Journal of Monetary Economics 6 (1980) 1-37. © North-Holland Publishing Company

ALTERNATIVE APPROACHES TO THE THEORY OF THE BANKING FIRM

Ernst BALTENSPERGER
University of Heidelberg, 6900 Heidelberg, Germany

During the past decade, the importance of a sound microeconomic foundation for aggregate economic analysis has been increasingly emphasized. In this context, a satisfactory theory of bank behaviour appears as an indispensable prerequisite for a clear understanding of the workings of the financial sector of the economy. This has led to the development of a substantial literature attempting to model and explain the behaviour of banking firms. This paper presents a survey and discussion of the various approaches which can be found in this literature. A special effort is made to present an integrated view of the real resource and financial aspects of banking activities.

1. Introduction

During the past decade, the importance of a sound microeconomic foundation for aggregate economic analysis has been increasingly emphasized. In this context, a satisfactory theory of bank behaviour appears as an indispensable prerequisite for a clear understanding of the workings of the financial sector of the economy in general, and of the money supply mechanism in particular.

This has led to the development of a substantial literature attempting to model and explain the behaviour of banking firms. This literature, however, is still unsettled and rather heterogeneous. There exist a number of rival models and approaches which have not yet been forged together to form a coherent, unified and generally accepted theory of bank behaviour. Of course, this reflects the difficult nature of the topic, as well as the different objectives pursued in different studies, and should not necessarily be viewed as an undesirable state of affairs. Nevertheless, it seems worthwhile to review the field and make an attempt to evaluate and compare the various models which have been proposed.

The main economic functions of financial firms are those of consolidating and transforming risks on the one hand, and of serving as dealers or 'brokers' in the credit markets (the basis of which is the existence of transaction and information costs in these markets) on the other hand.[1] A

[1]See Niehans (1978, ch. 9, p. 166 f.) for emphasis on the distinction between these two main functions. For a detailed discussion of the functions of financial intermediaries, see Gurley and Shaw (1960, p. 191 f.).

satisfactory theory of the banking firm which does not take into account these elements is inconceivable, therefore. Uncertainty, informational problems and adjustment costs should and will therefore play a central role in much of the following discussion.

As far as risk is concerned, we can distinguish between two different kinds, both of which are important to a bank. On the one hand, there is 'investment' or 'default' risk in connection with the assets held by the bank: as a creditor, the bank faces the risk that its debtors are not able or willing to meet their obligations at the agreed upon time and that the market's evaluation of its assets and thus their yield fluctuate. On the other hand, there is 'withdrawal' or 'liquidity' risk in connection with the bank's liabilities: as a debtor, the bank faces the possibility that its creditors are unwilling to extend or renew their credit to the bank, or that they are willing to do so at different terms only. This type of risk, of course, assumes a particular weight in the case of demand deposits, where the creditor has a contractual right to withdraw all of his funds at any time, without any restrictions and penalties.

The consolidation and transformation of risks, as well as the production and maintenance of financial contracts and transactions can be performed by the financial firm with the help of real resource (especially labour) inputs only. To a large extent, this is disregarded in the bank behaviour literature. In a way, this is surprising, since the amount of real resources absorbed by the banking industry is of a quite substantial order of magnitude. Presumably, the idea behind this neglect is that the 'real resource' aspects of banking and the 'financial' or 'portfolio' aspects, on which most of the existing literature concentrates, can somehow be separated. The view taken in this paper is that such a separation is not justified, and that, at least for certain types of questions, the real and the financial aspects of banking should be dealt with in an integrated way. This seems particularly clear in connection with all kinds of questions concerning the growth, size, structure and efficiency of the banking industry, including especially discussions of the effects of various types of regulatory constraints imposed on banking firms.[2] But, beyond that, such a view is also potentially important for certain kinds

[2]Much of the discussion in this paper, however, is in terms of a banking firm which is subject to no particular legal constraints and regulations (with some exceptions concerning reserve requirements and deposit insurance in sections 2.1 and 2.2, respectively). Of course, such regulations do exist and affect bank behaviour in most countries around the world, although to different degrees and in different ways. Abstracting from them is not meant to imply that they are unimportant or uninteresting. The justification is, rather, that their impact on bank behaviour cannot be understood unless a theory of bank behaviour in the absence of regulatory constraints has been developed first. That is, the latter should be logically prior, and a study of the effects of regulation then should follow in a second step.

of traditional monetary policy (money supply) questions, as the portfolio reaction of banks to monetary policy actions is influenced by real resource factors, too.

Our discussion will be organized into two main parts: Section 2 (sections 2.1 and 2.2) deals with models of bank portfolio management, i.e. with 'partial' models, in the sense that the total size of the bank's portfolio is assumed to be exogenously determined, so that only the question of the optimal allocation of this portfolio remains to be solved. Section 3 (sections 3.1 to 3.3) then deals with 'complete' models of the banking firm, i.e., models which attempt to explain the joint determination of not only the structure of assets and liabilities and their interaction, but also the total scale of the bank's operation and portfolio.

A large proportion of the existing literature on banking theory consists of partial models falling into the category of section 2. The most apparent need for dealing with bank behaviour on a firm-theoretic level arose in connection with a bank's reserve and liquidity management, which plays a central role in most economist's views of the money supply mechanism. It seems quite natural, therefore, that the first analytical models of bank behaviour were models of bank reserve (liquidity) management. This branch of the literature, which in quantitative terms has dominated up to this day, will be reviewed in section 2.1. It is characterized by the assumption that the total size, as well as the structure, of the bank's liabilities are exogenously determined and not subject to optimizing behaviour, the problem to be solved being the optimal allocation of the given funds among various assets, with particular attention being paid to the choice between earning assets and reserve (liquid) assets.

Important as this problem is, it is but one aspect of a full theory of the banking firm. First, an analysis of the activities making their appearance on the other side of the bank's balance sheet, and thus of its desired (optimal) liability side structure (including the desired relationship between deposit liabilities and capital account) in principal appears to be equally important as an analysis of bank asset choice for a satisfactory microeconomic foundation of monetary theory, in particular money supply theory. This question will be discussed in section 2.2, still under the assumption that total portfolio size is exogenous.

A complete theory of the banking firm, however, should not only provide an integrated view of the firm's asset and liability choice, but also allow an endogenous determination of the total scale of operation of the firm. A relatively small number of attempts has been made to deal with this problem. Three types of models will be distinguished, depending on different degrees of emphasis. The first, and probably best known among these, assigns a crucial role to the assumption that banks can operate as monopolistic price setters in deposit and/or credit markets. These 'monopoly models' will be discussed in section 3.1. The second type, in contrast, puts heavy emphasis on the assumption of subjective risk aversion on part of the bank (or its owners). These 'risk aversion models' will be reviewed in section 3.2. The third type, finally, can be referred to as 'real resource models', because it is characterized by the fact that, in contrast to most other models, it assigns a prominent role to the 'real resource' or 'real production' aspects of the banking business. This approach will be discussed in section 3.3. The characteristic elements of these different types of approaches are, of course, not mutually exclusive. The classification followed thus has to proceed according to the relative emphasis given to different factors in different models, and some overlapping can obviously not be avoided.

2. Bank portfolio management problems

2.1. Models of optimal asset choice

As mentioned, most models belonging to this group are in essence models of bank reserve and liquidity management. The basic model on which this literature is based can be traced back to Edgeworth (1888). In more recent times this approach has been taken up by a variety of writers.[3]

It essentially treats the bank's reserve and liquidity management decision as a problem of inventory optimization under stochastic demand. The idea of the approach can be summarized as follows.

Basic liquidity management model. Consider a bank which has a given amount of deposits D, and which can choose between two assets, namely reserves R[4] and an earning asset (loans) E. Let r denote the (net) yield on loans and assume that the bank is subject to withdrawal risk, i.e., it has a-priori-knowledge about deposit changes during the planning period (and thus about deposits at the end of the period) in a probabilistic form only, based on past experience. Let X denote the *outflow* of deposits, and thus reserves, during the period, with (estimated) density function $f(X)$. Suppose that the occurrence of a reserve deficiency, i.e., a situation where the reserve loss during the period X exceeds the beginning-of-period level of reserves R,

[3] See, e.g., Orr and Mellon (1961), Porter (1961), Morrison (1966), Poole (1968), Frost (1971), Baltensperger (1972a, b), Ritzmann (1973), Pringle (1974), Hester and Pierce (1975), Koskela (1976), and Niehans (1978).

[4] Note that R must not necessarily represent cash reserves. In some variants of the model [e.g., Porter (1961), and Pringle (1974)], the reserve function is performed by securities ('secondary reserves').

makes necessary certain costly adjustments for the bank (emergency borrowing, or emergency selling of assets). For simplicity, assume that these adjustment costs are proportional to the size of the reserve deficiency, with p denoting the factor of proportionality (i.e., the cost per dollar). The problem is to choose the optimal (beginning-of-period) allocation of the given funds D among reserves and loans.

If the bank believes the loan rate r to be independent of the volume of loans it extends (price taking behaviour in loan market), the two cost items which have to be balanced against each other can be expressed as rR for the opportunity cost of holding reserves, and

$$L = \int_R^t p(X-R)f(X)\,\mathrm{d}X \tag{1}$$

for the expected adjustment cost due to reserve deficiencies (='liquidity cost').

Holding an extra dollar of reserves thus implies a marginal opportunity cost of $r\,(>0)$, but a marginal reduction in liquidity cost $L_R = -p\int_R^t f(X)\,\mathrm{d}X < 0$. Optimization requires minimization of the sum of these two cost elements, i.e., equalization of marginal cost and marginal 'return' of holding additional reserves,

$$r = p\int_R^t f(X)\,\mathrm{d}X. \tag{2}$$

In words, this means that the bank must choose the level of reserves such that $\int_R^t f(X)\,\mathrm{d}X$, the probability of a reserve deficiency, is just equal to the ratio r/p. This condition defines the bank's desired reserves as a function of the data of the model: r, p and $f(X)$.

In discussing this result, care has to be taken in properly interpreting the parameters r and p. It is clear that in order to get a positive value for optimal reserves, we need $\int_0^t f(X)\,\mathrm{d}X > r/p$. If, e.g., $f(X)$ is symmetric with $E(X)=0$, we thus need $1/2 > r/p$, or $p > 2r$. Sometimes, p has been simply interpreted as the discount rate, and r as the loan rate, in which case fulfillment of the condition for a positive level of optimal reserves seems rather unlikely. Furthermore, this interpretation would imply that optimal reserves depend on the structure (ratio) of interest rates only, but not on their absolute level. However, it should be kept in mind that (a) r in this context cannot be identified with the total loan rate, but rather it is the loan rate net of all costs (including administration and information costs) of extending credit, and (b) p cannot be identified simply as the discount rate. Often a bank cannot borrow freely from the central bank, or has an aversion against borrowing from it. A reserve deficiency forces a bank to rearrange its portfolio on short notice. The rate p must reflect all the costs and inconveniences of such rearrangements, including transaction cost.

Given these additional considerations, fulfillment of the condition for a positive level of optimal reserves does not seem unlikely at all anymore. Also, if r and p are functionally related to, but not identical to the loan rate and the discount rate, it becomes clear that optimal reserves depend not only on the structure, but also on the level of interest rates. (These points become more apparent, of course, if real resource and transaction costs are modeled explicitly, as will be the case in some of the discussions in the following sections.)

The type of model just outlined can be and has been modified to take account of a variety of additional elements. Some of these modifications will briefly be summarized.

Declining demand function for loans. If the bank does not view r as a market determined parameter, but rather as being (negatively) related to the amount of credit which it extends, the marginal opportunity cost of reserve holdings is not given by r anymore, but rather by the marginal revenue lost due to reducing E by a unit. Otherwise, nothing is changed. The firm's optimality condition then becomes

$$\delta E(\pi)/\delta R = -\delta E(\pi)/\delta E = -[r(E)+Er'(E)] - L_R = 0. \tag{3}$$

Modifications of this nature are an element strongly emphasized in several models to be reviewed below (see section 3.1). Note also that the deficiency-cost function $p(X-R)$ could be made non-linear in a similar way, without affecting the logic of the model [see, e.g., Poole (1968)].

Reserve requirements. Legal reserve requirements have been disregarded in our discussion so far. They can be easily incorporated into this framework, however. Their major effect is to reduce the critical value of the reserve loss X beyond which a reserve deficiency and corresponding adjustment costs occur. Without reserve requirements, this critical level of X is equal to the beginning-of-period level of reserves R. If, e.g., the legal requirement is that reserves at the end of the period $(=R-X)$ must be at least equal to a specified proportion k of end-of-period deposits $(=D-X)$,[5] a reserve deficiency occurs whenever $R-X<k(D-X)$, or $X>(R-kD)/(1-k)\equiv\hat{X}$. If a reserve deficiency occurs, its size is $X(1-k)-(R-kD)=(X-\hat{X})(1-k)$. The expected value of these costs thus is

[5]The precise effect of including reserve requirements depends, of course, on the detailed specification of the legal requirements, including the definition of what is admissible as legal reserves, how required reserves are to be computed and held, penalties in case of violations, etc.

$$L = \int_{\hat{X}}^{l} p[X(1-k) - R + kD] f(X)\,\mathrm{d}X \tag{4}$$

with derivative

$$-L_R = p \int_{\hat{X}}^{l} f(X)\,\mathrm{d}X. \tag{5}$$

Thus, optimality requires again that the marginal opportunity cost of holding reserves ($= r$, if the bank is a price taker in the loan market) is equal to p multiplied by the probability of a reserve deficiency. This probability, of course, is given by the possibility of X exceeding $\hat{X}$ now, rather than of X exceeding R.

Costs of adjusting to the optimal portfolio (sS-strategies). If (beginning-of-period) adjustments to the optimal reserve position as determined above (denote it as R^*) were costless, the bank would continuously rearrange its portfolio so that it starts each planning period at R^*, independent of the level of reserves 'inherited' from the preceding period (R_0). In the presence of adjustment costs, however, an adjustment to R^* is profitable only if the resulting gain ($=$ reduction in $rR+L$) more than offsets the cost of the adjustment itself. In this case, there exists a range around R^* within which the bank will let its reserves fluctuate freely, without making adjustments.[6]

If adjustment costs are proportional to the size of the adjustment, the boundaries of this range are given by the points where the marginal gain resulting from moving towards R^*, i.e., the resulting reduction of the sum $(rR+L)$, is just equal to the marginal adjustment cost m.[7] If inherited reserves R_0 are outside this range (as a result of a corresponding value of X in the previous period), an adjustment to either the upper or the lower boundary will be made (depending on whether R_0 has left the critical range on the upper or lower side). Further adjustments towards R^* would result in a net marginal loss, taking into account the costs of the adjustment itself.

Furthermore, if adjustment costs also include a fixed element M which is independent of the size of the adjustment, an adjustment is profitable only if the resulting gain covers all the adjustment costs, including the fixed element. If an adjustment is made, it will still lead to the upper or lower boundary of the range just described, for the stated reasons. But R_0 lying outside this range does, in this case, not automatically imply that an adjustment is

[6]For an analysis of reserve management policies under these conditions, see, e.g., Frost (1971), Baltensperger (1972a), and Knobel (1977). In a general demand for money context, see Eppen and Fama (1969), and Miller and Orr (1966).

[7]That is, the upper and lower limits of the range are determined by the two conditions $r + L_R = \pm m$, or $-L_R = p\int_R^l f(x)\,\mathrm{d}X = r \pm m$.

profitable. For this to be the case, the total gain (i.e., reduction in $rR+L$) resulting from bringing reserves up to the lower boundary or down to upper boundary, respectively, must be at least as large as the fixed adjustment cost M, plus the proportional cost term m multiplied by the size of the adjustment. Thus, the size of the range within which no active adjustments are profitable is further increased.

Diversification of earning assets. So far we have allowed the bank to hold just one homogeneous earning asset. This is a simplification, of course. In reality, a bank holds a variety of different types of earning assets. One way to explain asset diversification by a bank is in terms of risk-return considerations along the lines of general portfolio theory (e.g. of the Tobin–Markowitz variety). The application of this approach, which is based on the idea of 'subjective' risk aversion by the bank (or its owners) will be discussed separately below (see section 3.2).

However, even if we stay within the expected profit maximization framework utilized so far, where the firm does not assign a negative value to the variability of profit *per se*, (i.e., as long as there is no feedback from this variability to expected profit), there are still ways to explain diversification of bank assets. One possibility is to extend the model of reserve management discussed above such that those earning assets which are characterized by a relatively high return at the same time are characterized by relatively high conversion cost into cash (i.e., are less 'liquid').[8]

Another way to explain asset diversification which is employed as an important element of some models is to introduce non-linearities by assuming monopoly power of the bank in (at least some of) the earning asset markets. The bank then optimizes the structure of earning assets by equalizing marginal revenues for the various assets. Models emphasizing this element will be discussed separately below (section 3.1).

Instead of introducing non-linearities on the revenue side, it is also possible, of course, to introduce them on the cost side. It has already been mentioned that the real resource cost of producing banking services have been rather neglected in much of the banking literature. Some models where these costs play a prominent role will be discussed separately below (section 3.3), however, so that a detailed discussion of this issue will be postponed, too.

Information costs. Uncertainty, i.e., incomplete information about various aspects of its activities plays a central role in many of the models of bank asset selection which have been mentioned. The degree of uncertainty is treated as completely exogenous in most of these models. One exception is provided by Aigner and Sprenkle (1968), who hypothesize that the bank, by

[8]See, e.g., Ritzmann (1973, p. 162 ff.), and Baltensperger and Milde (1976)

spending resources on the collection of information about its customers, can reduce the *expected* rate of default, thus earning a return on its information collection activity. Their specification is somewhat questionable, though, since it is based on the presumption that bankers always overestimate the 'true' probability of default, so that more information necessarily means a lower default probability. It is not clear (a) why this should be the case, and (b) why extra information should only be valuable if it results in a downward adjustment in the estimated default probability, as their model implies. The cost of informational errors arise because they lead to decisions which ex post will turn out to have been suboptimal. This is equally well possible, regardless of whether we initially have an under- or overestimation of the 'true' probability.

An example of an approach which links the cost of incomplete information and the return to more information with the *variability* of certain random variables (such as default losses and deposit fluctuations), and thus avoids the above problem, is given in Baltensperger (1972b, 1974), and Milde (1976).

Determinants of the distribution of deposit changes $f(X)$. The liquidity management model summarized above did express optimal reserves as a function of the parameters r and p, and the deposit fluctuation distribution $f(X)$. If $f(X)$ can be approximated by a normal distribution (which is to be expected, since X is the sum of a large number of independent changes in different deposit accounts),[9] and if, for simplicity, we assume $E(X)=0$, optimal reserves can be expressed as a multiple b of the standard deviation of X, with b being determined by the ratio r/p,

$$R=b\sigma_X. \tag{6}$$

The distribution of deposit fluctuations $f(X)$ should be expected to depend on the volume as well as the structure of the bank's deposits, of course. We are still treating these as exogenous; but we can nevertheless consider the effects of parametric changes in them on the distribution $f(X)$ and thus reserves R. The following hypotheses appear reasonable [for a formal analysis, see Baltensperger (1972a), and Miller (1975)]:

- An increase in the level of (initial) deposits D will raise σ_X, and thus R, but not in proportion, as long as it is the result of an increase in the number of (somewhat) independent causes of deposit changes (number of accounts).
- A redistribution of a given D in favour of more volatile deposits (e.g., from time deposits to demand deposits) will increase σ_X and thus R, and vice versa.

[9]Clearly, the normal distribution must be viewed as an approximation, though, since the upper limit of X is equal to initial deposits and thus finite.

2.2. Models of liability management

While quite a lot of work has been done on the question of the optimal asset choice of financial firms, relatively little attention has been paid to the structure of the other side of the firm's balance sheet and the question of liability management. Total volume as well as structure of liabilities usually are assumed to be exogenous and not subject to optimizing behaviour.

An argument which is often advanced in support of this assumption is that a bank does have no choice other than simply accepting all the deposits offered to it by the public at the ruling deposit rates. Thus volume and structure of deposits are viewed as being totally demand-determined and, from the point of view of the bank, exogenous. This view is not convincing:

First, it is not clear whether a typical bank is really as powerless and passive with respect to the size and structure of its liabilities as this argument has it. A bank has a variety of possibilities to influence the (relative and absolute) attractiveness of various types of liabilities, and thus the public's demand for them, of which explicitly paid interest is but one. This is manifested, e.g., in the fact that banks are often seen to advertise vigorously for their deposit business.

Second, and more fundamentally: even if the above argument were true, so that a bank has to accept completely passively the deposits which are offered to it by the public – at conditions, including deposit rates, which are, somehow, exogenously determined – it would still be sensible to ask for the volume and structure of liabilities which are optimal and thus *desired* by the bank at these conditions. Only supply considerations of this kind will enable us to know, e.g., whether a given situation has equilibrium character or not, i.e., is acceptable to bank customers as well as the bank itself. They are indispensable, therefore, for a full understanding of bank portfolio behaviour and the structure of deposit rates and other conditions, since these are ultimately determined by supply as well as demand factors in any case. In contrast to a frequently met view, there is no difference at all in this respect between the markets for bank liabilities and any other markets.

We will begin by concentrating on just two items on the liability side of the bank, homogeneous deposits D and equity capital W. Subsequently, we will discuss some elements of the determination of the structure of deposits.

The deposit–capital decision. The bank's capital account and its role has been almost completely disregarded in most analytical models of bank behaviour. Often, it is not mentioned at all, and in the remaining cases it is usually treated as exogenous and does not perform any meaningful role in the model.[10] We will begin with a model which is related to an approach

[10]The following discussion is based on Baltensperger (1973). See also Baltensperger (1972a, b) Niehans (1978, ch. 9), and Taggart and Greenbaum (1978).

which has in recent years been emphasized in the general finance literature. It essentially applies the inventory theoretic approach previously used in connection with reserve management to the determination of the bank's capital account and its supply of debt (deposits).

A bank's capital account performs an important economic function, fundamentally similar to that of liquidity reserves, namely protection against certain types of uncertainty and the associated possibility of emergency adjustments and costs. Consider a bank with a given (beginning-of-period) level and structure of assets A. The bank's income Y from these assets for the decision period is known a priori in probabilistic form only, with (estimated) density function $g(Y)$, partly due to default risk, partly due to uncertainty about (end-of-period) interest rates and asset prices [with $g(Y)$ obviously being dependent on volume and structure of the bank's asset portfolio]. If the bank issues an amount of deposits (=debt) D, and promises to pay an interest rate i on these deposits, its end-of-period indebtedness will be $D(1+i)$. If its end-of-period assets $A+Y$ are less than its end-of-period debt $D(1+i)$, it finds itself with a negative end-of-period net worth, i.e., is in a state of insolvency. The condition for this to occur can thus be summarized as $A+Y-D(1+i)=(Y-iD)+(A-D)<0$, or

$$Y<D(1+i)-A\equiv\hat{Y}. \tag{7}$$

It is clear that, ceteris paribus, the probability of this event is positively related to the size of D, or negatively to $W(=A-D)$. Insolvency is a costly affair for a firm, and thus an event which it prefers to avoid. It forces the firm into costly portfolio rearrangements under a short time constraint and may severely disrupt its regular activities, partly because of a loss of confidence by the public, partly because the bank may be suspended by supervising authorities. Creditor–debtor meetings will have to be arranged, legal services have to be employed, etc. It seems reasonable to assume that these costs of insolvency (which in part represent real resource costs) are positively related to the size of the capital deficiency $(\hat{Y}-Y)$. For simplicity, we will assume in the following a proportional relationship, with the cost per dollar of deficiency denoted as a. The expected cost of insolvency S then can be expressed as

$$S=\int_{-\infty}^{\hat{Y}} a(\hat{Y}-Y)g(Y)\,\mathrm{d}Y. \tag{8}$$

Maybe equally important as the costs of actual insolvency are the costs caused by the firm's efforts to avoid insolvency. A firm will start to rearrange and reorganize its portfolio even before insolvency actually occurs, if its capital position falls below some critical level as a consequence of a 'bad' year, in order to prevent insolvency. Also, the existence of legal capital requirements may impose similar costs on the bank even before insolvency actually occurs, whenever its capital account drops below or near to the required level. However, to keep things as simple as possible, we will let the term S as defined above represent all of these costs.

The bank's optimal decision with regard to its liability structure involves balancing these costs S against the cost of using equity capital rather than deposit funds. Let ρ be the opportunity cost of equity funds, and assume $\rho > i$. The marginal opportunity cost of increasing equity capital W by one dollar then is $\rho - i$. The corresponding marginal return is given by the associated reduction in S,

$$S_W = \int_{-\infty}^{\hat{Y}} a\hat{Y}_W g(Y)\,\mathrm{d}Y = -\int_{-\infty}^{\hat{Y}} a(1+i)g(Y)\,\mathrm{d}Y, \tag{9}$$

and optimately requires

$$\rho - i = -S_W. \tag{10}$$

The optimal demand for equity capital and supply of deposits thus is determined by the rates ρ, i, a and the distribution $g(Y)$, according to this approach.

In this discussion, we have still neglected two important aspects, however, which shall be taken up successively:

(a) Issuing and maintaining deposit contracts requires a real resource expense by the bank. Banks offer to their depositors a variety of services, the provision of which constitutes a major element of the contract agreement between the bank and the depositor (with the nature of these services being different for different types of deposit contracts). This aspect of banking, which has been emphasized, e.g., by Pesek (1970) and Saving (1977), will be taken up in more detail in section 3.3. At this place, we may simply introduce a cost function $C(D)$ as an explicit representation of this element. The critical level of Y becomes then

$$\hat{Y} \equiv D(1+i) + C(D) - A, \tag{7'}$$

and the optimality condition is

$$\rho - (i + C_D) = -S_W, \tag{10'}$$

with

$$S_W = -\int_{-\infty}^{\hat{Y}} a(1+i+C_D)g(Y)\,\mathrm{d}Y. \tag{9'}$$

Furthermore, the costs of holding reserves and adjusting to reserve deficiencies should, according to our previous discussion, be viewed as dependent on D. Since these costs have been dealt with in detail in section 2.1, we will, for simplicity, disregard them in the present context. They would, if taken into account, represent another marginal cost of increasing D.

(b) We have so far treated the interest rate i paid by the bank to its depositors as a given, market determined magnitude, independent of the deposit-equity ratio chosen by the bank. Given the existence of the insolvency cost S emphasized above, this does not seem justified. The possibility of insolvency implies that the *nominal* deposit rate i has to be distinguished from the *expected* rate of return received by depositors t. Given a positive probability of default, the latter is necessarily less than the former. The total interest payments paid out to the depositors of the bank are iD, if $Y > \hat{Y}$, and $iD-(1+a)(Y-\hat{Y})$, if $Y<\hat{Y}$,[11] with expected value

$$iD - \int_{-\infty}^{\hat{Y}} (1+a)(\hat{Y}-Y)g(Y)\,\mathrm{d}Y = iD - ((1+a)/a)S. \tag{11}$$

Thus, the expected rate of return paid to depositors is

$$t = i - \{(1+a)/a\}\cdot S/D < i. \tag{12}$$

Obviously, given the nominal rate i, t is a function of D. However, if the banking system is characterized by a reasonable amount of competition, the bank will be forced to grant the depositors a compensating adjustment in the nominal deposit rate i whenever the effective rate t is lowered due to an increase in the deposit-equity ratio. It should be noted that this has nothing to do with imperfect competition in deposit markets, but would be true particularly in a competitive environment. An increase in D, given $g(Y)$ implies a lessening of the quality of the bank's deposits. The bank thus cannot expect to sell these deposits at the same price as higher quality deposits.[12] We thus must see the deposit rate i as being functionally related to the firm's financial structure. We can express this in form of a 'price

[11]Assuming, for simplicity, that the minimum value is such that this expression remains positive.

[12]See Baltensperger (1976), and Barro (1976) for a discussion of credit market problems along these lines.

function' which assigns a price i to each deposit quality, and thus to each level of D,

$$i=i(D). \tag{13}$$

In a competitive environment, this function represents a market determined datum to the firm. If we disregard the effect of a change in D on the variability of the effective interest return to the depositors, or assume that the latter do not pay any attention to this variability, the market will see to it that a change in D (or $W=A-D$) will always be compensated such through an adjustment in i that t remains constant (as long as market constellations stay constant). The price function $i(D)$ is then implicitly determined by the expression for t [eq. (12)].

Considering that the use of equity capital implies opportunity costs equal to $\rho W=\rho(A-D)$, the firm's expected profit then can be expressed as

$$E(\pi)=E(Y)-tD-C(D)-S-\rho(A-D). \tag{14}$$

The optimal liability structure is reached at the point where

$$\mathrm{d}E(\pi)/\mathrm{d}D=-t-C_D-S_D+\rho=0,$$

or (15)

$$\rho-t=C_D+S_D.$$

That is, the bank will expand its deposit production up to the point where the 'marginal cost' of producing deposits $(=C_D+S_D)$ equals the corresponding 'marginal revenue' $(=\rho-t)$. An interior optimum, of course, also requires that the two functions intersect each other from the right side, i.e., $C_{DD}+S_{DD}>0$. Otherwise, corner solutions would be obtained.[13]

Deposit insurance. It seems worthwhile to briefly consider the institution of deposit insurance in the present context. From the point of view of the depositor, deposit insurance makes all deposits equally attractive, independent of the bank's insolvency risk, and it thus removes the necessity of a risk premium in the deposit rate. Thus, in this case $i'(D)=0$.

An efficiently organized insurance would graduate the premiums paid by the bank to the insurer according to insolvency risk and thus liability structure [given bank assets and $g(Y)$]. The bank would then simply pay a risk premium to the insurance company instead of depositors. To the extent that this is the case, the nature of the optimization problem of the bank is

[13]For a further discussion of these second order conditions, see Baltensperger (1973, p. 153).

unchanged. If the insurance company can offer a lower risk premium for some reason than the premium which the bank would have to pay to depositors directly otherwise, deposit production is made less expensive, of course, with a resulting change in the optimal capital–deposit structure.

The situation is somewhat different, if insurance premiums are determined independently of insolvency risk, e.g. as a fixed amount q per dollar of deposits. In the bank's expected profit function, the premium payment qD then replaces the risk compensation $((1+a)/a)S$ otherwise paid to depositors via price function (12), that is, we have

$$E(\pi)=E(Y)-iD-C(D)+S/a-qD-\rho(A-D) \tag{16}$$

in place of (14), with $i'(D)=0$. The condition for an (interior) optimum then is

$$\rho-i=C_D+q-S_D/a. \tag{17}$$

It is clear that under these circumstances $C_{DD}>0$ is a necessary condition for an interior optimum. Otherwise, corner solutions would be unavoidable, and a positive capital position would have to be forced upon the bank by legal restrictions (which are not infrequently found in this context).

An identical premium rate *q for all banks*, furthermore, obviously favours the relatively risky firms, i.e., those with a relatively high σ_Y, at the expense of those with a relatively low σ_Y. In a context where the bank's assets are endogenous choice variables, such a system results in a change in asset structure towards the relatively risky types of assets. This 'adverse selection' phenomenon is well known from the U.S. deposit insurance system.

Extensions and other approaches. Before we turn to the next topic, a few brief remarks concerning various possibilities to extend this approach should be made. One such possibility is to incorporate information collection and diversification of assets as additional endogenous elements. To the extent that these activities lead to a reduction of σ_Y, they will affect the optimal deposit/capital decision. Obviously, the costs of these activities have to be balanced against their returns. See Baltensperger (1972b), and Baltensperger and Milde (1976). Another possibility is the explicit incorporation of regulations and legal restrictions. Some authors discuss the bank's deposit/capital decision in terms of a framework relying exclusively on regulation. See, e.g. Mingo and Wolkowitz (1977), and Santomero and Watson (1977). Such an approach, although important in a context with deposit insurance, is less basic than the one outlined above.

Pringle (1974) has also discussed the bank's capital decision, but in a different model which disregards insolvency cost S as well as service production cost C, which were stressed above. His model treats the problem essentially as one of liquidity management, under the assumption of exogenous (but stochastic, i.e., uncontrollable) deposits and a less than perfectly elastic loan demand function facing the bank. The bank can obtain additional funds, however, in the form of equity capital, at a given marginal cost β. His model, therefore, is more a variation and extension of the liquidity management approach discussed in section 2.1.[14]

The firm's balance sheet constraint is $Z+E=D+W$, where 'liquid' assets Z in his model are not cash reserves, but securities (which can be held in negative amounts, representing bank borrowings). Deposit changes during the period, and thus end-of-period deposits are stochastic (but exogeneous) as in section 2.1. The bank chooses E and W (and thus initial Z) at the beginning of the period. If end-of-period deposits are less than $E-W$, it has 'used-up' its inventory of liquid assets and must borrow, at a given marginal cost p.

Essentially, since in this model the firm has no choice regarding D, it uses equity capital as a supplementary source of funds, up to the point where its (constant) marginal cost is just equal to the (declining) marginal revenue on loans. Of course, the use of equity capital is efficient only if its cost β is less than the marginal borrowing cost p, which is an assumption of the model.

Thus, what the Pringle model says is, in essence, that given fixed deposits, the bank will raise additional funds from the least cost source and make additional investments, as long as the marginal return from the latter exceeds the cost of financing. This least cost source of financing is assumed to be equity financing. In this way, the model does yield a determinate solution for the firm's capital position, without taking into account its risk bearing function. However, if the least cost source of financing were assumed to be something different, the firm in this model would use no equity capital. Also, if the firm's demand function for loans is such that, even with a zero equity capital (i.e., investing just the given deposit funds) the marginal revenue from loans is already driven down to or below β, the firm would use no equity capital. Furthermore, if the marginal revenue from extending loans is constant (price taking behaviour in loan market), the optimal capital position is indeterminate (either zero, or infinite, or irrelevant). Thus Pringle's approach to the determination of the bank's capital position is less basic than the one outlined above, in the sense that his bank uses equity capital only as a supplementary source of capital, because other sources of funds are assumed to be either exogenously fixed (deposits) or more expensive ($p>\beta$).

[14]These remarks refer to Pringle's formal model (his section I). On some further aspects of his discussion, emphasizing the maturity structure of liabilities, see footnote 21 below. It might also be added that his model allows for subjective risk aversion; but this is of no particular relevance for the essence of his arguments in the following discussion.

The deposit structure. So far, we have allowed the bank to issue only one type of deposits. The analysis can be extended to include different types of deposit liabilities. For simplicity, we will conduct the discussion in terms of just two types of deposits, say, demand deposits (D_1) and time or savings deposits (D_2). Generalization to more than two deposit types D is unproblematic, but does not yield additional insights of substance.

Our previous discussion has assigned an important role to the real resource cost of producing deposits (an element which will be emphasized again in section 3.3 below). These costs, together with the costs of liquidity management (emphasized in section 2.1 above) are also of primary importance in the determination of the structure of deposits. One of the main differences between different deposit categories clearly is the different nature of the services associated with the respective deposit contracts, and of the underlying production technologies. Without going into these matters in great detail, let us express the total resource cost of producing deposit accounts and the associated services as depending on the levels of both types of deposits (with positive and increasing marginal cost for each deposit type),

$$C=C(D_1,D_2). \tag{18}$$

Suppose that the interest rates paid by the bank on the two deposit types are i_1 and i_2, respectively, and assume that the firm treats these as market determined parameters. For simplicity, neglect the problem of equity capital and insolvency cost discussed in the preceding part of this section,[15] and treat the problem of reserve and liquidity management as already solved, yielding, in accordance with the discussion in section 2.1, both R and L as a function of σ_X, and thus

$$R=R(D_1,D_2) \quad \text{and} \quad L=L(D_1,D_2). \tag{19}$$

The bank's expected profit function then is

$$E(\pi)=E(Y)-i_1D_1-i_2D_2-C(D_1,D_2)-L(D_1,D_2), \tag{20}$$

with earning assets $A=D_1+D_2-R$ fixed exogeneously. The optimal structure of deposits then is determined by the condition

$$(i_1+C_1+L_1)/(1-R_1)=(i_2+C_2+L_2)/(1-R_2). \tag{21}$$

[15]An integration of these elements with the present model is possible, but at the expense of quite a bit more complexity.

where subscripts 1 and 2 denote partial derivatives with respect to D_1 and D_2, respectively. The appropriate second order conditions must also be satisfied, of course. Otherwise, the optimum is given by a corner solution.

If the resource cost C, as well as the dependence of the liquidity management cost (reflected in L and R) on deposit structure is neglected [as is the case, e.g., in Klein (1971), as discussed in section 3.1 below], a determinate interior optimum obviously cannot exist as long as i_1 and i_2 are market determined parameters. An interior solution can be generated in this case by making the deposit rates a function of the respective deposit levels, i.e., postulating monopoly power and thus price setting behaviour of the bank in its deposit markets. Whether this assumption is justified or not with respect to deposit markets is an open question and probably depends on specific circumstances. (It might be noted, however, that the mere fact that these markets do not precisely satisfy all the requirements of a textbook perfect competition market model does not yet imply that the monopoly model is better than the competitive model. There are virtually no real world markets which precisely meet all the requirements of the former, either – both models are extremes, and the question is which one is the better approximation.) In any case, the assumption of monopoly power can easily be incorporated into the analysis by simply replacing the levels of the interest rates with the respective marginal interest expenses for both deposit types in eq. (21). (By choosing D_1 and D_2, the bank then has also selected i_1 and i_2, of course.) It is clear that in this case a well defined interior optimum is possible even if C, L and R (or their dependence on deposit structure) is neglected. This element of deposit rate and structure determination has been emphasized, e.g., in the models by Klein (1971) or Monti (1972).

3. Complete models of the banking firm

The models discussed in the sections 2.1 and 2.2 of this paper are partial (portfolio structure) models dealing with questions of either asset choice (section 2.1) or liability management (section 2.2). This is comparable to the analysis of the minimum cost production of a given level of output in the general theory of the firm. A complete theory of the banking firm, of course, has to go beyond that and explain not only the bank's asset and liability choices and their interaction (if any), but also the determination of the total size of the firm. There are a number of different (but not mutually exclusive) factors which can be used to explain the scale of the firm: monopolistic market forms, risk aversion, and the real resource cost of producing banking services. We will distinguish three different approaches, depending on which of these factors is emphasized most.

3.1. Monopoly models

Most of our previous discussion has been in terms of a banking firm which behaves as a price taker in a competitive environment. Many authors, however, view the presence of monopoly power as something characteristic for banking markets, and thus prefer to work in terms of models incorporating this assumption. (On the appropriateness of this view, see the comments at the end of the preceding section.)

There is one set of models of the banking firm in which the assumption of monopoly power is assigned a dominating role for the logic of the model, in the sense that it explains not only firm scale, but also portfolio structure virtually on its own account. This group includes, e.g., the well known model by Klein (1971), as well as the model by Monti (1971, 1972).[16] These models are characterized by the practically complete neglect of the resource cost aspects of the banking business and, in essence, determine bank scale and portfolio structure via (net) revenue maximization along market determined demand functions by the public for bank products. Although this procedure does allow a solution to the problems of firm scale and portfolio structure, it does not *by itself* appear to be a very satisfactory solution as long as resource cost aspects are neglected, since it cannot tell anything about the production and supply characteristics of bank services. This is manifested by the fact that models of this sort completely break down if the firm is forced to behave as a price taker rather than a price setter (whereas an approach based on resource cost elements – at least in principle – can function regardless of form of market behaviour). It should be emphasized that this is not meant to be a critique of the assumption of monopolistic forms of behaviour in bank markets. Maybe this is the better approximation to the real situation in many cases. The point is that a model which functions *only because* of this assumption is lacking something and cannot provide a full and satisfactory understanding of bank behaviour.

The essence of the Klein model can be summarized quite easily, since it has a basically rather simple structure. The bank can choose among three assets: cash reserves, government securities, and loans. On the liability side, three items are again distinguished: two types of deposits, called demand deposits and time deposits, and equity capital. The latter is exogenously fixed and performs no real function in the model (i.e., it could be dropped

[16]A number of other authors also employ the monopoly power assumption, e.g., Towey (1974), Sealey and Lindley (1977), and Pringle (1974). However, in these models this assumption does not play as dominant a role as in those mentioned above, since they also include either resource cost or risk aversion considerations. A group of models which should also be mentioned in this context is the literature dealing with savings institutions in the U.S.: e.g. Goldfeld and Jaffee (1970), Slovin and Sushka (1975). This literature will not be pursued further, since it deals with rather special (U.S.) institutional problems, but it typically employs an approach of the type discussed here.

without any consequences for the logic of the model and its solution). The bank is assumed to maximize expected profit.[17]

The model determines the optimal structure of assets and liabilities (apart from the exogenous capital account) as well as the total size of the bank, under the assumption that it acts as a price taker in the market for government securities, but as a monopolistic price setter in the markets for bank loans as well as the markets for both deposit types. That is, the (average) rate of return on loans is considered by the bank to be negatively related to the amount of credit extended: $r_E = r_E(E)$, with $r_E' < 0$; and the (average) rates of interest which the bank has to pay on the two deposit types are viewed as being positively dependent on the levels of the respective deposit types: $i_1 = i_1(D_1)$ with $i_1' > 0$, and $i_2 = i_2(D_2)$, with $i_2' > 0$. The bank thus is seen as a monopolist optimizing along the loan demand and deposit supply curves of the public.

The bank's demand for reserves is determined by use of an inventory management approach with stochastic deposit withdrawals, as summarized in our section 2.1. The treatment of reserves is somewhat questionable, though, because the reserve flow distribution $f(X)$ is assumed to be a fixed distribution, independent of the composition of deposits, and homogeneous of degree one in total assets. This has the implication of making total reserve holdings R as well as the expected adjustment cost term L linear homogeneous in total portfolio size and completely independent of deposit mix. For an alternative treatment of these relationships, see Baltensperger (1972a), and Miller (1975).

Given these elements, the workings of a model of this type can be understood quite easily. Since the rate of return on government securities r_g is exogenous, it follows that the bank will extend loans until the resulting marginal revenue is equal to this exogeneous rate: $r_E + E r_E' = r_g$. This determines the bank's supply of loans. Furthermore, the bank will sell deposits (of either type) until the corresponding marginal expenditure is equal to $\hat{r}_g$ (= the exogeneous rate r_g, adjusted for the marginal cost of liquidity management, a factor independent of portfolio size and deposit structure under the assumption summarized above): $i_1 + D_1 i_1' = i_2 + D_2 i_2' = \hat{r}_g$. This determines the levels of D_1 and D_2 (or, expressed alternatively, the rates i_1 and i_2). These levels, together with the exogeneous capital account, determine total portfolio size. As explained above, reserves are determined as a given proportion of total portfolio (the size of this proportion depending on the parameters of the reserve flow distribution, r_g, and p, but independent of volume and structure of deposits). This determines total earning assets,

[17]The model by Monti is similar to Klein's, but does also consider the implications of different objective functions for the bank, e.g., size (deposit) maximization instead of profit maximization.

and thus, since the loan volume E has already been determined, the bank's demand for government securities.[18]

The Klein model has a number of weaknesses which make it – as it stands – less than completely satisfactory (which does not exclude, of course, that its elements can usefully be included in a more complete model). It has been mentioned already that it relies exclusively on the assumption of monopoly power by the bank. Klein's bank is in the market (i.e., earns a positive profit) only because of its monopoly power. The model breaks down completely as soon as the bank has to operate as a price taker in competitive markets. The bank would under such circumstances hold but one asset (either government securities or loans, depending on whether $r_g \gtrless r_e$), and it would issue one type of deposits only (D_1 or D_2, depending on whether $i_1 \gtrless i_2$). Its total scale, furthermore, would be either zero or infinite or undefined, depending on whether the higher asset rate is less or more than or equal to the lower deposit rate, corrected for liquidity management cost. Somehow, it seems that such an approach is not quite satisfactory, regardless of whether actual banking markets are typically monopolistic or not.

What is missing is a satisfactory analysis of the costs of banking activities. In connection with deposits, one should take into account the following aspects:

The cost of liquidity management should be included in a more meaningful way, by (at least) making it dependent on the structure of deposit liabilities (and possibly not homogeneous of degree one in scale). Differences in withdrawal risk are certainly one of the main distinguishing characteristics

[18]This summary of the model differs in one point from the exact specification by Klein. He actually made the loan rate r_E dependent not on the total volume of loans, but on the share of loans in the total portfolio. This seems peculiar in a model where total size is not fixed, but variable and endogeneously determined. It implies that the bank can sell loans at unchanged conditions, if only its portfolio structure remains unchanged, regardless of whether its total credit extension is $1 million or $500 million. On the other hand, the bank has to accept a worsening of the terms at which it can extend credit as a consequence of even the most marginal change in its asset structure. It seems that the logic of the model would require that the loan rate is viewed as a function of the total volume of loans, rather than their share, precisely as deposit rates are viewed as being dependent on the total volumes of deposits, rather than their share in total liabilities. If portfolio size is fixed, both specifications are equivalent, of course. But the Klein model explicitly attempts to determine portfolio size endogeneously. This specification problem has consequences for one aspect of the model's solution. With Klein's exact specification, the loan market optimality condition determines the share of loans, rather than the absolute level (as is the case with the alternative specification used above). Since deposit levels and reserve holdings are determined precisely as explained above, the only resulting difference is that total earning assets are distributed among securities and loans in a different way. With Klein's exact specification, the share of loans in total portfolio is independent of total portfolio size (and thus of the public's deposit supply functions), while with the alternative specification used above, this is the case for the level of loans (the remaining part of earning assets in both cases being held in the form of government securities.

of different types of deposit contracts, e.g. demand deposits versus savings deposits.

- The cost of producing and maintaining deposit contracts and the associated services should be taken into consideration, these again being among the characteristic differences which allow a meaningful distinction between different deposit types in the first place. Only by taking into account elements of this type can we meaningfully discuss the determinants of relative deposit supplies and rates. In Klein's model, there exists no difference between the two deposit types from the point of view of production and supply, the only differences between them being differences in demand.

For bank assets, again, the costs of producing and maintaining different types of credit contracts, as well as differences in default risk and consequently in insolvency costs (S) should be considered. Also, differences in the cost of adjustments when facing liquidity problems (i.e., differences in conversion costs into cash), as suggested in section 2.1, can be important. Inclusion of these elements allow a more meaningful differentiation, between different bank assets than is possible in Klein's model, where – from the point of view of production or 'supply' – there again exist no distinguishing features between them, the only such differences being differences in the market demand curves which the bank faces.

These aspects are important in connection with two results which were stressed by Klein and discussed repeatedly in the literature subsequently [see, e.g., Pringle (1973)]. These are (a) the question of the independence of asset and liability management, and (b) the question of the determination of relative deposit rates.

(a) Klein argued that the bank's optimal asset choice is independent of the optimal liability choice and thus of deposit market characteristics and deposit rates paid. He linked this with the discussion about deposit rate ceilings and prohibition, pointing out that the original justification for these regulations involved an argument connecting the composition and riskiness of bank asset portfolios with the degree of competition in deposit markets and thus the deposit rates paid by banks. He attempts to prove this argument wrong by demonstrating that in his model the shares of all three assets are independent of the size and structure of deposit liabilities and thus of deposit market features.

However, while it is true that this result does hold in his model (given his exact specification, see footnote 18), it is clearly the consequence of some very special and questionable assumptions and falls, as soon as the model is reformulated along the lines suggested above. E.g., we need only make $f(X)$, and thus R and L, dependent on deposit composition, as would appear most reasonable, to get a dependence of the shares of earning assets versus reserve assets on deposit structure. Introducing other cost elements including resource cost and insolvency cost, will bring further joint elements into the determination of optimal bank assets and liabilities. This will be stressed in the approaches to be discussed in sections 3.2 and 3.3 below. It should be emphasized, however, that all this is not meant to be a defense of the argument in favour of deposit rate regulation referred to above. That argument is silly, and even if it were correct, it could not justify deposit rate regulation any more than it justifies regulation of the price of any other bank expense, e.g., salaries. This discussion is aimed at the question of the relationship between optimal asset and liability structures only.

(b) In Klein's model, since the two deposit types are in no way distinguishable from the production point of view, it turns out, of course, that the optimal setting of the two deposit rates depends exclusively on the properties of the respective deposit supply functions of the public. More specifically, the relative levels of these two rates depend on the relative supply elasticities of the public for the two deposit types. If these are identical for both, both deposit rates would have to be equal. Since in the U.S. the explicit payment of interest on demand deposits is illegal, Klein cannot further disregard the resource costs of producing deposits (which otherwise are totally disregarded in his model) in this context. He calculates a 'cost rate' for demand deposits, by comparing the direct expenses allocated to demand deposit accounts, net of the bank's service charge income, with the stock of demand deposits (based on Functional Cost Analysis Data). The implicit rate obtained in this way is lower than the average rate paid on the time deposits. He explains this by arguing that banks typically face a less elastic supply in demand deposit markets as compared to other deposits markets. This may be true. However, it seems that another factor which can explain (at least) part of this difference is the difference in the cost of liquidity management (our terms R and L) between the two deposit types, which is disregarded by Klein. In view of the different nature of demand and time deposit contracts, this difference should be quite substantial. Furthermore, the statistical calculation of cost rates for individual banking services will always be problematic, due to the existence of a very substantial proportion of joint ('indirect') costs which cannot clearly be allocated to any particular bank activity.[19] Also, as Klein himself notes, another way for banks to pay their depositors in the presence of rate limitations is via preferential loan treatment. Klein's conclusion that demand deposits are 'more profitable' than other deposits is subject to some doubts, therefore. Nevertheless, these remarks are not meant to imply that questions of market form and differences in supply – or demand –

[19]On this question, see Adar, Agmon and Orgler (1975).

elasticities facing the bank may not be important. Their purpose is simply to draw attention to some further aspects of the problem.

In summing up our discussion, we can say that the approach reviewed in this section puts too much of a burden on the assumption of monopolistic market forms, at the expense of other relevant factors. This does not represent a disagreement with the view that financial institutions can perform a role in 'imperfect markets' only, and that otherwise they have no reason to exist [see, e.g., Pringle (1973)]. I totally agree with this (see the introduction of this paper), if by 'imperfect markets' we refer to the existence of incomplete information, uncertainty, adjustment costs, and the like. But this does not necessarily imply that the traditional monopoly model is always superior to the model of the price taker in describing and analyzing these markets. Markets featuring these elements can be atomistic markets characterized by a high degree of competition nevertheless.

Finally, it might be noted that a monopoly bank in the sense of being the only bank in the whole system is something different from a monopoly bank as discussed above. Such a bank presumably would have to take total reserves (or the monetary base, discount rate, or whatever else the monetary authorities fix) as given, take into account the interactions of the complete financial system, especially in the form of 'redeposits', in its optimization process [see Aftalion and White (1977) for a discussion of such a bank]. The bank in Klein's model is not a monopolist in this sense. It expects no redepositing, i.e., given the deposit rate, it expects no change in the public's supply of deposits as a result of an increase in its credit extension. Thus it has to be thought of as a monopolist in its own local market which, however, competes for reserves with other banks in other local markets. If redeposits are included (Aftalion and White), the question arises whether this represents an externality to the bank. This is the case only if, in equilibrium, deposits are obtained by the bank at lower costs (at the margin) than other funds. If effective deposit rates are not artificially fixed, there exists no reason for this to be so.

3.2. Risk aversion models

The models reviewed so far have worked with the assumption of expected profit maximization. This does not mean that risk is not important in these models. In fact, risk does play a crucial role in all of them, as should be clear from the preceding discussion. However, it does mean that the firm is assumed to care about risk only to the extent that it feeds back into, and thus is reflected in expected profit itself, e.g., via the possibility of costly emergency adjustments. The firm is assumed not to care about variability in its income and profit for its own sake.

Formally, this may be interpreted as implying a linear utility function for bank profit. This can be viewed as a simplification and approximation. In principle, the approach can be generalized to let the bank maximize the expectation of a utility function $U(\pi)$, instead of expected profit $E(\pi)$. Of course, this makes the detailed analysis much more complicated. Since it is not always clear how much of substance is gained in exchange, expected profit maximization may be justifiable in many cases as a first approximation, even if basically a more complex utility function is viewed as being more appropriate.

However, there exist a number of models which approach the theory of the financial firm by applying the general theory of portfolio behaviour under assumptions of subjective risk aversion, and which thus obviously are based on the latter assumption in a crucial way. These models represent another way of dealing with the question of asset and liability interactions and the determination of firm scale. Such an approach has been pursued by Parkin (1970), Pyle (1971), and Hart and Jaffee (1974).

These authors apply general portfolio theory (of the Tobin–Markowitz variety mostly), i.e., they treat the financial firm simply as a collection of financial assets with exogeneous (but stochastic) rates of return, and with liabilities treated as negative assets. The most basic of these models is the one by Pyle, since he discusses, in this type of a framework, the conditions under which intermediation will take place, i.e., under which a firm will sell deposit liabilities (= negative amounts of a specified asset) in order to acquire (positive amounts of) other financial assets.[20] His model has not been developed for commercial banks specifically, but is of interest in this context nevertheless and provides an interesting contrast to the approaches discussed above.

Pyle completely disregards liquidity and solvency considerations, as well as real resource costs. He considers an intermediary which has the choice between three securities: a riskless security, plus two securities with an uncertain yield over the given decision period, referred to as 'loans' and 'deposits' (all of which can, in principle, be held in positive as well as negative amounts). The question is under what circumstances the firm is willing to sell risky 'deposits' in order to buy risky 'loans'. Letting α_0, α_1 and

[20] More basically, he attempts to answer the question of which assets (in positive or negative amounts) the financial firm will hold. Cootner (1969) gives another discussion of the same approach. Parkin (1970), and Hart and Jaffee (1974), on the other hand, assume that the menu of assets and liabilities which are held by the firm is institutionally given, and go on to discuss the properties and comparative statics of portfolio choice under these conditions. Parkin's study is empirically oriented and applied to U.K. discount houses; Hart and Jaffee discuss theoretically the incorporation of certain institutional features special to depository institutions, such as reserve or liquidity requirements (but only as exogenous constraints), or constraints on the admitted range of assets and liabilities. See also Kane and Malkiel (1965) for another approach involving risk aversion.

α_2 denote the amounts of the three assets, with $\alpha_0+\alpha_1+\alpha_2=0$, and r_0, r_1 and r_2 the corresponding yields per decision period, with r_0 certain, and r_1 and r_2 random variables with given expectations and (joint) distribution, the firm's profit for the decision period is

$$\pi=r_0\alpha_0+r_1\alpha_1+r_2\alpha_2=\alpha_1(r_1-r_0)+\alpha_2(r_2-r_0). \tag{22}$$

The firm maximizes expected utility of π, where the utility function $U(\pi)$ is characterized by risk aversion (i.e., is concave). The uncertainties about the decision period yields r_1 and r_2 (as well as the relationship between them) are seen as arising from differences in the lengths of the respective maturities and the decision period.

Pyle shows (not quite surprisingly) that, in the case of stochastic independence between asset and liability yields, intermediation ($\alpha_1>0$, $\alpha_2<0$) will occur only if there is a positive risk premium on loans $[E(r_1)>r_0]$ and a negative risk premium on deposits $[r_0>E(r_2)]$. In other words, if there is a positive expected yield difference between assets and liabilities. If there is a positive dependence between the two rates r_1 and r_2, these conditions are still sufficient for intermediation to occur. This is seen clearly by comparison with the independence case: positive dependence is a more favourable environment for the intermediary than independence, since the probability of realizing a negative yield differential falls. Thus, intermediation can in this case be profitable even if there is a non-negative risk premium for deposits $[E(r_2)\geqq r_0]$, as long as the positive effect of (positive) dependence is strong enough. The probability for intermediation to be profitable, of course, increases ceteris paribus with the expected yield difference $E(r_1)-E(r_2)$, and with the degree of positive correlation.

In this model, the optimum asset and liability choices are clearly interdependent (except in the case of independence between r_1 and r_2, where they happen to be separable). Note also that in this model the assumption of risk aversion is crucial to ensure a finite firm size, as well as a place for the riskless asset in the portfolio. What is basically involved here is an arbitrage process exploiting the difference between $E(r_1)$ and $E(r_2)$ which is checked by the resulting increase in risk. As the degree of risk aversion declines towards zero, it becomes profitable to engage in this type of arbitrage ad infinitum – presumably until the yield differentials which started it are eliminated.

This raises the question of what gives rise to these differentials in the first place. It is clear that Pyle's results concerning the conditions under which intermediation will take place in his framework are correct. What is less clear – and not answered by Pyle, but in a sense the more basic aspect of the question he asked – is why such conditions should ever come into existence (and persist). Why will the bank find customers willing to hold a financial asset ('deposits') at an expected rate below the one which the bank can obtain itself, and others which are willing to indebt themselves to the bank at an expected rate exceeding the one which the bank has to pay itself? That is, the approach does not really, in this basic sense, make clear what makes the intermediary come into existence, and thus what function it performs.

In the context of the Pyle model, which disregards questions of resource (transaction, information, etc.) costs and liquidity and insolvency considerations, the only answer which can be given to this question is in terms of risk aversion. Without risk aversion arbitrage would (in the absence of transaction cost etc.) immediately eliminate the rate differentials which form the basis of Pyle's model. With risk aversion, this will not be the case, of course. For every individual, it will then be 'profitable' (in terms of expected utility) to engage in arbitrage up to a certain point only (determined by its degree of risk aversion). In a way, then, each of these individuals can be viewed as an intermediary in the sense of the Pyle model. (The question is whether the number of such individuals would not be sufficient to completely eliminate the need for further arbitrage, i.e., effectively eliminate expected yield differentials even in this case.)

Of course, alternative answers to this question can be given in terms of specialization and resource costs (including transaction and information costs), or in terms of the consolidation of liquidity as well as insolvency risks and costs. These are the elements which were stressed in the approaches discussed in the first sections of this paper (and which will again be stressed in the concluding section). Realistically, these must be viewed as major sources for the persistence of rate differentials which cannot be disregarded by a satisfactory analysis of intermediation.

In connection with the question asked by Pyle, it does not seem sufficient to just take the various yields and their interrelation as representing part of the state of nature. Rather they should somehow be endogenous to the model, not necessarily in the sense of price setting, but in the sense in which even in a purely competitive system equilibrium prices are determined by cost as well as demand conditions. This requires that somehow the nature of the services produced by the firm makes an appearance in the model, in one form or another. This is not the case in models which restrict themselves to a direct application of traditional portfolio theory to the financial firm, and it is hard to achieve in such a framework. Of course, it is true in a formal sense that a financial firm is nothing but a collection of assets and liabilities. But so is General Motors, and any other economic unit. It does not follow from this that a direct application of portfolio theory (of, say, the Tobin–Markowitz kind) does provide a full understanding of all activities of General Motors (which, again, does not mean that it cannot contribute something to such an understanding). What is unsatisfactory is the assumption of exogenous (albeit stochastic) *net* yields for all the different assets and liabilities. One of the major tasks of a theory of the firm must be to explain how the firm combines resources of various kinds in order to generate these net yield and profit streams. This requires going beyond a pure traditional portfolio approach. These comments, however, are not meant to imply that portfolio theory and risk aversion cannot play a useful role as an element of a more complete model.

There is another point concerning the Pyle model which deserves some comment. It was mentioned before that uncertainty about deposit rates in that model is seen to result from differences in the length of the decision period and the maturity of deposits. If the decision period is longer than deposit maturity, the decision period rate of interest on deposits is uncertain because the future course of deposit rates is uncertain. In all this, the length of the decision period, and its relationship to the maturity of deposits and assets is exogenous (in the sense that a commitment to hold α_1 and α_2 for the entire period has to be made at the beginning). The relationship between these periods is a much neglected question. But suppose that the length of the decision period is chosen as being equal to the maturity of the deposit liabilities (and that, as in Pyle's model, just one deposit maturity is allowed). Then there is no uncertainty about the (decision period) rate of interest on deposits r_2 and, of course, there can be no dependence between r_1 and r_2. The (decision period) rate of return on assets is uncertain (unless asset maturity is just equal to deposit maturity), but the portfolio decision is renewed in regular intervals, with no uncertainty about deposit rates over each of the successive decision periods. As long as the bank is restricted to the given menu of assets (and liabilities) and thus to just one deposit maturity, this seems superior to making a decision at the beginning over several deposit maturity periods together (which was the justification for viewing deposit rates as random).

However, if deposit instruments as well as assets with various maturities (e.g., 'short' and 'long') are available, the question of deposit rate uncertainty becomes meaningful again. This case is discussed by Niehans and Hewson (1976, appendix), in a 'term structure' model where the bank has to choose between long- and short-term deposits and loans and the direction and extent of maturity transformation (i.e., different amounts of various maturities on the two sides of the balance sheet).[21] The problem then becomes one of evaluating the (uncertain) future short rates and their relationship to the current rate structure. Maturity transformation will take place if there are deviations from 'term structure parity' which can be exploited. As long as such deviations persist, arbitrage will be profitable, if the (transaction) cost of the arbitrage process itself and the associated risk of an open position are covered by the expected gain.

[21] Pringle (1974, pp. 790–791) also has some interesting thoughts along these lines, by treating the question of bank capital management as one of the maturity structure of liabilities.

Even if only one deposit maturity is available, it can be efficient to make portfolio decisions over several maturity periods together, and consequently treat deposit rates as uncertain, if the costs of making decisions and portfolio adjustments are taken into account. This, however, leads back to the type of approach employed in the preceding sections, where variability is undesirable because of (real) expected costs associated with it, rather than a (subjective) dislike of variability per se.

In connection with models based on subjective risk aversion, the question of the time horizon over which the firms expect to operate and its relationship to the length of the decision period of the models under discussion should be raised, finally. Even if basically risk averse behaviour for the bank or its owners is assumed, it is, presumably, not the variability of a single period (e.g., yearly) profit figure which is of primary relevance in this respect, but rather the variability (i.e., unpredictability) of profit over the entire period during which the firm expects to operate. Expressed differently, the relevant variable should be the variability of 'terminal wealth' or, in terms of a short period (e.g., yearly) profit figure, the variability of *average* period profit over the long run, which will tend towards zero, if the horizon is long enough and single period results are independent. In other words, as long as the time horizon is long enough, expected period return is an accurate measure of what the firm will actually earn per period, on the average, and in this sense is an appropriate magnitude for firm owners to be concerned with, maybe more so than a criterion involving also a (single period) variability figure. Note again, however, that expected profit itself must incorporate the fact that extreme realizations of certain random variables, such as deposit outflows, default losses, or profits, can cause emergency problems and thus costs to the firm (namely the cost of bringing the firm back to the optimal 'starting position'). The (single period) variability of these variables thus feeds back into expected costs. If (single period) variability appears to be of importance to the bank empirically, this is perfectly consistent with such a view, therefore.

Of course, the owners of the firm as individuals do also face the possibility of emergency problems and adjustment costs as the result of (single period) variability of their income. They have to take into account these costs and balance them against the cost of devices helping to contain this risk (e.g., 'precautionary' savings), precisely as analyzed for the firm before. Taking into account these costs, they should then maximize expected utility for their entire planning horizon. This element can give an explanation for the existence of an aversion to (single period) variability of income.

3.3. *Real resource models*

The major part of the literature concerned with the theory of the banking firm treats the real resource or 'production' aspects of banking as of secondary importance only, and most often completely refrains from explicitly introducing them into the analysis. This is an undesirable state, since the banking business is actually quite resource (especially labour) intensive, and since there is no reason to believe that the real and the financial aspects of banking can be treated separately.

There exists one strand of the literature, however, which views the real resource aspects as a crucial element of any attempt to understanding the role and behaviour of banks and the financial system. This approach has been particularly emphasized by Pesek (1970) and Saving (1977), but also by a number of other authors, including Towey (1974) and Sealey and Lindley (1977).[22] These models essentially represent pure 'production cost' models of banking, i.e., they explain size and structure of bank liabilities and assets purely in terms of the flows of real resource costs of generating and maintaining those stocks (emphasizing in particular the cost of deposit production). Almost all of the other elements characteristic for the approaches discussed in the previous sections, including the cost of liquidity management and solvency protection (which, as emphasized before, also represent real resource expenses in part) are completely disregarded.

Such a production cost approach starts with the view that each type of deposit liability (and, in some cases, each type of asset) is characterized by the provision of a specified combination of services to the respective bank customers: 'The financial institution can maintain a stock of earning assets or deposits on its balance sheet only by constantly producing a flow of services to its customers and thus constantly incurring a flow of costs' [Sealey and Lindley (1977, p. 1255)]. These services (e.g., some combination of check clearing, withdrawals and deposits of funds, bookkeeping, safety, etc. in the case of deposits, or some combination of evaluating credit risks, bookkeeping, etc. in the case of bank assets) are produced by the bank with inputs of real resources under a given technology. That is, the bank has a production function relating different combinations of liabilities and assets to corresponding feasible combinations of inputs. E.g., if the production and maintenance of all the different liability and asset types are viewed as independent processes (see Sealey and Lindley), we have a separate production function for each of the i assets and j liabilities,

$$E_i = E_i(v_{i,h}) \quad \text{and} \quad D_j = D_j(v_{j,h}), \tag{23}$$

where $v_{i,h}$ denotes the quantity of input type h employed in connection with asset type i, etc. More generally, a production function allowing technical interaction between various activities can be specified (in implicit form),

$$H(E_i, D_j, v_h) = 0. \tag{23'}$$

[22]See also Kareken (1967), Stillson (1974), and Adar, Agmon and Orgler (1975).

The bank then is viewed as maximizing its profit

$$\pi=\sum_i r_i E_i-\sum_j i_j D_j-\sum_h w_h v_h \tag{24}$$

(where w_h is the decision period rental price of input h), subject to the technical constraints (23) or (23′), a balance sheet constraint $\sum_i E_i=\sum_j D_j(1-k_j)$ (where k_j is the – exogenous – cash reserve requirement for liability type j), and market determined prices r_i, i_j and w_h (or corresponding demand and supply functions of the public for E_i, D_j and v_h, if the bank is able to act as a price setter in any of these markets). This yields the optimal levels and combinations of all the E_i, D_j and v_h, according to the usual theory of the firm principles (with second order conditions requiring decreasing returns to each activity and scale, as in any other production process).[23] It is clear that in such an approach, decisions regarding asset and liability structure and scale generally will not be independent of each other, but made jointly.

Alternatively, and to simplify the following discussion, the minimum cost combinations v_h^* for producing each E_i – and D_j – combination can be found first, and profit expressed in terms of a cost function $C=C(E_i,D_j)\equiv \sum_h w_h v_h^*(E_i,D_j)$,

$$\pi=\sum_i r_i E_i-\sum_j i_j D_j-C(E_i,D_j). \tag{24′}$$

The real resource costs emphasized in this approach have already played an important part in some of the previous discussions in this paper, especially in connection with deposit supply in section 2.2. In this writer's opinion, it is highly desirable to assign more weight to models of this type.[24] However, they should be more fully integrated with some of the other elements traditionally stressed in banking models, especially with liquidity and solvency management aspects, than is the case in the studies mentioned above.

Such an approach, which combines several of the aspects discussed in the previous parts of this paper, was developed by Niehans (1978, ch. 9, pp. 175 ff.), and will be sketched below. The presentation followed here is based on a

[23] Instead of defining each deposit (and asset) type in terms of a specified bundle of associated services (and letting the customers choose between these), one can also treat changes in this bundle as changes in the quality of a given type of deposits. Towey (1974) deals with the problem in this way. Also, charges for specified services can be introduced into such a framework, of course.

[24] Another approach for which 'real resource' aspects are crucial is a 'brokerage' or 'distribution' approach which stresses differences in the costs of (credit) transactions between different pairs of transactors. Such an approach has been emphasized by Niehans and Hewson (1976) in connection with the Eurodollar business. This type of approach has been unduly neglected in the banking as well as the non-banking area.

similar model by Baltensperger and Milde (1977), which was influenced by an earlier version of the Niehans model. The analysis is based on the assumption of expected profit maximization and price taking by the firm in all markets. However, it can easily be modified to take account of price setting along the lines of the standard monopoly model (which seems to be preferred by many authors, and may be necessary to limit the scale of the firm in the presence of non-decreasing returns to scale in the technical sphere, just as in any other area), and should, in principle, be amenable to reformulation to take account of profit variability, if so desired (although at the expense of greater analytical complexity).

Joint determination of asset structure and firm scale. A simplified version of the model will be discussed first, which integrates the resource cost elements mentioned above with an explicit treatment of liquidity management costs. Questions of default and insolvency risk and capital management thus will first be disregarded for simplicity's sake. To further simplify the discussion, the model will be in terms of just one type of deposit liability (with a specified package of services) and one type of earning asset.[25] The bank's balance sheet constraint thus is $R+E=D$. Deposit changes during the decision period are uncertain, so that there are liquidity management costs as summarized in section 2.1. The present model can be considered an extension of the basic model discussed there to a situation where total portfolio size D and thus firm scale is not fixed, but a choice variable, and where the real resource cost of producing and maintaining deposit and credit accounts $C=C(D,E)$ are taken into account.[26] It thus allows to determine endogenously not only the optimal allocation of assets (between R and E), but – in contrast to the model of section 2.1 – also the optimal size of total assets and liabilities. The firm's expected profit can be written

$$E(\pi)=rE-iD-C-L, \tag{25}$$

where $C=C(D,E)$, $R+E=D$, and L is defined in eq. (1). However, in contrast to the discussion in section 2.1, we cannot treat the density function $f(X)$ as fixed in the present context, but must keep in mind the dependence of its parameters on the size of the bank's deposit liabilities.

It is convenient to model this explicitly by expressing L in terms of standardized units. Let $x\equiv X/\sigma_X$, and $b\equiv R/\sigma_X$, and assume – to simplify the presentation – that X is distributed normally with zero mean. Then we can

[25]Generalization to more assets and liabilities is formally quite straightforward, however.

[26]Maybe C should be made dependent on the number of accounts rather than the total volumes of D and E. For simplicity, a constant average size of deposit and credit accounts is assumed in the following so that this distinction becomes irrelevant.

write

$$L=\int_R^{\infty} p(X-R)f(X)\,\mathrm{d}X=\sigma_X\int_b^{\infty} p(x-b)\phi(x)\,\mathrm{d}x, \tag{1'}$$

where $\sigma_X=\sigma_X(D)$, with $\sigma'_X>0$ (and ϕ denotes the standard normal distribution). Thus we have $L=L(D,b)$, with $L_D>0$ and $L_b<0$.

Maximization of (25) with respect to R and D (and thus E) yields the following first order condition:

$$\frac{\delta E(\pi)^{\mathrm{d}D=0}}{\delta R}=-\frac{\delta E(\pi)^{\mathrm{d}D=0}}{\delta E}=-r-C_R^{\mathrm{d}D=0}-L_R^{\mathrm{d}D=0}=0, \tag{26}$$

$$\frac{\delta D(\pi)^{\mathrm{d}R=0}}{\delta D}=r-i-C_D^{\mathrm{d}R=0}-L_D^{\mathrm{d}R=0}=0. \tag{27}$$

Condition (26) says that the bank should change its asset structure in favour of earning assets until the resulting marginal revenue is just equal to the corresponding marginal production and liquidity cost. This corresponds to our condition (2) in section 2.1, except that we did not explicitly include the production cost term C there. This becomes clear if we note that

$$C_R^{\mathrm{d}D=0}=-C_E^{\mathrm{d}D=0} \quad \text{and} \quad L_R^{\mathrm{d}D=0}=-p\int_R^{\infty} f(X)\,\mathrm{d}X=-p\int_b^{\infty}\phi(x)\,\mathrm{d}x$$

to obtain

$$r-C_E=p\int_R^{\infty} f(X)\,\mathrm{d}X=p\int_b^{\infty}\phi(x)\,\mathrm{d}x. \tag{26'}$$

Condition (27) states that the bank should expand its production of deposits and thus its scale until the resulting marginal gain is just equal to the corresponding marginal cost. Note that

$$C_D^{\mathrm{d}R=0}=C_D^{\mathrm{d}E=0}+C_E^{\mathrm{d}D=0} \qquad (>0),$$

so that we can alternatively express this condition as

$$(r-C_E)=(i+C_D)+L_D^{\mathrm{d}R=0}. \tag{27'}$$

These two conditions together determine the optimal size and asset structure of the bank, as well as – through the cost function – its resource input (assuming that the appropriate second order conditions for a maximum are satisfied, of course). We can further evaluate

$$L_D^{dR=0} = (\delta L/\delta D)^{db=0} + (\delta L/\delta b)(\delta b/\delta D)^{dR=0}$$
$$= \sigma'_X \cdot L/\sigma_X + p \int_b^{\infty} \phi(x)\,dx \cdot b\sigma'_X \qquad (>0).$$

If we substitute (26′) into (27′), and use this expression for $L_D^{dR=0}$, we have

$$(r - C_E)(1 - b\sigma'_X) = (i + C_D) + \sigma'_X \cdot L/\sigma_X. \tag{27″}$$

The left- and right-hand side in this equation represent the marginal gain and marginal cost resulting from increases in the scale of the bank under the assumption that the optimal structure of assets is continuously maintained [i.e., eq. (26′) is continuously satisfied].[27] It might be noted, finally, that certain extensions of this model are quite straightforward. E.g., it can easily be reformulated for the case of a monopolistic bank. The relevant marginal revenue and expenditure expressions then simply have to incorporate the (positive or negative) elasticities of the respective supply and demand functions faced by the bank. The model can also be extended to include different types of deposits and assets, along the lines discussed previously. This leads to further interactions between liability and asset choice, because of resource cost interactions on the one hand, and because of the dependence of $f(X)$ and thus reserve management and liquidity costs on the structure of deposits on the other hand. The model by Klein (1971) can be obtained as a special case from such an approach, namely if resource costs are disregarded ($C=0$), and if $f(X)$ and thus the L-function is treated as independent of deposit structure and homogeneous of degree one in total size.[28]

Joint determination of capital structure and firm scale. The discussion above has disregarded default risk, insolvency costs, and equity capital considerations, and determined the optimal size and asset structure of the firm in a model incorporating real resource as well as liquidity costs. We can consider an analogous but opposite case by turning these assumptions around and jointly determining the optimal equity-deposit structure and firm size in a model which incorporates real resource and insolvency costs, but (for simplicity) disregards liquidity costs and reserve management considerations. The bank's balance sheet constraint then is $E = D + W$ (as in

[27]Inspection of (26′) shows that if C_E is independent of firm scale, the optimal asset structure is such that a constant value of b (i.e., of reserves relative to σ_X) is maintained.

[28]It is clear that under such circumstances a finite size of the firm is determined only if $r' < 0$ and/or $i' > 0$.

section 2.2), but with E variable and endogeneous (in contrast to section 2.2). The distribution of the return Y flowing from earning assets E is stochastic again (as in section 2.2), but now also depends on the size of E [and its structure, if endogeneous, with $E(Y)=rE$, and $\sigma_Y=\sigma_Y(E)$, $\sigma_Y'>0$].

The bank now maximizes

$$E(\pi)=rE-tD-C(D,E)-S-\rho(E-D), \tag{28}$$

where S, t and ρ are as defined in section 2.2, with respect to E (=total portfolio size, in this case) and D (and thus, implicitly, W). A detailed analysis of this model is analogous to the one for the first model discussed above.

Joint determination of asset-structure, liability-structure and firm scale. It should be clear that the elements of the two models discussed above can be combined to allow a simultaneous analysis of firm size, assets-structure and liability-structure, taking into account real resource costs, liquidity costs and insolvency costs all at the same time. The bank's balance sheet constraint then reads $R+E=D+W\equiv A$ (assuming just one type of asset and deposits), and the firm maximizes

$$E(\pi)=rE-tD-C(D,E)-L-S-\rho W. \tag{29}$$

The firm has three choice variables now. E.g., we can view it as choosing total portfolio size A and two variables characterizing the structure of asset and liability side, respectively, e.g., $\alpha=E/A$ (implying $1-\alpha=R/A$) and $\delta=D/A$ (implying $1-\delta=W/A$). Expected profit then can be expressed as

$$\begin{aligned}E(\pi)&=r\alpha A-t\delta A-C(A,\alpha,\delta)-L(A,\alpha,\delta)-S(A,\alpha,\delta)-\rho(1-\delta)A\\&=[\alpha r-\delta t-(1-\delta)\rho]A-C(A,\alpha,\delta)-L(A,\alpha,\delta)-S(A,\alpha,\delta).\end{aligned} \tag{30}$$

The expression $[\alpha r-\delta t-(1-\delta)\rho]$ measures the difference between the expected rate of return on assets r, weighted with the asset structure parameter α, and the weighted (with the liability structure parameter δ) sum of the two expected cost rates t and ρ. The cost expressions C, L and S are as previously defined, but expressed in terms of the portfolio allocation parameters α and δ, too (so that the partial derivatives of these functions with respect to A measure the respective marginal costs under conditions of constant portfolio structure). Optimization with respect to A, α and δ determines the optimal structure of the bank's asset and liability portfolio as well as its optimal scale in terms of the parameters of the underlying cost and return functions. It is clear that in a model of this type, all of these decisions will be made in an interdependent way.

References

Adar, Z., T. Agmon and Y.E. Orgler, 1975, Output mix and jointness in production in the banking firm, Journal of Money, Credit and Banking 7, May, 235–243.

Aftalion, F. and L.J. White, 1977, A study of a monetary system with a pegged discount rate under different market structures, Journal of Banking and Finance 1, 349–371.

Aigner, D.J. and C.M. Sprenkle, 1968, A simple model of information and lending behavior, Journal of Finance 23, March, 151–166.

Baltensperger, E., 1972a, Economies of scale, firm size and concentration in banking, Journal of Money, Credit, and Banking 4, Aug., 467–488.

Baltensperger, E., 1972b, Costs of banking activities - Interactions between risk and operating costs, Journal of Money, Credit, and Banking 4, Aug., 595–611.

Baltensperger, E., 1973, Optimal bank portfolios: The liability side, Jahrbücher für Nationalökonomie und Statistik 187, 147–160.

Baltensperger, E., 1974, The precautionary demand for reserves, American Economic Review 64, March, 205–210.

Baltensperger, E., 1976, The borrower-lender relationship, competitive equilibrium, and the theory of hedonic prices, American Economic Review 66, June.

Baltensperger, E. and H. Milde, 1976, Predictability of reserve demand, information costs, and portfolio behavior of commercial banks, Journal of Finance 31, June, 835–843.

Baltensperger, E. and H. Milde, 1977, Aktivstruktur, Passivstruktur und Bilanzvolumen einer Geschäftsbank, Zeitschrift für die Gesamte Staatswissenschaft 133, Dec., 681–701.

Barro, R.J., 1976, The loan market, collateral, and rates of interest, Journal of Money, Credit and Banking 8, Nov., 439–456.

Cootner, P.H., 1969, The liquidity of the savings and loan industry, in: Study of the savings and loan industry (The Federal Home Loan Bank Board, Washington, DC).

Edgeworth, F.Y., 1888, The mathematical theory of banking, Journal of Royal Statistical Society 51, March, 113–127.

Eppen, G.D. and E.F. Fama, 1969, Cash balance and simple dynamic portfolio problems with proportional costs, International Economic Review 10, June, 119–133.

Freedman, Ch., 1977, Microtheory of international financial intermediation, American Economic Review 67, 172–179.

Frost, P.A., 1971, Bank's demand for excess reserves, Journal of Political Economy 79, July/Aug., 802–825.

Goldfeld, S.M. and D.M. Jaffee, 1970, The determinants of deposit-rate setting by savings and loan associations, Journal of Finance 25, June, 615–632.

Gurley, H.G. and E.S. Shaw, 1960, Money in a theory of finance (Washington, DC).

Hart, O.D. and D.M. Jaffee, 1974, On the application of portfolio theory to depository financial intermediaries, Review of Economic Studies 41, Jan., 129–147.

Hester, D. and J.L. Pierce, 1975, Bank management and portfolio behavior (New Haven, CT).

Kane, E.J. and B.G. Malkiel, 1965, Bank portfolio allocation, deposit variability and the availability doctrine, Quarterly Journal of Economics 79, Feb., 113–134.

Kareken, J.H., 1967, Commercial banks and the supply of money: A market-determined demand deposit rate, Federal Reserve Bulletin 53, Oct., 1699–1712.

Klein, M.A., 1971, A theory of the banking firm, Journal of Money, Credit, and Banking 3, May, 205–218.

Knobel, A., 1977, The demand for reserves by commercial banks, Journal of Money, Credit, and Banking 9, Feb., 32–47.

Koskela, E., 1976, A study of bank behavior and credit rationing (Helsinki).

Milde, H., 1976, Kreditrisiko und Informationsaktivität im Bankbetrieb, Zeitschrift für Wirtschafts- und Sozialwissenschaften, no. 2, 127–142.

Miller, S.M., 1975, A theory of the banking firm: Comment, Journal of Monetary Economics 1, Jan.

Miller, M.H. and D. Orr, 1966, A model of the demand for money by firms, Quarterly Journal of Economics 80, Aug., 413–435.

Mingo, J. and B. Wolkowitz, 1977, The effects of regulation on bank balance sheet decisions, Journal of Finance 32, Dec., 1605–1616.

Monti, M., 1971, A theoretical model of bank behavior and its implications for monetary policy, L'Industria 2, 3–29.

Monti, M., 1972, Deposit, credit and interest rates determination under alternative bank objective functions, in: G.P. Szegö and K. Shell, eds., Mathematical methods in investment and finance (Amsterdam).

Morrison, G.R., 1966, Liquidity preferences of commercial banks (Chicago, IL).

Niehans, J., 1978, The theory of money (Baltimore, MD).

Niehans, J. and J. Hewson, 1976, The eurodollar market and monetary theory, Journal of Money, Credit, and Banking 8, Feb., 1–28.

Orr, D. and W.G. Mellon, 1961, Stochastic reserve loans and expansion of bank credit, American Economic Review 51, Sept., 614–623.

Parkin, M., 1970, Discount house portfolio and debt selection, Review of Economic Studies 37, Oct., 469–497.

Pesek, B., 1970, Bank's supply function and the equilibrium quantity of money, Canadian Journal of Economics, Aug., 357–385.

Poole, W., 1968, Commercial bank reserve management in an uncertain world: Implications for monetary policy, Journal of Finance 23, Dec., 769–791.

Porter, R.C., 1961, A model of bank portfolio selection, Yale Economic Essays 1, Fall, 323–359.

Pringle, J., 1973, A theory of the banking firm: Comment, Journal of Money, Credit, and Banking 5, Nov., 990–996.

Pringle, J., 1974, The capital decision in commercial banks, Journal of Finance 29, June, 779–795.

Pyle, D.H., 1971, On the theory of financial intermediation, Journal of Finance 26, June, 737–747.

Ritzmann, F., 1973, Die Schweizer Banken: Geschichte – Theorie – Statistik (Bern).

Santomero, A.M. and R.D. Watson, 1977, Determining an optimal standard for the banking industry, Journal of Finance 32, Sept., 1267–1282.

Saving, T., 1977, A theory of money supply with competitive banking, Journal of Monetary Economics 3, July, 289–303.

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, Sept., 1251–1266.

Slovin, M.B. and M.E. Sushka, 1975, Interest rates on savings deposits (Lexington, MA).

Stillson, R.T., 1974, An analysis of information and transaction services in financial institutions, Journal of Money, Credit, and Banking 6, Nov., 517–535.

Taggart, R.A. and St.I. Greenbaum, 1978, Bank capital and public regulation, Journal of Money, Credit, and Banking 10, May, 158–169.

Towey, R.E., 1974, Money creation and the theory of the banking firm, Journal of Finance 29, March, 57–72.