from (C.4), meaning that the marginal value of securities holdings is the after-tax return. Once λ_t is known, and given $(r_t^D D_t)'$, (the derivative of total deposit interest payment with respect to Y_t^C), (C.5) and (C.6) combined solve for L_t^C and K_t^C . Rearranging (C.5), we get $w_t = [r_t^F(1-k_t^*) D_t'(Y_t^C) - (r_t^D D_t)'] F_L^C(.,.)$, where the righthand side is the marginal revenue product of labor, which is the net of two effects: the first term denotes the marginal return from securities funded by the extra dollar of deposit raised, and the second is the bank's interest payment on that deposit. A similar relation exists for r_t^K , and the ratio of these two equations will give us the familiar relation: $w_t / r_t^K = F_L^C(L_t^C, K_t^C) / F_K^C(L_t^C, K_t^C)$; i.e., the ratio between the two factor prices equals their respective marginal physical products. This is consistent with the separability between $F^C(.,.)$ and funds (i.e., D and Y^B), as established above. The second case is where $\eta_t > 0$, so $S_t = 0$; i.e., the optimal deposit balance if Division C were stand-alone is a binding constraint on the volume of loans.⁸ Then, $\lambda_t = \mathbf{E}_{t_0} [\rho_t r_t^F(1-\tau)] + \eta_t$, meaning that, at the corner solution, the shadow value of securities is now greater than (the PV of) the after-tax safe return. Equivalently, this case can be characterized as a downward shift of the effective marginal cost curve (of transaction services) in Figure B.1, because of the higher shadow value of the marginal dollar raised.

(C.3) is the Euler equation characterizing the intertemporal choice involved in issuing loans, and ϕ_t —the shadow value of one dollar of loan in period t —depends on $\phi_{t+j}, j > 0$:

$$\phi_t = \phi_{t+1}(1-\delta_t) + \mathbf{E}_{t_0} [\rho_t (R_t^B - r_t^F)(1-\tau)] - \eta_t$$

⁷ For brevity, I treat the deposit insurance premium (i^D) as fixed and omit it in the solution, since it is insensitive to the risk of loan portfolios. To focus on the choice of L_t and K_t , I assume that constraint (16) is always satisfied.

$$\begin{aligned}
&= \mathbf{E}_{t_0} \left\{ \sum_{j=1}^{T-1} [\rho_{t+j} (R_{t+j}^B - r_{t+j}^F)(1-\tau) - \eta_{t+j}] \prod_{i=0}^{j-1} (1-\delta_{t+i}) \right. \\
&\quad \left. + [\rho_t (R_t^B - r_t^F)(1-\tau) - \eta_t] \right\} + \lim_{T \rightarrow \infty} \phi_{t+T} \prod_{j=0}^{T-1} (1-\delta_{t+j})
\end{aligned} \tag{C.8}$$

The intuition for the first equality is that ϕ_t is the PV of the marginal loan balance in the next period (net of “depreciation”), plus its value in the current period in excess of the shadow value of securities, which depends on whether D_t is binding.⁹ The value of ϕ_t is lower (by η_t) when D_t is binding. The second equality is derived through recursive substitution. Since $1 - \delta_{t+j} \in [0, 1)$ ($j \geq 0$), the last term in (C.8) converges to zero for finite ϕ_{t+T} . Then, ϕ_t equals the PV sum of loans’ excess return over the shadow value of securities in the current and all the future periods.

Once given ϕ_t ’s, L_t^A and K_t^A can be solved from (C.1) and (C.2). Here the marginal revenue product of an input (say, labor) contains two parts: the additional service income for Division A ($= P_t^A F_L^A$) and the value of marginal loans for B ($= \phi_{t+1} \bar{A}_t F_L^A(\cdot, \cdot)$). The intuition for ϕ_{t+1} ’s role in determining Y_t^A is that loans are “durables” to the bank, so the optimal amount of origination in a period depends partly on the expectation of a loan’s shadow value in future periods. In contrast, deposits are one-period contracts. As long as $D(Y^C)$ is sufficient to fund the current loan balance, only within-the-period constraints are relevant for deciding Y^C . Once Y_t^A is solved, and given Y_t^B , Y_{t+1}^B is determined by (17).

Clearly, when $\eta_{t+j} = 0$ for all $j \geq 0$, i.e., funds are never binding, and the optimal choice of Y_t^C is independent of that of Y_t^A for all t . However, when there exist $\eta_{t+j} > 0$ for some $j > 0$, then the optimal Y_t^A will depend on all those $\eta_{t+j} > 0$, and it will thus be tied to all those Y_{t+j}^C ’s, each of which depends only on the η_{t+j} of that period, since Y^C is optimized period by period. (C.1), (C.5), and (C.8), which connect Y_t^A and Y_{t+j}^C , show that it pays for C to increase Y_t^C and thus $D(Y_t^C)$ as long as its marginal loss is more than offset by A’s marginal gain from additional Y_t^A and, in turn, future Y_j^B ($j > t$). So, A makes fewer loans in period t than if it were stand-alone, in anticipation of tight funding in certain future periods; later, in those periods, C will provide more services than otherwise optimal, in order to raise more deposits.

⁸ It is assumed that $Y^{C*} < \bar{Y}^C$, where Y^{C*} is the optimal choice of Y^C , and \bar{Y}^C is the upper limit that may be imposed by external factors such as limited local demand for cash-equivalent liquid assets. Otherwise, the solution of Y^{C*} would be trivial.

⁹ Without the subsidy owing to deposit insurance, ρ_t is set exactly to offset the return differential between loans and deposits so that the second term yields the PV of bank shareholders’ equity (per dollar of loan) that is invariable with respect to loan risk. However, with deposit insurance, it pays to issue more risky loans.

Bibliography

- Akerlof, G. (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84(3), p. 488-500.
- Allen, F. (1990). "The Market for Information and the Origin of Financial Intermediation," *Journal of Financial Intermediation* 1(1), p. 3-30.
- Barnett, W. A. (1980). "Economic Monetary Aggregates," *Journal of Econometrics* 14(1), p. 11-48.
- Bartel, A. P. (1999). "The Performance of Retail Bank Branches: Does Human Resource Management Play A Role?" *NBER Summer Institute*.
- Bassett, W. F. and M. Carlson (2002). "Profits and Balance Sheet Developments at U.S. Commercial Banks in 2001," *Federal Reserve Bulletin*, p. 259-288.
- Basu, S. and J. G. Fernald (1995). "Are Apparent Productive Externalities a Figment of Specification Error?" *Journal of Monetary Economics* 36(1), p. 165-188.
- Baumol, W. J. (1952). "The Transaction Demand for Cash: an Inventory-Theoretic Approach," *Quarterly Journal of Economics* 66(4), p. 545-56.
- Berger, A. N. and T. H. Hannan (1989). "The Price-Concentration Relationship in Banking," *Review of Economics and Statistics* 71(2), p. 291-99.
- Berger, A. N. and D. B. Humphrey (1992). "Measurement and Efficiency Issues in Commercial Banking," in Z. Griliches (eds.), *Studies in Income and Wealth*, vol. 56, p. 245-79, NBER.
- Berger, A. N. and L. J. Mester (1997). "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance* 21(7), p. 895-947.
- Bernanke, B. (1983). "Non-Monetary Effects of the Financial Crisis in the Propagation of the Great Depression," *American Economic Review* 73(1), p. 257-76.
- Bhattacharya, S. and A. Thakor (1993). "Contemporary Banking Theory," *Journal of Financial Intermediation* 3(1), p. 2-50.
- Black, F. (1975). "Bank Fund Management in an Efficient Market," *Journal of Financial Economics* 2(2), p. 323-39.
- Boot, A. W. A. and A. V. Thakor (1996). "Can Relationship Banking Survive Competition?" Working Paper 96-6, University of Michigan Business School.
- Bruno, M (1978). "Duality, Intermediate Inputs, and Value Added," in M. Fuss and D. McFadden, eds., *Production Economics: A Dual Approach to Theory and Applications*, vol. 2, New York, North-Holland.
- Dean, E. R. (1999). "The Accuracy of the BLS Productivity Measures," *Monthly Labor Review* 122(2), p. 24-34.
- Diamond, D. W. (1984). "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies* 51(3), p. 393-414.
- Diamond, D. W. and P. H. Dybvig (1983). "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* 91(3), p. 401-419.
- Dixit, A. K. and Stiglitz J. E. (1977). "Monopolistic Competition and Optimum Product Diversity," *American Economic Review* 67(3), p. 297-308.
- Donovan, D. J. (1978). "Modeling the Demand for Liquid Assets: An Application to Canada," *International Monetary Fund Staff Papers* 25(4), p. 676-704.
- Fama, E. F. (1985). "What's Different about Banks?" *Journal of Monetary Economics* 15(1), p. 29-36. *Federal Reserve Bulletin*: various issues.
- Fixler, D. J. and K. D. Zieschang (1991). "Measuring the Nominal Value of Financial Services in the National Income accounts," *Economic Inquiry* 29(1), p. 53-68.
- Fixler, D. J. and K. D. Zieschang (1992). "User Costs, Shadow Prices, and the Real Output of Banks," in Z. Griliches (eds.), *Studies in Income and Wealth*, vol. 56, NBER.

- Fixler, D. J. and K. Zieschang (1998). "The Productivity of the Banking Sector: Integrating Financial and Production Approaches to Measuring Financial Services Output," *BLS Working Paper #307*.
- Freixas, X. and J.-C. Rochet (1997). *Microeconomics of Banking*, MIT Press, Cambridge and London.
- Froot, K. A. and J. C. Stein (1998). "Risk Management, Capital Budgeting and Capital Structure Policy for Financial Institutions: An Integrated Approach," *Journal of Financial Economics* 47(1), p. 55-82.
- Goodhart, C. A. E. (1989). *Money, information and uncertainty*, MIT Press, Cambridge, MA.
- Grenadier, S. R. and B. J. Hall (1995). "Risk-Based Capital Standards and the Riskiness of Bank Portfolios: Credit and Factor Risks," *NBER Working Paper #5178*.
- Hancock, D. (1985). "The Financial Firm: Production with Monetary and Nonmonetary Goods," *Journal of Political Economy* 93(5), p. 859-80.
- Hancock, D. (1991). *A theory of production for the financial firm*, Norwell, Mass. and Dordrecht: Kluwer Academic, 1991.
- Hannan, T. (1991). "Bank Commercial Loan Markets and the Role of Market Structure: Evidence from Surveys of Commercial Lending," *Journal of Banking and Finance* 15(1), p. 133-49.
- Humphrey, D. B. (1992). "Flow Versus Stock Indicators of Banking Output: Effects on Productivity and Scale Economy Measurement," *Journal of Financial Service Research* 6(2), p. 115-35.
- James, C. (1987). "Some Evidence on the Uniqueness of Bank Loans," *Journal of Financial Economics* 19(2), p. 217-35.
- Jensen, M. C. and W. H. Meckling (1976). "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics* 3(4), p. 305-60.
- Kane, E. and B. Malkiel (1965). "Bank Portfolio Allocation, Deposit Variability and the Availability Doctrine," *Quarterly Journal of Economics* 79(1), p. 113-134.
- Leland, H. E. and D. H. Pyle (1977). "Informational Asymmetries, Financial Structure, and Financial Intermediation," *Journal of Finance* 32(2), p. 371-87.
- Lummer, S. and J. McConnell (1989). "Further Evidence on the Bank Lending Process and the Reaction of the Capital Market to Bank Loan Agreements," *Journal of Financial Economics* 25(1), p. 99-122.
- Moulton, B. R. and E. P. Seskin (1999). "A Preview of the 1999 Comprehensive Revision of the National Income and Product Accounts: Statistical Changes," *Survey of Current Business* 79(10), p. 6-17.
- Myers, S. C. and N. S. Majluf (1984). "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics* 13(2), p. 187-221.
- Pesek, B. P. (1970). "Bank's Supply Function and the Equilibrium Quantity of Money," *Canadian Journal of Economics* 3(3), p. 357-85.
- Ramakrishnan, R. and A. V. Thakor (1984). "Information Reliability and A Theory of Financial Intermediation," *Review of Economic Studies* 51(3), p. 415-432.
- Ramey, V. A. (1989). "Inventories as Factors of Production and Economic Fluctuations," *American Economic Review* 79(3), p. 338-54.
- Rogers, K. E. (1998). "Nontraditional Activities and the Efficiency of US Commercial Banks," *Journal of Banking and Finance* 22(4), p. 467-82.
- Roll, R. (1977). "A Critique of the Asset Pricing Theory's Tests: Part I: On Past and Potential Testability of the Theory," *Journal of Financial Economics* 4(2), p. 129-76.
- Ross, S. A. (1976). "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* 13(3), p. 341-360.
- Sealey, C. W. and J. T. Lindley (1977). "Inputs, Outputs and a Theory of Production and Cost at Depository Financial Institutions," *Journal of Finance* 32(4), p. 1251-1266.
- Stiglitz, J. E. and A. Weiss (1981). "Credit Rationing in Markets with Imperfect Information," *American Economic Review* 71(6), p. 393-410.
- Tobin J. (1956). "The Interest Elasticity of the Transactions Demand for Money," *Review of Economics and Statistics* 38(3), p. 241-47.

- Townsend, R. M. (1979). "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory* 21(1), p. 265-293.
- Triplett, J. E. (1992). "Comment: Measurement and Efficiency Issues in Commercial Banking," in Z. Griliches (eds.), *Studies in Income and Wealth*, vol. 56, p. 287-96, NBER.
- Triplett, J. E. (1998). "Concepts of Banking Output: What Do Banks Do?" *Brookings Workshop on Measuring Banking Output*.