MANAGERIAL AND DECISION ECONOMICS, VOL. 13, 183–200 1992

A Service-based Theory of Retail Banking

Ellen E. Hanak

University of Maryland, College Park, MD, USA

INTRODUCTION

It is a well-accepted proposition of economic theory that financial intermediaries owe their existence to the presence of imperfections in the markets for the transfer of positive and negative assets among economic agents. In particular, costs arising from asset indivisibilities and from the acquisition of information on potential borrowers and lenders pose obstacles to the conduct of individual agents in these markets (Santomero, 1984). Successful intermediaries add value to the economy by providing cost-reducing services to individuals. For bank-type intermediaries that pool funds to produce new financial assets,[1] this feat is achieved by exploiting scale economies in the gathering and processing of information, and by combining existing assets so as to create divisibility and reduce the influence of their specific risk on the expected returns of the new assets.

Yet the development of a comprehensive firm-theoretic approach to the bank's decision-making problem has proceeded with insufficient grounding in this essential service-based reason for intermediaries' existence. Bank outputs have been defined in manners inconsistent with the basis for their demand by consumers; banking technology has been simplified to the point of rendering operating costs irrelevant; and the structure of banking markets has been specified in alternative forms without clear regard to theoretical underpinnings. The consequence has been a theory whose general relevance is open to doubt. As Santomero noted in his 1984 review:

> . . . [It] applies Economics 1 to the bank to obtain results that border on the trivial in restrospect, that is, marginal revenue equals marginal cost. It depends, to an excessive extent, on demand and supply curve slopes that are not well motivated or understood (p. 590).

This paper develops a formal model of banking, the centerpoint of which is the interaction of supply and demand for the range of intermediation services banks are able to offer their customers. Specifically, banks provide the explicit service of intermediating a loan or deposit account, in conjunction with a host of implicit services that enhance the quality of that account, such as accessibility, information, and availability of an assortment of other explicit services. The focus on the service-based nature of intermediation establishes a sound basis for applying the theory of the firm to banking, and generates a number of important insights into the operation of the markets for financial services.

Because it explicitly derives the demand for financial services, the framework opens up new areas of research on the optimal combinations of banking services and their relationship to output pricing. Most banking products are shown to be gross complements with each other, rather than substitutes as is typically assumed in the literature. Most banking services are also shown to substitute for the own time input of customer households into the achievement of intermediation objectives. Using wage rates as a measure of the opportunity cost of time, the model predicts variations in service levels across different types of communities. An intuitively appealing result is that as consumers' opportunity cost of time increases, so do banks' incentives to create highly service-oriented financial supermarkets. Yet somewhat surprisingly, these are the conditions that lead to downward pressure on pricing of financial services; lower wage areas may experience both lower services levels and higher prices.

Other results of the model shed new light on past ambiguities or controversies in the literature. For one, the common practice of specifying the bank's cost function as separable (Towey, 1974; Sealey and Lindley, 1977; Startz, 1979; Mitchell, 1979) is shown to be theoretically invalid once the production of

0143–6570/92/030183–18$14.00

services is properly incorporated. Moreover, past empirical work that has attempted to measure cost complementarity between explicit bank products has been biased against detecting it, by the failure to consider separately the cost of increasing assortment (an implicit service of the bank).

The recognition of demand and supply interdependencies among bank products provides new evidence to the long-standing controversy about the conditions under which banks' deposit and loan pricing decisions are made separately (Klein, 1971; Pringle, 1973; Baltensperger, 1980; Langohr, 1982; Santomero, 1984). Even if risk considerations that might link the two sides of the balance sheet are ignored, they are linked by the nature of the technology and the demand functions facing the bank.

In two respects, the presumption of a perfectly competitive paradigm in banking is rejected on theoretical grounds. One concerns the determination of bank fees for depository services. A number of authors assume fees would be at levels sufficient to cover the marginal costs of service provision in the absence of binding interest rate ceilings on bank deposits (Klein, 1974; Black, 1975; Startz, 1979; Mitchell, 1979, 1988; Merris, 1985; Fama, 1980, 1985). This framework shows that, in general, banks will continue to pay 'implicit interest' in the form of service fees below marginal costs even when rates are unregulated. The infeasibility of explicitly charging for many services that are consumed jointly by deposit customers leads banks to recover large portions of their costs through interest rate margins.

The second concerns the broader question of the market structure in which banks operate. Until now, the debate on whether banks face perfectly or imperfectly elastic demands for their output has hinged on the presumption of absence or existence of barriers to contestability. This model demonstrates that the nature of competition in banking markets is, *at most*, monopolistic: the multiple possibilities for combination of implicit services with any given deposit or loan category will, in general, prevent a convergence of service levels and of output pricing within this industry. Factors limiting the contestability of banking, such as entry or exit costs, can generate economic rents to firms whose product diversity characteristics already permit them to set prices.

Since the demand side is derived for household—as distinct from corporate—customers of banks, the model is specifically addressed to retail banking, as reflected in the title of the paper. This does not mean the conclusions are less general than those from earlier firm-theoretic banking models. Prior studies have generally specified demand functions without concern for the distinction between household and firm clients. Indications of how this framework could be extended to incorporate formally the demand of corporate clients are provided briefly in the concluding section.

The paper is organized as follows. The next section provides an overview of the analytical framework. Because the approach used is new to banking, the logic behind the model is explained in some detail in non-technical fashion. The third and fourth sections formalize the demand and supply sides of the model, respectively. The fifth section presents the results of the profit maximization problem of the firm. The main implications of the framework are presented in each of the sections as relevant. The final section presents a brief conclusion, noting valuable avenues for future research.

OVERVIEW OF THE ANALYTICAL FRAMEWORK

Model of the Retail Firm

The model developed here draws on a recent literature on the economics of retailing by Betancourt and Gautschi (B&G) (1986, 1988, 1989, 1990a, b). This framework was designed to explain the decision-making problem of a firm whose outputs consist of some explicit products (the retail items) along with a set of implicit products (distribution services) that are not marketed separately, but purchased in a bundle together with the explicit product. The framework's major innovation is the way in which it captures the demand for and supply of these services. Households are recognized to incur a range of 'distribution costs' in their interactions with the retail system. B&G have identified six broad cost categories: *direct time and transportation* costs, *adjustment* costs incurred as a result of unavailability of products at the desired time or in the desired form, costs incurred from *storage* and acquisition of *information*, and *psychic* costs arising from unappealing characteristics of the retail environment.[2]

Over a certain range, firms can provide distribution services that directly reduce these costs to households. The five broad service categories are:[3]

accessibility and *assortment of retail items* (both reducing direct time and transportation costs), *assurance of product availability* in the desired time and form (reducing adjustment costs, including higher expenditures for substitutes; indirect time, transport and information costs to search for alternatives; and storage costs if consumers are forced to hold larger quantities than desired), *information* (reducing various costs associated with information gathering about product characteristics), and *ambience* (reducing psychic costs of transacting). Since households' own costs are reduced by the presence of these services, they are willing to pay a higher price for explicit products marketed in conjunction with higher service levels. Also, since firms cannot provide the services without incurring additional operating expenses, they generally will need to receive higher prices for the explicit product in order to raise implicit service levels.

Formally, the demand side is specified as a household production problem. Distribution costs enter the household's utility-maximization problem as the inputs of own time and other 'market goods' needed to produce utility-yielding 'commodities'. Distribution services provided by the firm appear to the household as fixed inputs into household production, reducing household costs of producing any given level of commodities. The services enter the supply side as cost-increasing outputs. The firm's decision, then, is to choose the profit-maximizing levels of distribution services and either the explicit products or their prices, given its own cost function and households' price and service elasticities of demand.

Application to Financial Intermediation

Although the examples presented by Betancourt and Gautschi are drawn from retail merchandizing, their framework is particularly illuminating for any industry wherein an important component of costs is devoted to the provision of distribution services. It is ideal for financial intermediation, where the very reason for the industry's existence is its ability to lower the distribution costs facing individual agents. Moreover, the specification of demand as one derived from household production is appropriate to this industry. Agents purchase financial services strictly according to how well the intermediated funds can be used to smooth inter- and intra-temporal consumption of commodities (Benston and Smith, 1976).

To distinguish among the types of costs and services involved it is useful to consider briefly what takes place in the intermediation process. If a household wishes to transfer assets or liabilities without recourse to an intermediary it will incur the following types of distribution costs. Very large time and transportation resources will need to be devoted to acquisition of information to locate a counterpart borrower or lender. Massive adjustment costs will be incurred when the assets transferred are not available in the amounts, for the duration, or with the risk characteristics the household desires. These costs will be reflected in a mismatch between actual and desired consumption both now and in the future. For net lenders unable to find suitable repositories for their assets, considerable storage resources may be needed to provide own safekeeping of the funds. Finally, the high stress arising from the uncertainties of transacting in a world without intermediaries is likely to impose high psychic costs.

The existence of intermediaries centers on their ability to significantly reduce these costs by realizing scale economies in information processing, risk pooling, and safekeeping. In particular, they reduce households' information and adjustment costs. While intermediaries in general provide some levels of all five types of distribution services, the two most central to the intermediation function are information and assurance of product availability at the desired time and in the desired form.

The relative importance of these two services depends on the class of intermediary. The simplest class, typified by stock exchanges, provides a forum for interaction among borrowers and lenders, and is primarily providing an information service. If the forum is sufficiently well attended, however, it also ensures high levels of assurance of product availability for most agents. Going a step beyond that, brokers, who 'make the market' by matching borrowers and lenders, provide both higher levels of assurance of product delivery and high levels of information.

The most sophisticated intermediary is one, typified by banks, that creates new assets from combinations of existing ones. As a result, banks provide substantially higher levels of assurance of product availability, with regard not only to the divisibility and duration but also to the risk characteristics of the new financial products. By creating new products, banks subsume many of the household's information costs into this assurance of

availability service. The direct information service of banks is correspondingly lower than that of an exchange or a broker. It consists mainly of explaining characteristics of their own financial products rather than the individual characteristics of the multitude of individual borrowing and lending options agents would face if they transacted with each other.

These products are primarily distinguished from one another by variations in the level of assurance of product availability. A central aspect of this service is the provision of *liquidity*, defined broadly as the flexibility of customers' access to assets deposited with or borrowed from the bank. In general, the more liquid asset is one with fewer restrictions on timing or purpose of use. More liquid assets reduce households' costs of adjustment to uncertain expenditure patterns; greater availability as means of payment also reduces direct time and transportation costs of obtaining cash.

A closely related area of assurance of product availability is *insulation against interest rate and default risk*, in the first case through contracts for fixed rates of interest, and in the second through the provision of deposit insurance. Here, too, the value to the household lies in the reduction in adjustment costs in the face of uncertainty.

Conceptually, the liquidity and risk insulation services taken together would seem the most logical candidates for defining what banks are explicitly in the business of selling. This represents a departure from the standard model of retailing presented above. Since banking is a pure distribution services activity, the explicit product is itself a distribution service. As in the standard model, the explicit product can be marketed in conjunction with a variety of levels of implicit distribution services.

An area of variety is in the assortment of the explicit product. By the very nature of intermediation, all banks provide at least some range of liquidity options to their customers, and many supply risk-insulation options. A broad assortment of deposits might range from an ordinary checking account (with few restrictions on the timing or amount of withdrawal) to large denomination time deposits (with substantial penalties for early withdrawal and inaccessibility as a means of payment). Intermediate options might be money market demand accounts and small time deposits. A broad loan assortment might span from unsecured consumer lines of credit (highly liquid) to highly structured, secured real estate loans.

Depth in assortment is reflected in the range of options available for products with similar uses, e.g. the choice among fixed or variable rate mortgages of various durations, among checking instruments with different minimum balance requirements, or between money market accounts with and without deposit insurance. Broad assortments reduce direct time and transportation costs by enabling households to conduct multiple purchases at the same place. Deep assortments lower adjustment costs by increasing the likelihood of finding the desired type of transaction among the selection. Throughout this paper, the terms 'generalist' and 'specialist' will refer to banks with broad and deep assortments, respectively.

Banks increase accessibility by expanding locations and increase assurance of product availability at the desired time by expanding opening hours. They can expand information services with financial planning advice and record-keeping services such as cancelled checks and electronic balance information by phone. Ambience can be improved with increased resources for customer services and more comfortable surroundings. Increases in any of the services—whether explicit or implicit—raise bank costs.

Relation to Prior Banking Models

Although many of the distribution services discussed above have been noted in the prior literature on banking, attempts to incorporate services into the formal optimization process have been limited in scope. Following Tobin's (1956) and Baumol's (1952) theory on the transactions demand for currency, there is a literature explaining consumer willingness to hold lower-yielding assets as a function of the liquidity services banks provide them (Pesek, 1970; Klein, 1974; Towey, 1974; Sealey and Lindley, 1977; Barnett, 1981). Other models (Ali and Greenbaum, 1977; Williamson, 1987) explicitly capture the accessibility of location service, but under very restrictive assumptions about the behaviour of banks in other respects. The present framework is consistent with the liquidity services tradition, but considerably broader in scope—both by its inclusion of a range of additional services and by its consideration of a wider range of bank products, including loans and less liquid deposit accounts.[4]

This last point inevitably positions the model in the long-standing debate on which items in the

balance sheet constitute outputs of the banking system. This model stands in contrast to those that argue that only loans (Sealey and Lindley, 1977), or deposits (Pesek, 1970; Towey, 1974), or only some rather than all of both (Hancock, 1985; Fixler and Zieschang, 1990) are bank outputs. The inclusion of the full range of deposits and loans as outputs of the banking system follows naturally from the basis for their demand. Each is sought for the reduction in distribution costs it provides to the customers; banks sell the associated distribution services in order to earn profits.

A divergence between the present model and much of the literature lies in the formal modelling of uncertainty. Most firm-theoretic models capture some element of uncertainty through a stochastic optimization problem, where the firm maximizes either expected profit or, in the case of risk aversion, the expected utility of profit. Here, by contrast, the formal model is deterministic; uncertainty is introduced on the demand side through the presence of adjustment costs and on the supply side through the costs of providing the service of assurance of product availability at the desired time and in the desired form.

The approach here is complementary to, rather than a substitute for, the stochastic models. This is best seen by comparing the implications for the three main types of uncertainty that banking models consider: output demand shocks (through uncertain deposit supply or loan demand), input supply shocks (through movements in the market rate of interest), and probability of loan default.

In this model, banks may choose to take on higher levels of the first two types of uncertainty as a way of providing higher levels of assurance of product availability to households. The provision of increased deposit and loan liquidity raises the variance of output demand shocks, and the guarantee of an output price (fixed interest rate) for a specified period of time makes the bank more vulnerable to input supply shocks. Increases in both types of services are assumed to increase operating costs, if for no other reason than that management has to consider whether or not to make adjustments in inventories, to immunize the portfolio, etc. Whether the increased uncertainty also raises interest costs (in the first case through foregone earnings from increased inventories in low-yielding reserves, and in the second through losses due to unfavorable swings in the market interest rate) is not determined formally in this model. Stochastic models provide various answers to this question, depending on the nature of risk preference of the bank and the types of adjustment costs that are assumed to exist (Santomero, 1984). Stochastic models do not capture the fact that uncertainty raises operating costs.

The question is different for default risk, because the bank does not take on additional risk in order to provide assurance of product delivery. The bank's risk-pooling activities provide this service by lowering the information and adjustment costs to individual depositors, but the depositors (and equity holders) bear any remaining portfolio risk.[5] Increases in uncertainty of loan repayment are assumed to raise the operational costs of successfully risk pooling. How the risk premia on individual loans should be set, as well as how they collectively affect the risk-adjusted returns on various deposit categories, would need to be determined in a stochastic framework.

HOUSEHOLD DEMAND FOR FINANCIAL SERVICES

An appropriate framework in which to examine the demand for financial services is a life-cycle model of consumption. In the perfect capital market version of that model, households can borrow and lend at a single market rate of interest, equal to the rate of discount. If the term structure of the interest rate is flat, they have no reason to simultaneously hold positive and negative assets. Assets are homogenous, and provide no distribution services (there being no distribution costs).

The present model introduces imperfections in capital markets by imposing costs on individuals for engaging in intermediation. In order to lend or borrow at the rate of interest identified in the perfect capital market world, households would have to incur very high distribution costs, in terms of their own time and other resources. The demand for bank deposits and loans is a function of their ability to help households reduce their own costs of transferring funds in the pursuit of their objectives of maximizing the intertemporal utility of consumption. Because assets take on diverse characteristics, it is quite likely that households will hold various types—both positive and negative—at any given time.

The Household's Optimization Problem

The optimization is set up as a two-stage, representative agent, household production problem.[6] The

household first minimizes expenditures of producing given levels of commodities for each period of life, and then chooses the utility-maximizing levels of those commodities. Commodities may include such items as nutrition, shelter, leisure, and entertainment, in each period and at various levels of aggregation. For simplicity of exposition, the problem is expressed here in two periods,[7] with a stable price level.

Stage one For a given level of commodities in each period the household chooses the expenditure-minimizing levels of the bank's loan and deposit products and other market goods (including own time used in household production), subject to its production technology. Levels of implicit distribution services supplied by the bank are given and enter as *fixed inputs* into household production. That is, the household cannot vary the level of implicit services accompanying the explicit products it chooses from the bank. To do so, it would have to find a competing bank with a different service offering. Formally, the competitors' products appear in the environmental vector affecting household production:

$$\begin{aligned}\underset{\{D,L,Q_{1,2}\}}{\text{Min}}\, E &= R^*D + I^*L + P_1Q_1 + P_2^*Q_2 \\ &\quad - x(Z_1(D, L, Q_1; S, N) - Z_1^0) \\ &\quad - y(Z_2(D, L, Q_2; S, N) - Z_2^0) \qquad (1)\end{aligned}$$

wnere

subscripts 1 and 2 denote the periods,
D and L are vectors of deposit and loan products, measured in currency units,
Q_1 and Q_2 are vectors of non-financial market goods,
S is a vector of implicit distribution services accompanying the bank's deposits and loans,
Z_1 and Z_2 are vectors of commodities yielding utility,
N is a vector of natural and social environmental factors such as topography, education levels, and the price and service offerings of other banks,
P_1 and P_2^* are the transposed vectors of prices of non-financial market goods (for labor, the wage rate),
R^* and I^* are the transposed vectors of deposit and loan prices,
*denotes discounting by $1/(1+r^m)$,
r^m is the market rate of interest, and
x and y are the Lagrangian multipliers.

Because the household chooses for both periods at the beginning of period 1, prices of market goods paid for in period 2 are discounted by the market rate of interest. Since this is the rate that can be obtained with sufficient investment in intermediation activities, it can be thought of as an interbank rate. The prices of deposits and loans consist of an interest margin in relation to the market rate, plus any per-volume non-interest fee. Thus, even though households purchase and have access to these assets in period 1, they do not pay for them until the beginning of period 2 when the interest has accrued:

$$I_j^* = (i_j - r^m + h_j)/(1+r^m) \quad \text{for } j = 1, \ldots J$$

$$R_k^* = (r^m - r_k + h_k)/(1+r^m) \quad \text{for } k = 1, \ldots C$$

$$r_k - h_k \leq r^m \leq i_j + h_j$$

wnere

i_j is the interest rate charged on loan L_j,
r_k is the interest rate paid by the bank on deposit D_k, and
$h_{j,k}$ are the respective non-interest fees.

The rationale for defining these prices in terms of the interest margin follows from the nature of the market imperfections in the model. At some (perhaps exhorbitant) level of expense, households could achieve the market rate for loans and deposits; they pay banks the spread between that rate and the one actually charged (for loans) or paid (for deposits) in exchange for the intermediation services.[8] The specification of the non-interest (often called 'service') fee as applicable to the whole package of services accompanying the loan or deposit is consistent with the standard assumption made in the literature (Klein, 1974, Startz, 1979; Ho and Saunders, 1981). Charges for individual distribution services are considered below.

In reality, the household may choose assets other than those provided by banks. These are generally abstracted from here, on the assumption that their qualities can be duplicated by some types of intermediated products.[9] The exception is currency, D_c. Currency has a higher price than any deposit category which pays a positive interest rate net of service charges, since $R_c^* = r^m/(1+r^m)$. Thus, to be held, currency must provide some types of services superior to deposit accounts. In addition to its greater liquidity under some circumstances (when

one can only pay in cash), it has high accessibility, entails low information costs, and may have desirable risk properties for some households' portfolios (Tobin, 1958).

The solution of the first-order conditions for the endogenous variables and their insertion into the expenditure equation generates the household's expenditure function:

$$E = E(R^*, I^*, P_1, P_2^*; S, Z_1, Z_2, N) \quad (2)$$

The application of Shephard's Lemma yields the conditional demand functions for the market goods. For deposits and loans these are:

$$D_k = E_{R^*k}(R^*, I^*, P_1, P_2^*; S, Z_1, Z_2, N), \quad k = 1, \ldots C \quad (3a)$$

$$L_j = E_{I^*j}(R^*, I^*, P_1, P_2^*; S, Z_1, Z_2, N), \quad j = 1, \ldots J \quad (3b)$$

Stage two The household maximizes intertemporal utility by choosing commodities to be consumed in each period, subject to the constraint that full lifetime wealth is at least as great as optimal expenditures on market goods (Eqn (2)). Full lifetime wealth includes any initial positive or negative endowment, plus the present discounted value of lifetime human capital:[10]

$$\underset{\{Z_1, Z_2\}}{\text{Max}}\ U(Z_1, Z_2) + u[A_0 + w_1 T_1 + w_2^* T_2 - E(R^*, I^*, P_1, P_2^*; S, Z_1, Z_2, N)] \quad (4)$$

where

A_0 is initial non-human wealth,
w_1 and w_2^* are wages, discounted in period 2 as above,
T_1 and T_2 are hours of productive household time,
$W = A_0 + w_1 T_1 + w_2^* T_2$ is lifetime wealth,[11] and
u is the Lagrangian multiplier.

The solution is a set of commodity demand functions:

$$Z_{ij} = Z_{ij}(R^*, I^*, P_1, P_2^*; S, W, N), \quad i = 1, 2; j = 1, \ldots M, N \quad (5)$$

Substitution of Eqn (5) into the conditional demand system in Eqn (3) yields the unconditional (Marshallian) demands for market goods. For deposits, these are:

$$D_k = E_{R^*k}(R^*, I^*, P_1, P_2^*; S, N, Z_1 (R^*, I^*, P_1, P_2^*; S, W, N), Z_2(\cdot)) \quad (6a)$$

$$D_k = f(R^*, I^*, P_1, P_2^*; S, W, N) \quad \text{for } k = 1, \ldots C \quad (6b)$$

Properties of the Demand for Financial Services

The demand for bank products differs from standard Marshallian demands in two key respects. First, because the implicit distribution services are not priced separately, they do not have price elasticities of demand. Rather, their demand is expressed in relation to that for explicit loan or deposit product(s) with which they are bundled. These 'service elasticities of demand' measure the percentage change in demand for the *explicit product* with a percentage change in the level of the implicit service. The presence of output levels rather than prices in this term causes it to behave inversely to a price elasticity.

Second, because this is a derived demand system, the price and service elasticities consist of two sorts of effects—a 'direct production effect' on demand for given levels of commodity production (Z_{li}), and a 'consumption effect' that captures the influence on demand of optimally adjusting the levels of Z_{li} to changes in prices and service quantities. In a manner reminiscent of the Slutsky decomposition of the standard Marshallian demand function, the direct production effects can be thought of as the 'net' effects of price and service level changes. The 'gross' effects relevant to the banker depend on the interaction of the two effects. The decomposition follows from Eqn (6a):

$$E_{kk} = e_{kk} + \sum_i \sum_l w_{kli} n_{lik}, \quad (7a)$$

between a market good and a market price

$$E_{kh} = e_{kh} + \sum_i \sum_l w_{kli} n_{lih}, \quad (7b)$$

between a market good and an implicit distribution service

where

$$E_{kk} = (\partial D_k / \partial R_k^*)(R_k^* / D_k)$$
$$= [1/(1+r^m)](\partial D_k / \partial h_k)(R_k^* / D_k)$$
$$= -[1/(1+r^m)](\partial D_k / \partial r_k)(R_k^* / D_k)$$
$$E_{kh} = (\partial D_k / \partial S_h)(S_h / D_k)$$

$$e_{kk} = (\partial D_k/\partial R_k^*)|_z (R_k^*/D_k)$$
$$= [1/(1+r^m)][(\partial D_k/\partial h_k)|_z (R_k^*/D_k)]$$
$$= -[1/(1+r^m)][(\partial D_k/\partial r_k)|_z (R_k^*/D_k)]$$
$$e_{kh} = (\partial D_k/\partial S_h)|_z (S_h/D_k)$$
$$w_{kli} = (\partial D_k/\partial Z_{li})(Z_{li}/D_k)$$
$$n_{lik} = (\partial Z_{li}/\partial R_k^*)(R_k^*/Z_{li})$$
$$n_{lih} = (\partial Z_{li}/\partial S_h)(S_h/Z_{li})$$

The net and gross price elasticities of demand with respect to changes in interest rates (i_j and r_k) and non-interest fees ($h_{j,k}$) are identical for loans, and identical but of opposite sign for deposits.

The direct production effects are as follows:[12]

(1) Non-positive own-price elasticities of demand ($e_{kk} \leqslant 0$) are guaranteed by concavity of the expenditure function.
(2) The signs of the implicit distribution service elasticities of demand with respect to a market good (e_{kh}) are positive, zero, or negative as the services are complements, independents or substitutes with that good in the production of Z_{li}.
(3) The signs of the cross-price elasticities of demand with respect to other market goods (e_{kj}) are negative, zero, or positive as these are complements, independents, or substitutes in the production of Z_{li}.

Since increases in market good prices and reductions in service levels make it more costly for a household to produce commodities, the consumption effect will generally be negative with respect to price changes and positive with respect to service level changes.[13] As a result, this effect reinforces the negativity of the net own-price elasticity of demand, and the positivity of the cross-price and service elasticities of demand for net complements. Moreover, it *drives toward gross complementarity* any market goods and distribution services that are net independents or substitutes.

Complementarity or its absence determines the demand-side incentives for the bank to expand its assortment. Both the consumption and direct production effects suggest that these incentives are great. The negativity of the consumption effect pushes all but strong net substitutes toward gross complementarity, and reduces the latter to highly imperfect gross substitutes. This effect is larger for bank products with high budget shares in the production of commodities, and for commodities that have high income elasticities. Thus mortgage loans (for home owners) and consumer loans (for renters) should have particularly strong consumption effects. So should highly liquid financial products that are used for transactions.

The two main factors determining the direct production effects are the definition of the Z's and the degree of liquidity distinctions among the loans and deposits. Whatever these conditions, few of the financial products considered here are likely to be net complements,[14] so the issue is whether they are independents of substitutes, and to what degree. As the liquidity distinctions widen, products move from substitutability toward independence. Substitutability also diminishes, the more disaggregated the consumption objectives are with respect to time, place, and purpose. For meeting nutrition needs over the next year, the money market demand account is sufficiently liquid to be able to substitute for an ordinary demand account. For buying this weeks' groceries, the two are probably independent. For buying tonight's milk from the corner grocery, the substitutability between the ordinary demand account and currency may even disappear, assuming the grocer does not accept checks. If, however, he or she accepts credit cards, this liquid loan account substitutes for the liquid deposits.

It seems reasonable to think of households as pursuing Z's with various levels of aggregation at any given point in time. Even financial products with very similar liquidity characteristics may be gross complements or independents over highly disaggregated Z's, although they may function as strong net, and hence gross, substitutes when held in higher quantities. As the corner grocery example illustrates, institutional factors may cause markets to vary substantially in the degree of asset similarity.

The bank's incentive to raise the levels of implicit distribution services depends on their complementarity with the explicit products and on their substitutability with another market good, the household's own time. Both factors suggest strong demand-side pull for service expansion. In general, the bank's implicit distribution services will tend to complement the liquidity and risk-insulation services of the explicit products in the reduction of household distribution costs.[15] Since many of the distribution costs take up household time, both explicit and implicit products will tend to be strong substitutes for that resource. As the value of household time increases, so should the demand for bank

services. This effect is reinforced by the non-negative relationship between wealth and the demand for the explicit bank outputs (B&G, 1988).

The model's result that the demand for explicit bank products is generally positively related to increases in implicit services is consistent with previous studies that incorporate a quality dimension into bank products (Klein, 1974; Towey, 1974; Barnett, 1981, Fixler and Zieschang, 1990). This framework provides deeper insights into the nature of demand for services by exposing its underlying relationship to the opportunity cost of the household's own time. A recent study by Dowd (1990) makes this connection in the context of the demand for money as a function of the direct time and transportation costs of transacting. The model here demonstrates the much wider validity of the relationship for most, if not all, services, and consequently, for the full range of explicit products.

Formal analysis of the interaction of demand patterns *across* explicit products in prior studies has focused on a very limited range—typically between a demand deposit and a savings instrument (Baumol, 1952; Tobin, 1956; Mitchell, 1979; Barnett, 1981). The general assumption made, either explicitly or implicitly, is that bank products are gross substitutes. This model demonstrates that there are in fact strong tendencies toward gross complementarity among all except the closest of substitutes in a bank's assortment. It thereby provides a basis for explaining the demand-side phenomenon that individual consumers hold multiple asset types at any given time, and for analyzing what Litan (1987) has dubbed 'economies of scope in consumption' as a factor influencing the multiproduct nature of banking.

THE TECHNOLOGY FOR PRODUCING FINANCIAL SERVICES

The bank's optimization problem can also be viewed as having two stages—cost minimization and profit maximization. The first stage reveals the special properties of bank technology in this model. The bank's problem is to minimize the costs of producing given levels of deposits, loans and distribution services, subject to its technology and input prices:

$$\underset{\{X\}}{\text{Min}}\, C = VX - z[H(X, D, L, S)] \qquad (8)$$

where

V is the transposed vector of non-financial input prices,
X is the vector of non-financial inputs,
$H(X, D, L, S)$ is the continuous transformation function, decreasing in X and increasing in D, L, S, and
z is the Lagrangian multiplier.

The solution is a set of conditional input demand functions, which can be inserted into the cost equation to yield the bank's joint cost function:

$$C(V; D, L, S) \qquad (9)$$

By construction, this function has the standard properties: non-decreasing, linear homogeneous and concave in input prices, and increasing in at least one input price and in the explicit and implicit products of the bank.

Two types of production interrelationships cause this cost function to be *inherently* joint in ways not previously recognized in the literature. First, many implicit services are not specific to a single explicit product, but common to several or many items in the assortment. The accessibility of bank branches, ambience, and the level of assortment itself is common to the entire range of products. Information services tend to be common over some range of deposit or loan categories relevant to the customer's selection process. Accessibility through automated teller machines tends, by contrast, to be specific to a narrow range of highly liquid deposit accounts. Unlike specific services, the costs of common services can be recouped over the full range of explicit products they accompany. This property has significant implications for equilibrium bank pricing behaviour, to be discussed in the following section.

Second, some implicit services are jointly produced, in the sense that an increase in one automatically raises the level of the other. An expansion of assortment also increases assurance of product availability and information services at the bank. Raising information services may also increase the ambience of transacting at the bank. Services with this property give the bank cost incentives to expand their output in favor of other services.

The model also provides insights into the technology debate that has occupied the literature—the existence or absence of cost complementarities among explicit bank products. Various authors (Adar *et al.*, 1975; Lawrence and Shay, 1986; Berger *et al.*, 1987; Hunter and Timme, 1989) have made a

theoretical case against the common practice of assuming cost function separability in banking models (e.g. Towey, 1974; Sealey and Lindley, 1977; Mitchell, 1979). They argue that indivisibilities in inputs such as computer technology and the possibility of sharing the same information about customers for different types of accounts would give cost advantages to multi-product banks over specialist, single-product institutions.

The results of the numerous empirical studies on this question have been mixed, frequently rejecting the cost-complementarity hypothesis (Clark, 1988). These studies, however, do not consider the provision of assortment as an implicit, cost-increasing output of the bank. Assortment costs are those arising from the increasing operational complexity of an expanded number of explicit products. In empirical work the presence of assortment costs will appear as a *diseconomy* of scope between explicit products if they are not controlled for separately. This suggests that the existing empirical work testing for jointness between explicit bank products has been biased against its presence. If the assortment costs are sufficiently high (as they might well be between the broad output categories typically considered in empirical work), they could substantially mute, and even reverse, the effects of joint production between the explicit products themselves.

IMPLICATIONS OF PROFIT MAXIMIZATION

The Bank's Optimization Problem

In the second stage the bank chooses the profit-maximizing level of output prices and distribution services, subject to the cost-minimizing choice of inputs, the demand functions for explicit and implicit outputs, and a balance sheet constraint. Several aspects of this problem stand for elaboration.

Price setting Product bundling, with implicit products unable to be marketed separately, leaves the bank facing imperfectly elastic demand curves. The model does not assume additional market power owing to non-contestability, but extensions to incorporate oligopolistic conditions should yield qualitatively similar results. The bank's decision to choose over price rather than quantity is considered more realistic, particularly for deposits.

The balance sheet The bank's balance sheet constraint is that the total value of loans equals the total value of deposits (less any reserve requirements (d_k)) plus net worth (NW).[16] It is assumed that the bank has access to a wholesale source of funds, A^w, at the market rate of interest, r^m, which it can hold in positive or negative amounts. This specification is similar to that in other models that allow banks to engage in both deposit and loan production (Klein, 1971; Baltensperger, 1980; Slovin and Sushka, 1983; Prisman *et al.*, 1986). It also allows for special cases where a bank concentrates its production on the deposit side (Sealey and Lindley, 1977; Towey, 1974; Mitchell, 1979, 1988) or the loan side. Examples of the former type are depository institutions like the European postal savings banks, as well as the proposed 'narrow bank' (Litan, 1987), and of the latter, consumer finance companies.

Timing The bank's timing conforms to the household's. Deposits and loans are issued in period 1, but the bank receives its revenue on them at the beginning of period 2. Production costs are incurred in period 1, financed by working capital available at the interbank rate of interest, r^m. From the perspective of period 1, the bank is maximizing the discounted profits it will earn in period 2.

Substituting in the balance sheet constraint,

$$\sum_j L_j - \sum_k D_k(1-d_k) - NW = A^w$$

the problem takes the form:

$$\underset{\{i_j,\, r_k,\, h_{j,k},\, S_h\}}{\text{Max}}\ \pi^* = I^*L + R^*D + r^m NW - r^m dD - C(V;\, D,\, L,\, S) \quad j=1, \ldots J;\ k=1, \ldots C-1;\ h=1, \ldots H \tag{10}$$

The first-order conditions can be expressed as follows:

$$a_j = \sum_i a_i M_i(-E_{ij}) + (R^*D/I^*L)\sum_k a_k(M_k)(-E_{kj})$$

$$j,\ i=1, \ldots J;\ k=1, \ldots C-1, \text{ for loans} \tag{11a}$$

$$a_k = (I^*L/R^*D)\sum_j a_j M_j(-E_{jk}) + \sum_i a_i(M_i)(-E_{ik})$$

$$j=1, \ldots J;\ k,\ i=1, \ldots C-1, \text{ for deposits} \tag{11b}$$

$$SE_h = \sum_j b_j M_j(E_{jh}) + \sum_k b_k M_k(E_{kh})$$

$$h=1, \ldots H, \text{ for implicit services} \tag{11c}$$

where

$a_i = I_i^* L_i / I^* L$; $R_i^* D_i / R^* D$, the share in net financial revenues from loans or deposits of the ith loan or deposit product,

$b_i = I_i^* L_i / C$; $R_i^* D_i / C$, the share in production costs of the ith loan or deposit product,

$M_i = (I_i^* - C_{Li}) / I^* i$; $(R_i^* - C_{Di} - r^{m*} d_i) / R_i^*$, the marginal profit rate on the ith loan or deposit, including interest loss on reserves in the latter case,

$SE_h = C_{Sh}(S_h / C)$, the proportionate contribution to production costs of the hth distribution service, and

E_{ij} are the elasticities of demand (Eqn (7)).

The discussion of the results proceeds by first indicating the general implications, and then applying them in the context of three areas of controversy in the banking literature: the linkages between deposit and loan pricing, the nature of banking market structure, and the optimal level of service fees.

General Results

In general, the banker will need to adjust both output pricing and implicit service levels in response to an exogenous change in demand. As the following results show, however, the optimal response varies according to the nature of the change:[17]

(1) An exogenous increase in *demand for the explicit product* (for given price and service elasticities of demand) generally allows the firm to increase its profits by raising prices and/or lowering distribution services.

(2) By contrast, exogenous increases in the *price sensitivity of demand* (reflected in higher absolute values of own-price elasticities of demand, and greater degrees of gross complementarity with other items in the assortment) lead, other things being equal, to lower prices and/or lower levels of the distribution services.

(3) Exogenous increases in the *sensitivity of demand to service levels* (reflected in higher values of the elasticities of demand for explicit products with respect to services) lead to higher levels of services and/or lower prices.

In the last two cases, the altered relationship between price and service elasticities limits the bank's ability to increase profits through price increases, and lowering services, respectively.[18] With increased price sensitivity, for a given level of prices, the household demands relatively lower levels of distribution services. With increased sensitivity to the services, a given level of implicit services will only be demanded at a lower price.

These conditions have interesting implications for the levels of services and prices under different types of overall demand conditions. An exogenous increase in demand might arise from an increase in non-human wealth or a rise in the wage rate. On this basis, the bank in the area with higher wealth could expect to earn higher marginal profits on deposit and loan products, and/or provide lower implicit services.

Wealth increases occasioned through the wage channel should also result in relatively higher service elasticities of demand, as higher wages denote a higher opportunity cost of household time. This puts an opposite pressure on both prices and service levels. If the latter effects dominate, one can expect to see higher service levels in the high-wage areas relative to lower-wage ones. This conclusion corresponds to the view that highly service-oriented banking, such as 'one-stop financial shopping' and 'financial supermarkets' are likely to gain in importance as the opportunity cost of households' time rises (Litan, 1987; Benston and Smith, 1976). What these hitherto informal observations have not been able to surmise, however, is the somewhat counter-intuitive result that this type of demand shift may also lower prices. Thus, the result is also consistent with the observation of consumer advocates such as the Consumer Federation of America that banks provide lower levels of service and charge higher prices to consumers in poorer urban neighborhoods. No collusion among bankers is necessary for this to occur.

A further implication of the first-order conditions concerns the bank's ability to realize scale and scope economies in the provision of services. The most general source of cost-spreading potential lies in the common distribution services of the bank. An intuitively straightforward result of profit maximization is that output price increases will need to be greater for increases in specific distribution services than in common services (B&G, 1989). Put otherwise, banks that provide high levels of common distribution services will generally pursue lower price policies than those with high levels of specific distribution services. This divergence will be proportionately greater, the larger the assortment over which the services are common.

Deposit and Loan Pricing Policy

The results provide a basis for re-examining an area of considerable controversy in the banking literature—the extent to which deposit and loan-pricing decisions are separate. In one of the most often-cited theoretical models of the banking firm, Klein (1971) constructs a scenario in which separability is maintained. The contours of his model are similar to this one in two key respects: imperfect elasticity of demand for the bank's own deposit and loan products, and a perfectly elastically supplied source of wholesale funds to the bank.

Klein's result caused a stir because it questions a primary justification for regulatory price controls on deposit interest rates. If loans and deposits are priced separately, regulators need not be concerned that allowing interest rate competition on deposits would drive banks into more highly priced lending activity (with, by implication, higher risk of default). Various authors have contributed to the subsequent literature demonstrating the extent to which Klein's separability result is a special case. Separability can be rejected with the introduction of some risk-related linkage between the two sides of the balance sheet, as in loan and deposit duration mismatch (Benston and Smith, 1976; Ho and Saunders, 1981). It can also be rejected by modifying Klein's assumptions on the elasticities of demand for bank outputs and the supply elasticity for wholesale funds (Pringle, 1973; Langohr, 1982; Slovin and Sushka, 1983).

The present model demonstrates that separability is invalid even if one abstracts from risk issues and maintains Klein's assumptions regarding the elasticities of asset demand and supply. Demand and supply interdependencies will, in general, require banks to make deposit and loan-pricing decisions jointly. The demand side linkages suggest that generalist banks with a higher proportion of gross complements among their explicit product assortment will charge lower prices, and have lower marginal profit rates, than those more specialized institutions with less complementarity among products. Jointness in supply will allow banks that provide relatively high levels of common distribution services to operate with lower margins than those with high levels of specific services. Since the ability to spread costs of common services is proportional to the level of assortment, this suggests again that the more generalist, common services-oriented institution will charge lower prices than the specialist who concentrates on specific services over a narrow assortment of explicit products. How these pricing patterns affect the level of bank-specific default risk depends on the relationship of loan pricing to the probability of default, a matter that is beyond the scope of this model.[19]

The Nature of Banking Market Structure

Whether and how different types of firms can co-exist is a subject on which this model is also able to provide new insights. The first concerns the nature of competition. Product bundling was noted earlier as a condition presenting banks with imperfectly elastic demand curves. Contrary to some of the literature (Towey, 1974; Klein, 1974), there is no reason to expect, under these circumstances, that the market will naturally converge to one where banks offer identical service mixes at a perfectly competitive price. Even under the simplest of cases, where technologically homogenous banks offer a single explicit product to a homogenous set of households, different banks would only choose the same combination of service levels and output price by chance; households are indifferent between a range of such combinations (B&G, 1988). Under the more realistic assumption that banks can also vary the number of explicit products, the likelihood of service level and price convergence is virtually nil (B&G, 1989). Thus the model provides a theoretical foundation for the frequently made assumption of imperfectly elastic demand for banking products. In general, retail banking will be, *at most*, monopolistically competitive, with product and price diversity.

The model can also provide guidance on the feasible types of co-existence among banks with different service offerings. In banking, a particular area of interest is the relationship among banks with different ranges of assortment. The interplay between the value to households of an expanded assortment, and costs to banks of providing this expansion, determine the market outcomes.

Consider two scenarios whereunder two banks face identical technology and a homogenous clientele. In case one, the generalist and the specialist are distinguished only by the former's providing a higher level of assortment. Demand conditions require the specialist to charge *lower* prices than the generalist, since a smaller assortment imposes higher distribution costs on consumers. This is only

feasible if the generalist's costs of expanding assortment are high enough to counterbalance the cost savings of common services over the larger assortment and any other scale economies. In case two, the scenario is the same except that the specialist also provides higher levels of some other distribution service. This can allow the specialist to co-exist and charge *higher* prices.

The feasibility of co-existence is interesting, because it occurs despite the restrictive assumption of homogenous consumers, and in the absence of regulatory restrictions that confer advantages to either type of institution. Thus the model provides a market-based explanation for co-existence of generalists and specialists which has hitherto been cast in regulatory terms (e.g. Benston and Smith, 1976). For instance, the first scenario could well account for the co-existence of commercial banks with the more specialized thrift institutions in the United States. The latter have consistently charged lower interest spreads than commercial banks on time deposit accounts (Mahoney *et al.*, 1987). A sufficient explanation for the feasibility of co-existence would be the (not-unlikely) possibility of sufficiently higher assortment costs at commercial banks.

Optimal Service Fees

Finally, this model makes it clear that regulatory obstacles are not a necessary condition for banks' failure to recover their operating costs through non-interest or 'service' fees.[20] The presumption in much of the literature is that binding interest-rate ceilings on deposit accounts are what push banks to pay 'implicit interest' by undercharging customers for the services they receive on these accounts (Klein, 1974; Black, 1975; Startz, 1979; Ho and Saunders, 1981; Mitchell, 1979, 1988; Merris, 1985; Fama, 1980, 1985). In the more common formulation of the problem, typified by the work of Startz, services are not considered in any explicit form, and are assumed to be purchased in a bundle with the deposit account. Other models, such as Mitchell's, have addressed the question by separating out a single service (check cashing) and relating its pricing to bank processing costs.

Both types of models assume that in an unregulated environment the bank would prefer to use a non-interest fee rather than the interest margin, $(r^m - r_k)$, to recoup operating expenses from service provision. In the Startz framework, the bundle of services includes liquidity, so the optimal interest margin is zero ($r^m = r_k$), and the full price of the deposit is $R_k^* = h_k^*$. In an unregulated banking environment, h_k^* would be at a level sufficient to cover marginal and average costs. Services are bundled with distinct deposit categories in a separable cost function. In terms of the present model this is equivalent to a marginal profit rate, $M_k = (h_k^* - C_{Dk})/h_k^* = 0$, where C_{Dk}, the marginal cost of producing D_k, incorporates the costs of producing all implicit services as well as the explicit product. The non-interest fee, h_k^*, would only fall below the level of C_{Dk}, if the bank were constrained to setting $r_k < r^m$. In this case, it would fall directly by the amount of the gap, $h_k^* = C_{Dk} - (r^m - r_k)^*$.

One deficiency in this formulation is the assumption of cost function separability. As long as there are implicit services provided jointly over a range of bank explicit products, the optimal price cannot be expressed as equivalent to the marginal cost of producing the deposit and its associated services. For a given marginal cost of implicit service, C_{Sh}, the optimal prices on deposits will be higher if the service is specific than if it is common to a number of explicit bank products, other things being equal.

Even if joint production is taken into account, however, the assumption that banks would prefer to levy the optimal charge through a non-interest fee rather than the interest rate margin seems unfounded. Once one recognizes that the *total* price of a deposit to the household is $R_k^* = (r^m - r_k + h_k)/(1 + r^m)$, it is clear that the price elasticity of demand with respect to an increase in the fee, h_k, is equivalent to the price elasticity of demand with respect to an equivalent decrease in the deposit rate of interest, r_k (Eqn (7)). If there are no other factors to consider,[21] the bank should be indifferent between charging the consumer through the interest margin or a service fee.

If regulations set the interest margin so low that the total desired price, R_k^*, cannot be attained without setting $h_k < 0$, then from Eqn (11c), the bank will be forced to raise the level of services beyond the desired level in order to stay on the demand curve.[22] This result is consistent with Neven's (1989) conclusion that banks constrained not to compete on price will compete by extending services beyond the level that would be optimal in an unconstrained situation.[23]

The Mitchell framework poses a quite different problem, because it assumes that the service and the deposit account are related, but unbundled. That is, the consumer can choose over both D_k and the level of a specific service, S_k (check clearing) which can,

but need not, accompany it. Each product is explicitly priced—D_k through the interest rate and S_k through a per unit service fee, g_k. The linkage between the two products is a high degree of gross complementarity. If the bank were unconstrained with respect to the price on D_k, it would set g_k equal to the marginal (here, equivalent to average) cost. With a binding interest ceiling on r_k, the bank is willing to sustain net revenue loss on S_k, by lowering g_k. In his framework, this net loss is the 'implicit interest' paid by the bank.

Mitchell then demonstrates that there need not be a one-to-one correspondence between a rise in the explicit rate of interest on D_k and a decline in the implicit rate on S_k, such that the combined marginal earnings to the bank remains constant. The bank's marginal revenues rise with an increase in r_k, if D_k and S_k are not strong complements, and fall under conditions of strong complementarity. Since the two products are related, but separate, this should hardly be surprising. In the framework of the present model, the results follow directly from the first-order conditions, which show that marginal profit rates will be lower for products that are strong gross complements.

When Mitchell's single explicit service is incorporated into the more general model presented here, two further observation are warranted. First, many (if not most) of the implicit services offered by banks will not lend themselves to unbundling from consumption with the explicit product. Typically, only highly transactions-specific services such as check clearing, international transfers, and use of an automated teller machine are readily separable. Some services, such as information, might be priced separately, but with difficulty.[24] Others, like ambience, assortment, assurance of product delivery through reputation, and accessibility to general bank premises, are basically inseparable. Thus even if the bank prefers to charge the consumer directly for some services, it will still price a large number in a bundle with the explicit product. For this latter group, the above discussion on interest margin versus service fee pricing applies, and banks can be expected to 'undercharge' for services.

Second, Mitchell's conclusion that an unregulated bank's optimal choice of the service fee g_k would be set equal to C_{Sk} breaks down if one incorporates a more realistic, multi-product joint cost function.[25] Even a bank that is subject to a long-run market constraint of zero profits will not need to meet this on each individual product; the equilibrium condition is total revenues equals total costs. If some separately priced services are sufficiently strong gross complements with other outputs of the bank, the optimal service fee could be set below marginal cost, in the interest of 'traffic building'. The bank that did so would be paying 'implicit interest'.

Much of the literature dealing with this question preceded the nearly full-scale deregulation of deposit interest rates in the United States over the 1980s. Since 1980, some shifting has occurred in revenue sources on deposits, suggesting that deposit rate ceilings shifted the optimal balance between non-interest and interest margin pricing.[26] Some types of service reduction are also evident, particularly with branch closings, suggesting that the rate restrictions forced banks to provide supra-optimal service levels (Berger and Humphrey, 1990). American banks, however, continue to take in the great majority of their revenues on deposit products through the interest margin, rather than non-interest fees.

CONCLUSION

This paper has demonstrated that a service-based approach to retail banking offers many new insights into the financial firm's decision-making environment. In particular, the model substantially enhances our theoretical perspectives on both supply and demand side factors determining what banks produce and how they charge for it. The result is a framework that is sufficiently rich to be of practical relevance. This conclusion addresses the types of applications to which the framework lends itself.

First, the model could be usefully extended to consider the demand of commercial clients. The analogy of the household problem to the firm is straightforward: in the firm's case the demand for financial services is derived from profit rather than utility maximization. The wage rate will play a similarly important role in determining the demand elasticities for distribution services and explicit outputs, but with the difference that the firm does not face a fixed supply of labor as does the household. The commercial client will also face a different environmental vector (N). Sufficiently large firms can be expected to have access to the types of financial technology that make it feasible to conduct their financial transfers without intermediaries; the result may be substantially different price

and service elasticities than household clients. The importance to banks of determining the nature of these differences is highlighted by the wave of disintermediation by commercial clients that has occurred over the last decade (Litan, 1987).

Exploration of the characteristics of specific bank products within the context of this model would be a useful task for the many bankers engulfed by the resulting wave of attention on the retail client.[27] The implications of the model are important, because they indicate a trade-off between profit margins and volume of demand: items that would be good traffic builders by virtue of high gross complementarity will have lower margins than less complementary items. By the same token, margins are likely to be lower in areas where the opportunity cost of time is higher, precisely those areas where the demand for service intensity will be the strongest.

Finally, the framework can offer guidance to both bankers and policy makers concerning the likely effects of increased competition. B&G (1989) have demonstrated that, in general, the welfare effects of increased competition are ambiguous, in the sense that it will tend to lower prices but may not raise distribution services. If prices are fixed, competition will, by contrast, always be beneficial in the sense of raising services. Application and extension of these results to consider the effects of deregulation, as, for instance, that which is occurring within the European Community, would provide a fruitful avenue of research.

Acknowledgements

Thanks for helpful comments are due to Roger Betancourt, Dennis Fixler, David Gautschi, Kim Kowalewski, Tony Santomero, Kim Zieschang, and the participants of seminars at the European Institute for Advanced Studies in Management, the Brookings Institution, and the NBER Summer Institute. Financial support from the Brookings Institution, the Economic Club of Washington, and the University of Maryland is gratefully acknowledged. The usual disclaimers apply.

NOTES

1. The term 'bank' is used here for all asset-creating intermediaries, including consumer finance companies, executors of mutual funds, and the various bank-like institutions that go by different legal designations, such as thrifts, savings and loans, building societies, and credit unions. 'Financial intermediary' is the more general category, also including firms strictly performing brokerage and information-exchange functions.
2. Readers may recognize these as 'transactions costs'. B&G (1988) use the more general term 'distribution costs' to emphasize that they can be incurred in both market and non-market (e.g. intra-household) exchanges.
3. The corresponding distribution costs in parentheses are illustrative, not exhaustive. For a more detailed discussion of both costs and services see B&G (1986, 1988, 1990a).
4. Another recent approach to bank services aiming at more generality is Fixler and Zieschang (1990), which collapses the multi-dimensional aspects of service into a single-service distance function, thus assuming a perfectly competitive market structure.
5. With deposit insurance, the insurer bears the risk. This analysis ignores moral hazard issues associated with incorrectly priced insurance.
6. See Deaton and Muellbauer (1980, Ch. 10) for a clear exposition of the two-stage technique.
7. Thus assets with different liquidity characteristics are not formally required to be held for different durations. Pissarides (1978) imposes this constraint for deposits. Here it is assumed that the more liquid account will be priced higher, other things equal, because it subjects the bank to greater balance variability.
8. In the literature, loan demand is generally expressed as a function of the loan rate rather than the spread, and consumers' interaction on the deposit side as a supply schedule upward sloping in the deposit rate of interest (Klein, 1971; Sealey and Lindley, 1977; Sealey, 1980; Baltensperger, 1980). The exception is the 'user cost' approach, an early example of which is Barnett's (1981) model of the demand for deposit accounts as a function of the opportunity cost of funds. Although the expressions for prices are similar to this model, user cost models (Hancock, 1985; Fixler and Zieschang, 1990) do not place a non-negativity constraint on them as is done here. Any account categories appearing with a negative price in empirical work are then classified as 'inputs' rather than 'outputs'. This is difficult to justify on theoretical grounds because the negative prices imply both potentially limitless demand and that the bank is choosing an input more expensive than its opportunity cost of capital. More likely explanations for the empirical phenomenon of negative prices (which appear, in some periods, for less liquid deposit accounts) are: (1) measurement of *ex post* rather than *ex ante* prices, where market interest rates have moved unfavorably against a bank locked into fixed-rate contracts; (2) mismatching maturities between the market rate and the consumer rate of interest; and (3) longer-run decision-making horizons, where such factors as adjustment costs lead the bank to take short-run losses on some outputs (see e.g. Flannery, 1982).
9. Thus non-financial assets are redundant given the possibility to purchase shares in the market portfolio. Loans at below-market rates contain a grant element equivalent to a positive endowment.

10. Since this measure values all productive time (including that used in household production), it is the intertemporal concept that corresponds to full income in the single-period model.
11. The value of lifetime wealth is not directly affected by the imperfections in the capital market which drive a wedge between the borrowing and lending rate. These effects enter as prices for financial products, and so have standard income and substitution effects, but no money income effects. The only interest rate directly affecting lifetime wealth is r^m, which discounts future human capital. Hence $dW/dr^m < 0$.
12. Formal proofs are found in B&G (1988, 1990b).
13. By the expenditure function's property of being increasing in output, the first term of the consumption effect, w_{kH}, is non-negative unless the market good is a regressive input into commodity production. Under most circumstances, the second term, the elasticity of demand for the final consumption good, will react negatively to price changes (n_{HK}), and positively to implicit service changes (n_{Hh})(B&G, 1988, 1990b).
14. By contrast, ancilliary items such as mortgage and home owner's insurance would be net complements with the real estate loans themselves. So would transactions accounts with separately marketed transactions services, such as the check-clearing example considered below.
15. Since explicit products with similar liquidity characteristics can be substitutes, an expansion of assortment depth may cause demand to fall for some items in the assortment. Similarly, an increase in accessibility or any other service that lowers the cost of switching between different deposit types will tend to complement the less liquid (higher-yielding) explicit product and substitute for the more liquid (lower-yielding) one. This result corresponds to a reduction in transactions costs in a transactions demand for money model (Baumol, 1952; Tobin, 1956).
16. Net worth is assumed constant. Reserves could include (exogenously determined) excess amounts to respond to output demand shocks.
17. See B&G (1989) for a technical exposition of the propositions summarized here. One difference in this multi-period framework is that terms relating revenues to costs—M_i and b_i—incorporate the discounting of the former in relation to the latter. Also, as a result of imposing a balance sheet constraint, the bank's marginal profit rate, M_i, cannot be decomposed into a retail margin less marginal production costs as it can for the retail merchandizing firm. Here the profit rate is defined in relation to the net price, I_i^* or R_i^*, rather than the gross price.
18. This is seen in Eqns (11a-c) in a lower M_i with a higher absolute price elasticity $|E_{kj}|$, or service elasticity E_{kh} for given revenue (a_i) or cost (b_i) shares. Attention here is confined to the general cases of gross complementarity among explicit products ($E_{kj} < 0$) and between explicit products and services ($E_{kh} > 0$). Extensions of the results to cases of gross substitutability are the subject of current research.
19. For instance, default risk might increase for a stable customer pool by virtue of the higher amount of repayment that a higher loan rate entails. Alternatively, increases in the loan rate may alter the customer pool in a risk-increasing direction (Stiglitz and Weiss, 1981).
20. 'Non-interest' fee is considered preferable because it clarifies the bank's ability to charge for either explicit or implicit services through the interest margin as well.
21. For instance, greater flexibility to adjust to changes in the market rate might favor pricing through the interest margin. If there are taxes on interest income (at rate t), $(\partial D_k/\partial r_k)(R_k^*/D_k) = (t-1)\ (\partial D_k/\partial h_k)(R_k^*/D_k)$, and the bank's preferred pricing tool will be asymmetric. For price increases, more revenue can be earned for the same demand response with a lowering of r_k than a raising of h_k, while for price decreases the opposite holds.
22. With a fixed M_h greater than the desired M_h, the bank will increase S_h. Under the assumptions that the bank has exhausted scale economies in this service and that the household's demand for it is less sensitive as its level increases, this will raise SE_h and lower E_{kh}, respectively.
23. Startz's empirical work concludes that banks pay less implicit interest than they would pay in unregulated explicit interest. His preferred explanation is that there are upper limits on the amount of services banks are able to provide to compensate deposit holders for the lack of interest payments. In terms of this model, that suggests the unlikely outcome that consumers had reached near-saturation in service provision (reflected in very low values of E_{kh} for all h). A more likely explanation, which Startz acknowledges as possible, is that implicit interest—or, more accurately, the value of services to households—is undermeasured in the estimations.
24. This was, for instance, the conclusion of a French government task force on the prospects for charging explicitly for information services in retail banking (Conseil National du Crédit, 1987).
25. Mitchell's cost function consists of constant unit costs for the single output of check clearing.
26. The percentage of revenue on deposits from non-interest fees is estimated at 5% in 1980, 11% in 1984, and 18% in 1988 (Berger and Humphrey, 1990).
27. Articles like 'Banks discover the consumer' (*Fortune*, 12 February 1990) and 'To market to market . . .' (*Banking World*, March 1990), which describe banks' extensive service-oriented retail strategies, are common fare in the trade journals.

REFERENCES

Z. Adar, T. Agmon and Y. E. Orgler (1975). Output mix and jointness in production in the banking firm. *Journal of Money, Credit and Banking* **7**, 235–43.

M. Ali and S. Greenbaum (1977). A spatial model of the banking industry. *Journal of Finance* **32**, No. 4, 1283–1303.

E. Baltensperger (1980). Alternative approaches to the theory of the banking firm. *Journal of Monetary Economics* **6**, 1–37.
W. A. Barnett (1981). Aggregation of Monetary Assets. In *Consumer Demand and Labor Supply: Goods, Monetary Assets, and Time*, New York: North-Holland.
W. J. Baumol (1952). The transactions demand for cash: an inventory theoretic approach. *Quarterly Journal of Economics* **66**, 545–56.
G. J. Benston and C. W. Smith, Jr (1976). A transactions cost approach to the theory of financial intermediation. *Journal of Finance* **31**, 215–31.
A. N. Berger, G. Hanweck and D. B. Humphrey (1987). Competitive viability in banking; scale, scope, and product mix economies. *Journal of Monetary Economics* **20**, 501–20.
A. N. Berger and D. B. Humphrey (1990). Measurement and efficiency issues in commercial banking. Paper presented at NBER Conference on Research on Income and Wealth: Measurement in the Service Sector, April.
R. Betancourt and D. Gautschi (1986). The evolution of retailing: a suggested economic interpretation. *International Journal of Research in Marketing* **3**, 217–32.
R. Betancourt and D. Gautschi (1988). The economics of retail firms. *Managerial and Decision Economics* **9**, 133–44.
R. Betancourt and D. Gautschi (1989). Two essential characteristics of retail markets and their economic consequences. University of Maryland, Dept of Economics Working Paper No. 89–6.
R. Betancourt and D. Gautschi (1990a). Demand complementarities, household production, and retail assortments. *Marketing Science*, Spring.
R. Betancourt and D. Gautschi (forthcoming, 1990b). The demand for retail products and the household production model: new views on complementarity and substitutability. *Journal of Economic Behavior and Organization*.
F. Black (1975). Bank funds management in an efficient market. *Journal of Financial Economics* **2**, 323–39.
J. A. Clark (1988). Economies of scale and scope at depository financial institutions: a review of the literature. *Economic Review*, Federal Reserve Bank of Kansas City, September/October, 16–33.
Conseil National du Crédit (1987). L'Incidence des Technologies Nouvelles sur l'Activité des Intermediares Financiers. Documentation de la Banque de France, April.
A. Deaton and J. Muellbauer (1980). *Economics and Consumer Behavior*, New York: Cambridge University Press.
K. Dowd (1990). The value of time and the transactions demand for money. *Journal of Money, Credit and Banking* **22**, No. 1, 51–64.
E. Fama (1980). Banking in the theory of finance. *Journal of Monetary Economics* **6**, 39–57.
E. Fama (1985), What's different about banks? *Journal of Monetary Economics* **15**, 29–39.
D. Fixler and K. Zieschang (1990). User costs, shadow prices and the real output of banks. Paper presented at NBER Conference on Research on Income and Wealth: Measurement in the Service Sector, May.
M. Flannery (1982). Retail bank deposits as quasi-fixed factors of production. *American Economic Review* **72**, No. 3, 527–36.
D. Hancock (1985). The financial firm: production with monetary and nonmonetary goods. *Journal of Political Economy* **93**, No. 5, 859–80.
T. S. Y. Ho and A. Saunders (1981). The determinants of bank interest margins: theory and empirical evidence. *Journal of Financial and Quantitative Analysis* **16**, 581–600.
W. C. Hunter and S. G. Timme (1989). Does multiproduct production in large banks reduce costs? *Economic Review*, Federal Reserve Bank of Atlanta, May/June, 2–11.
B. Klein (1974). Competitive interest payments on bank deposits and the long-run demand for money. *American Economic Review* **64**, No. 6, 931–49.
M. A. Klein (1971). A theory of the banking firm. *Journal of Money, Credit and Banking* **3**, 205–18.
H. Langohr (1982). Alternative approaches to the theory of the banking firm: a note. *Journal of Banking and Finance* **6**, 297–304.
C. Lawrence, and R. P. Shay (1986). Technology and financial intermediation in multiproduct banking firms. In *Technological Innovation, Regulation, and the Monetary Economy* (edited by C. Lawrence and R. P. Shay), Cambridge, MA: Ballinger.
R. Litan, (1987). *What Should Banks Do?* Washington, DC: Brookings Institution.
P. Mahoney *et al.* (1987). Responses to deregulation: retail deposit pricing from 1983 through 1985. Federal Reserve Board Staff Papers No. 151, January.
R. C. Merris (1985). Explicit interest and demand deposit service charges: a note, *Journal of Money, Credit and Banking* **17**, No. 4, 528–33.
D. W. Mitchell (1979). Explicit and implicit demand deposit interest. *Journal of Money, Credit and Banking* **11**, 182–91.
D. W. Mitchell (1988). Explicit interest and demand deposit service charges: a comment. *Journal of Money, Credit and Banking* **20**, No. 2, 270–74.
D. J. Neven (1989). Structural adjustment in European retail banking: some views from industrial organisation. Centre for Economic Policy Research, Discussion paper series No. 311.
B. Pesek (1970). Bank's supply function and the equilibrium quantity of money. *Canadian Journal of Economics* **3**, No. 3, 357–85.
C. A. Pissarides (1978). Liquidity considerations in the theory of consumption. *Quarterly Journal of Economics* **92**, 279–96.
W. Poole (1968). Commercial bank reserve management in a stochastic model: implications for monetary policy. *Journal of Finance* **23**, 769–91.
J. J. Pringle (1973). A theory of the banking firm: comment. *Journal of Money, Credit and Banking* **5**, 990–96.
E. Z. Prisman, M. B. Slovin and M. E. Sushka (1986). A general model of the banking firm under conditions of monopoly, uncertainty, and recourse. *Journal of Monetary Economics* **17**, 293–304.

A. M. Santomero (1984). Modeling the banking firm, *Journal of Money, Credit and Banking* **16,** No. 4, 576–616.

C. W. Sealey, Jr (1980). Deposit rate-setting, risk aversion, and the theory of depository financial intermediaries. *Journal of Finance* **35,** No. 5, 1139–54.

C. W. Sealey, Jr and J. T. Lindley (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *Journal of Finance* **32,** 1251–66.

M. B. Slovin, and M. E. Sushka (1983). A model of the commercial loan rate. *Journal of Finance* **38,** No. 5, 1583–96.

R. Startz (1979). Implicit interest on demand deposits. *Journal of Monetary Economics* **5,** 515–34.

J. E. Stiglitz and A. Weiss (1981). Credit rationing in markets with imperfect information. *American Economic Review* **71,** 393–410.

J. Tobin, (1956). The interest elasticity of transactions demand for cash. *Review of Economics and Statistics* **38,** 241–47.

J. Tobin, (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies* **25,** 65–86.

R. W. Towey (1974). Money creation and the theory of the banking firm. *Journal of Finance* **29,** 57–72.

S. Williamson (1987). Transactions costs, inflation, and the variety of intermediation services. *Journal of Money, Credit and Banking* **19,** No. 4, 484–98.