An Incentive Problem in the Dynamic Theory of Banking*

by

Ernst-Ludwig von Thadden**

– Revised, December 2000 –

* I am grateful to Abhijit Banerjee, Ed Green, Martin Hellwig, Ken Judd, Nobu Kiyotaki, Hyun Song Shin, Scott Stern, and Neil Wallace for helpful comments or dicussions. For part of the research reported here, I have benefited from the hospitality of the Federal Reserve Bank of Minneapolis and of the Department of Economics at Stanford University. Financial support by Schweizerischer Nationalfonds is gratefully acknowledged.

** DEEP, Universitè de Lausanne

Abstract

This paper develops a continuous-time model of liquidity provision by banks, in which customers can deposit and withdraw their funds strategically. The strategic withdrawal option introduces an incentive-compatibility problem that turns the problem of designing deposit contracts into a non-standard, non-convex optimal control problem. The paper develops a solution method for this problem and shows that, in this more general framework, the insights obtained from the traditional banking models change considerably, up to the point of liquidity provision becoming impossible. The continuous-time framework allows to discuss the problem elegantly and may help to make this part of the banking literature more operational in the sense of modern asset pricing theory.

Key words: Liquidity, deposit contracts, banking, incentive compatibility, continuous time, dynamic programming

JEL classification: D 51, D 92, G 20, G 21

I) Introduction

In an influential paper, Diamond and Dybvig (1983), building on earlier work by Bryant (1980), have argued that banks can provide intertemporal risk sharing possibilities to investors by taking on an illiquid portfolio. In their model, efficient intertemporal consumption allocations in a simple two-period general equilibrium model can be achieved through standard demand deposit contracts, and despite its illiquidity the banking system is essentially stable. This insurance argument provides the theoretical underpinning for what banking theory has usually called maturity transformation: the creation of short-term liabilities backed by long-term investments.

The original Bryant-Diamond-Dybvig model has been further developed and extended to deal with, among others, interbank coordination (Bhattacharya and Gale (1987)), intergenerational lending (Qi (1994)), banking regulation (Wallace (1996)), and economic growth (Bencivenga and Smith (1991)), and has become a standard work horse for modelling liquidity problems.

The present paper generalizes Diamond and Dybvig's model to the case of continuous time and studies the scope and structure of liquidity provision in a fully dynamic framework. This is of interest, because the Diamond-Dybvig model is, as far as incentive problems are concerned, essentially static. In that model, investors choose intermediated investment in an ex-ante sense, because it provides higher expected utility than direct investment. Once the uncertainty concerning future consumption is resolved, there is no more interaction in the model, and therefore, the investors trivially adhere to their choice.

However, if one introduces the possibility of repeated investment and ongoing uncertainty, such investor behavior cannot be taken for granted. Precisely because bank deposits provide greater liquidity than the underlying direct investment opportunities, a depositor has an incentive to withdraw her deposit, even without liquidity needs, thus realizing the liquidity premium the deposit provides, and to re-invest it directly. This arbitrage behaviour has a potentially destabilizing effect, because liquidity provision implies a transformation of the return structure of the intermediary's asset base that reflects the liquidity preferences of depositors. If the depositors have incentives to misrepresent their preferences, the intermediary itself risks illiquidity.

This is the basic incentive problem studied in this paper. I have presented a special, simple case of the present problem in an earlier paper (von Thadden, 1998), which, however, neither sheds light on the general structure of the incentive problem, nor provides a generalizable solution. The problem studied here does not arise in the original model by Diamond-Dybvig (1983), because reinvestment after one period there makes no sense. The problem is, however, a principal one: liquidity provision is by its very nature a dynamic (i.e. ongoing) phenomenon, and incentive issues are at the core of banking. Hence, a theory of liquidity provision that does not address dynamic incentive issues is at best incomplete.

A three-period or four-period model can, in principle, describe this incentive problem, by allowing for the possibility of re-investment at an interim period.[1)] Such a model would yield upper bounds on the extent of feasible liquidity provision. Yet, this type of model would suffer from the opposite drawback of the Diamond-Dybvig model: it would overstate the benefits from deviations to direct investment, as the new time horizon would, again, be too short to capture the full post-deviation history. In fact, liquidating a deposit contract for investment purposes has the downside that the deviant direct investment is less liquid. Hence, there is a tradeoff between sticking to intermediated investment and withdrawing the funds for direct investment, and this tradeoff is present at all dates at which the investor has not yet consumed her deposits. If the investor leaves her funds with the bank she gets high liquidity at lower levels of overall returns; if she withdraws and re-invests them she obtains extra returns which are less liquid.

It is, therefore, natural to study liquidity in a model with infinitely many periods. The continuous-time model developed here is a convenient and relatively elegant limiting case, with advantages similar to those of continuous-time formulations a la Merton in asset pricing. The model developed here, however, is much simpler than the advanced models of dynamic asset pricing. I consider no uncertainty on the asset side and only very specific processes on the demand side, for which, moreover, a Law of Large Numbers holds that eliminates all uncertainty in the aggregate.[2)] More work needs to be done to bring this theory to the level of generality of modern asset pricing theory. Yet, even this simple continuous-time model, a direct generalization of Diamond and Dybvig (1983), allows for a relatively rich analysis of the instabilities inherent in the bank-depositor relationship, and of the interplay between technological factors, consumer preferences, and uncertainty influencing them.

It is worth emphasizing that the incentive problem studied here is in a sense more fundamental than the type of problems analyzed by Jacklin (1987), Haubrich and King (1990), or Diamond (1997) that arise from the coexistence of banks and markets. In those models, there exist markets for intertemporal exchange next to the deposit facility offered by the banking sector, and agents have access to both instruments. As these papers show, the access to these markets can undo the risk sharing benefits of the banking arrangements, because, in equilibrium, the possibility of frictionless intertemporal exchange drives the rates of return of deposit contracts down to the technical rates of substitution in production.[3)]

Yet, as Wallace (1988) has argued, banks may be precisely an alternative to market-based intertemporal exchange in situations in which such markets either do not exist at all or are difficult to access for unsophisticated investors. In this perspective, the incentive problems caused by market investment opportunities are of little concern to banking theory. However, the incentive problem discussed in the present paper is still relevant: a deviation to autarky is possible even in the most segmented environment. In this sense, the present paper studies a more basic problem than the Jacklin (1987) - type literature.

Although simple in its outset, the analysis of the present paper becomes quickly complicated. In fact, taking the rather natural incentives for strategic deviation to autarky into account, turns the problem of designing deposit contracts from a standard, convex problem into a non-convex second-best control problem with a complex constraint set. Most of Section III of the present paper is devoted to solving this problem, for which the literature does not seem to provide a solution method. In particular, the standard Lagrangian methods used, e.g., in the literature on optimal consumption under borrowing constraints (e.g., Scheinkman and Weiss (1986), He and Pagès (1993)) do not apply here, because the control's permissible value at one single point in time is restricted by the path of all future values.

The method used here is to not use Lagrangian or recursive methods directly, but to first determine the region of binding incentive-compatibility in the optimization problem. Formally, if $H(t;r) \geq 0$, $t \in [0,1]$, describes the incentive-compatibility of deposit contract $r(\cdot)$ over the time horizon $[0,1]$, this amounts to characterizing the set $\{t \in [0,1]; H(t;r) = 0\}$. This method allows to break the problem into two types of sub-problems, one in which the incentive-compatibility constraint binds and the other in which any solution is "locally optimal" in the sense that it can be characterized by first-order conditions. This provides a full characterization of the solution, which can be shown to be unique by consecutively "stitching together" the solutions of the two types of sub-problems, very much as in standard control problems such as Guesnerie and Laffont (1984). The approach is therefore constructive, but relies neither on the maximum principle nor on convexity. As the solutions to the sub-problems are either given in explicit functional form or by a linear ordinary differential equation, we can easily compute the solution numerically for given specifications of the data of the economy.

Several features of the solution are of interest. First, there is little general structure. In particular, the set of dates at which the incentive compatibility constraints bind may consist of arbitrarily many time-intervals, alternating with intervals in which first-best liquidity provision can be achieved locally. While the (unconstrained) first-best solution is largely unaffected by the distribution of households' consumption needs, the second-best strongly depends on them.

Second, while there is little general structure, two alternative additional restrictions on the data of the problem (explored in Propositions 3 and 4) yield a remarkably simple structure. First, incentive compatibility is binding over the whole time-horizon if and only if the investors' intertemporal risk aversion is sufficiently large and the degree of irreversibility of investment not too large. In particular, the question of whether incentive compatibility is always binding is determined solely by a simple condition on preferences and technology.

Somewhat orthogonally, Proposition 4 presents a condition on the distribution of consumption uncertainty – regardless of preferences and technology – that implies a simple temporal structure of incentive compatibility. If the distribution of investors' consumption needs over time has a log-concave density (i.e., is essentially single-peaked), then there is only one regime switch: early on, incentive compatibility binds, later on, when investment has been in place for a sufficiently long time, incentive compatibility does not bind.

The main economic conclusion from this analysis is that the scope for liquidity transformation, as determined by the second-best, may be much very different from what the first-best suggests is desirable. From the first-best analysis (a la Diamond and Dybvig (1983)), liquidity transformation is, ceteris paribus, socially the more valuable the larger the investors' intertemporal risk aversion. However, as the second-best analysis shows, the scope for liquidity transformation decreases with intertemporal risk-aversion and is, independently of investors' preferences, restricted by the degree of irreversibility of investment. The lower the latter, the less liquidity provision is possible. In the extreme case in which the economy has only short-term investment possibilities (the productivity of long-term investment equals the productivity of short-term investment, which means that investment is fully reversible), liquidity transformation is completely impossible, and deposit arrangements can only replicate the autarky allocation.[4)]

These results cast doubts on the robustness of the dynamic features of the Diamond-Dybvig (1983) model, and therefore, in the spirit of finance theory after Merton (1990), on the practical usefulness of the basic version of that model. On the other hand, the results are compatible with the empirical observation that deposit contracts are typically not front-loaded, as the notion of liquidity by Diamond and Dybvig (1983) would suggest. The model may therefore provide a useful starting point for making Diamond and Dybvig (1983) operational.

The remainder of this paper is organized as follows. Section II sets out the model. Section III provides the main analysis of incentive-compatible deposit contracts. Section IV discusses the results, and Section V concludes. Longer proofs are in the appendix.

II) The Model

We consider a simple general-equilibrium economy with a continuum of identical households $a \in [0,1]$ who each live from time $t = 0$ to $t = 1$. Time is measured continuously, with $t \in [0,1]$.

There is one good in the economy. Each household is endowed with 1 unit of the good at time $t = 0$ that it can invest or store without depreciation, and with nothing thereafter. Everybody in the economy has access to the same constant-returns-to-scale investment opportunity, which, over any time-interval $[\tau, t]$, $0 \leq \tau < t \leq 1$, has an own rate of return of $R(t-\tau)$. In other words, an investment of $a \geq 0$ units of funds invested at time τ yields $aR(t-\tau)$ units of funds when liquidated at time t. Investment and its liquidation are costless.[5)] The return function R is assumed to have the following properties:

$$
\begin{aligned}
(i) \quad & R(0) = 1, \\
(ii) \quad & R \in \mathcal{C}^2([0,1], I\!R^+) \text{ with } R'(t) > 0 \text{ on } (0,1), \\
(iii) \quad & g := \frac{R'}{R} \text{ is non-decreasing on } (0,1).
\end{aligned}
$$

By (i), investment needs time to produce returns and there are no transactions costs, and by (ii) this return is positive and increases over time. (iii) implies that investing funds for a time-span of t yields at least as much as investing them for a time-span of $\tau < t$, liquidating them, and investing the proceeds for another period of $t - \tau$. Formally, this means $R(t) \geq R(t-\tau)R(\tau)$ for all $t > \tau$. Absent fixed costs for liquidation or investment, (iii) obviously holds if funds are used efficiently. If the inequality in (iii) is strict, "the real investment technology has an irreversibility, or goods-in-process, feature" (Wallace, 1988): a sequence of short-term investments is strictly inferior to a long-term investment.[6)]

Following Diamond and Dybvig (1983) and much of the literature, households' preferences exhibit the following strong intertemporal asymmetry. In "normal" times, households consume a constant, perfectly predictable flow of funds, not modelled here and normalized to zero. However, for each household a there is a time T_a at which its demand is singular and where it needs to consume all its wealth. From the perspective of the individual household, the time of consumption is an exogenous random event. However, the random variables $\{T_a\}_{a \in [0,1]}$, $0 < T_a \leq 1$, are assumed to be identically and independently distributed and to satisfy the Law of Large Numbers (see Judd, 1985). Hence, there is no uncertainty in the aggregate. To simplify some parts of the exposition, let the c.d.f. of the T_a, $F : (0,1] \to [0,1]$, be differentiable (everywhere) with $F' = f$. Without loss of generality we can assume that $F(1) > F(t)$ for all $t < 1$.

To emphasize the importance of the unforeseen consumption shock, a household's utility over its lifetime is assumed to depend solely on what it can consume at time T_a. Households' expected utility as of time $t = 0$ then is

$$U = \int_0^1 u\big(w_a(T_a)\big) dF(T_a),$$

where $w_a(t)$ is household a's wealth at time t, and its instantaneous utility $u : I\!R_+ \to I\!R$ satisfies $u' > 0$, $u'' < 0$, $u'(0) = \infty$, $u'(\infty) = 0$ (see Jacklin (1987) and Haubrich and King (1990) for more symmetric and general preferences).

This completes the description of the model, a straightforward generalization of the model of Diamond and Dybvig (1983) to continuous time. The four-period production model of Postlewaite and Vives (1987) is also easily translated into the above framework if $R(0) < 1$ is admitted.

Absent any interaction in the economy, each household can invest its funds in $t = 0$ in order to liquidate the investment in T_a. This yields the autarkic utility level $\bar{U} := \int_0^1 u\big(R(t)\big) dF(t)$.

However, if the consumption shocks T_a are commonly observable and contractible, households can usually do better by acting collectively. Under autarky, an individual household is forced to consume a random amount $\tilde{c}$ with c.d.f. $F \circ R^{-1}$. Since there is no uncertainty in the aggregate and the fluctuations in consumption given by $F \circ R^{-1}$ will typically be neither in line with the households' risk preferences nor with their preferences for intertemporal substitution, there are gains from reallocating funds over time between households. Such a reallocation can be achieved by investing endowments collectively at $t = 0$ and liquidating them according to a collectively agreed upon rule, subject to the constraint that aggregate consumption be feasible.

Therefore, suppose that the households pool their funds at $t = 0$ and invest collectively. In the collective, it is certainly a dominated strategy to liquidate assets in order to reinvest them (strictly dominated if g in (iii) above is strictly increasing). Hence, all funds are invested in $t = 0$ and liquidated only for consumption purposes. Let $S(t)$, $0 \le t \le 1$, denote the aggregate amount of funds available at time t and $r(t)$ the aggregate rate of funds consumed at time t. The evolution of S is determined by r. In the absence of liquidation, S would instantaneously evolve as R, i.e. with a growth rate of g. Hence, for a given r, the evolution of S is given by

$$S'(t; r) = g(t)S(t; r) - r(t)f(t) \quad \text{a.e.} \tag{1}$$

Straightforward integration of (1), together with the initial condition $S(0) = 1$, yields

$$S(t; r) = R(t)\Big(1 - \int_0^t \frac{r(\tau)}{R(\tau)} dF(\tau)\Big). \tag{2}$$

The problem of finding the first-best optimal aggregate consumption rate $r^\star(\cdot)$ can now be stated as

$$\max \quad \int_0^1 u(r(t))dF(t) \tag{3}$$

$$\text{subject to} \qquad r \geq 0 \;\text{ integrable on } [0,1], \tag{4}$$

$$S(t;r) \geq 0, \;\; t \in [0,1]. \tag{5}$$

Let $q := (u')^{-1}$ denote the inverse of the marginal utility function. The following proposition provides a simple characterization of the first-best.

Proposition 1: The (dF-a.e.) unique solution to problem (3) – (5) is

$$r^\star(t) = q\Big(\frac{C^\star}{R(t)}\Big), \tag{6}$$

where the constant $C^\star$ is given by

$$\int_0^1 \frac{r^\star(t)}{R(t)} dF(t) = 1. \tag{7}$$

Proof: By straightforward verification. Uniqueness follows from the strict concavity of u.[7)]

∎

Proposition 1 has a simple economic interpretation. (6) states that at the optimum, marginal utility of consumption at time t, $u'(r^\star(t))$, weighed with the marginal rate of transformation from time 0 to time t, $R(t)$, must be constant over time. In particular, the gross rate of return on collective investment, $r^\star$, is a smooth and strictly increasing function of time, although the distribution of shocks, given by f, may be very irregular. The distribution of T_a influences $r^\star$ only through the constant $C^\star$ determined by the resource constraint (7).

Notice that the economy could choose a constant consumption flow $r(t) \equiv \bar{r}$. However, although this would completely eliminate consumption risk, this is not optimal. This simple observation suggests that it is not only the households' risk aversion which makes collective investment attractive, but also the fact that the timing of investment returns does not correspond to the households' intertemporal consumption preferences. To see the impact of risk aversion more clearly, differentiate (6) to get

$$\frac{r^{\star\prime}(t)}{r^\star(t)} = -g(t)\frac{C^\star}{R(t)}\frac{1}{r^\star(t)u''(r^\star(t))} = \frac{g(t)}{\gamma(r^\star(t))}, \tag{9}$$

where

$$\gamma(c) := -\frac{cu''(c)}{u'(c)}$$

is the absolute value of the elasticity of households' marginal utility with respect to consumption. In decision theory under uncertainty, γ is the households' (static) coefficient of relative risk aversion. In deterministic problems of intertemporal consumption choice with time separable utility, γ denotes the relative importance of the income effect as compared to the substitution effect, and also the inverse of the intertemporal elasticity of substitution. Hence, with the simple preferences employed here, several aspects, in addition to risk sharing, determine the shape of $r^\star$.[8)]

An important insight of Bryant (1980) and Diamond and Dybvig (1983) has been to relate the demand for liquidity to γ. To see this most easily, we shall restrict attention to the following class of utility functions:[9)]

$$u(c) = -\frac{1}{\gamma - 1} c^{1-\gamma}, 1 < \gamma \in I\!R. \tag{10}$$

By (9), the first-best consumption profile now is always flatter than the return path available from the economy-wide investment opportunity. The resource constraint (7) therefore implies that $r^\star(0) > R(0)$ and $r^\star(1) < R(1)$. In other words, in an economy with intertemporal relative risk aversion of more than unity, consumption along the first-best path is shifted from later to earlier times, in the sense that a household hit by an early consumption shock consumes more than its funds have physically produced up to that time, at the expense of households who will consume later. Conversely, in an economy with $\gamma(c) < 1$ for all $c > 0$ consumption would optimally be shifted towards later time points.

III) Incentive-Compatible Deposit Contracts

If the consumption shocks T_a are not observable, markets for contingent claims cannot exist, because individuals with no consumption shock will never give away their investment for what an individual hit by a shock is able to offer. The central contribution by Bryant (1980) and Diamond and Dybvig (1983) has been to show that demand deposit contracts can be interpreted as mechanisms that provide intertemporal substitution possibilities for the economy which markets cannot provide.

Following them, suppose that the households in the economy set up an organization – henceforth called "bank" – that collects the households' funds in $t = 0$ as deposits and invests them collectively. If a household wants to consume a fraction δ of its wealth at time t, it can withdraw it from the bank and obtains $\delta r(t)$, where $r(t)$ is the gross interest rate at time t, pre-specified in the deposit contract in $t = 0$.

Here, a deposit contract between a bank and its customers should be properly viewed as a mechanism defining a "withdrawal game" among depositors. In this game, the return to each customer from depositing her funds in the bank depends on the deposit contract and the withdrawal decisions of all other customers. The question posed by Diamond and Dybvig (1983) is whether the first-best path $r^\star$ can be implemented as a Nash equilibrium of the withdrawal game.

In their model, the answer is affirmative. The only reason for an individual household to withdraw funds at a date before T_a would be the belief that other households will do so as well. To prevent an equilibrium in which such beliefs are correct (a "bank-run"), the mechanism only has to include a provision such as "suspension of convertibility". If the bank stops paying out funds to depositors ("suspends convertibility") once withdrawal becomes excessive (compared to F), there is no danger that the bank will run out of funds prematurely, and everybody's demand for funds can be satisfied at T_a. Of course, this argument only holds if there is no aggregate risk in the economy. If everybody knows that the bank behaves this way, it is individually strictly optimal to withdraw only at T_a, and the first-best allocation of funds results.

However, in a model with more than two periods, the first-best can generally not be implemented through deposit contracts. The reason is that the option of withdrawing and reinvesting funds individually dominates the banking option.

To see this formally, assume that $F(t) > 0$ for all $t > 0$, i.e. that liquidity needs arise right after $t = 0$ already. In any equilibrium that implements the first-best the bank has to pay out $r^\star(t)$ at $t > 0$ according to the distribution F. Since $F > 0$ and the T_a are non-contractible, there is a neighborhood of $t = 0$ such that the bank must repay every depositor with probability 1 whenever she demands her funds in this neighborhood.

Consider a household that has not been hit by a consumption shock up to time t. Funds left in the bank yield future returns of $r^\star(\tau)$, $\tau > t$, per unit. Funds withdrawn can be reinvested privately where they will yield $r^\star(t)R(\tau - t)$, $\tau > t$, per unit. By (6), the latter return stream pointwise dominates the former iff

$$q\Big(\frac{C^\star}{R(t)}\Big)R(\tau - t) \;\; > \;\; q\Big(\frac{C^\star}{R(\tau)}\Big), \quad \tau > t,$$

which is equivalent to

$$h_t(\tau) := R(\tau)u'\Big(q\big(\frac{C^\star}{R(t)}\big)R(\tau - t)\Big) \ < \ C^\star, \quad \tau > t.$$

For $t = 0$ we have $h_0(0) = C^\star$ and

$$h_0'(\tau) = R'(\tau)u'\big(R(\tau)q(C^\star)\big)\Big(1 - \gamma\big(R(\tau)q(C^\star)\big)\Big) \ < \ 0, \quad \tau > 0$$

by assumption (10). Hence, for $t = 0$,

$$\int_t^1 u\big(r^\star(t)R(\tau - t)\big)dF(\tau) \ > \ \int_t^1 u\big(r^\star(\tau)\big)dF(\tau), \tag{11}$$

and by continuity, there is a $\bar{t} > 0$ such that (11) holds for all $0 < t < \bar{t}$. For all these t, each household prefers to withdraw its funds over leaving leaving them in bank.

This argument shows a basic tradeoff for the individual household. Since its elasticity of intertemporal substitution is relatively small ($\gamma > 1$), the first-best flattens the aggregate return path by shifting returns forward in time. However, these higher early returns, when strategically reinvested, generate a new individual return path which had not been feasible before. Now the household must decide whether it prefers the relatively flat first-best path or a steeper (i.e. riskier) individual path on a higher level. The premium of the first-best over autarky is highest at $t = 0$ and, as the above argument shows, high enough to even push the new individual path uniformly above the first-best path. Hence, if withdrawal around $t = 0$ is possible, everybody will demand it. In other words, any attempt to implement the first-best induces a bank-run in strictly dominant strategies in $t = 0$.[10)]

The reason for the apparent discrepancy between this observation and the result of Diamond and Dybvig (1983) is the following. The Diamond-Dybvig model is obtained as a special case of the present framework by concentrating the mass of $\mathrm{d}F$ around $t = 1/2$ and $t = 1$ and by having $R(1/2) \approx 1$. If $f \equiv 0$ on $(0, 1/2)$ this has two consequences: the bank can refuse to pay out at early t, and reinvestment, which therefore can only happen at later t, is unattractive for depositors. Thus, in this case depositors will leave their funds in the bank. However, if consumption needs arise over a longer time-span and the bank is forced to satisfy also early withdrawal demands (because $F(t) > 0$ for all $t > 0$), withdrawal and reinvestment is always more profitable for depositors.

The negative observation above, that the first-best is not implementable, holds regardless of risk preferences or technology. We now turn, more constructively, to the structure of incentive-compatible liquidity provision. Now all three factors – preferences, technology, and the distribution of consumption shocks – will play a role.

At each time-point, a household not yet hit by a consumption shock must decide whether to leave its funds with the bank or whether to withdraw and reinvest them privately. Defining, for a given schedule r, the function

$$H(t) := H(t;r) := \int_t^1 \big[u(r(\tau)) - u(r(t)R(\tau - t))\big]dF(\tau),$$

$t \in [0,1]$, the expected net gain from leaving the funds with the bank in $t \in [0,1]$ is $\frac{H(t)}{1-F(t)}$. Incentive-compatible liquidity provision therefore poses the following problem:

$$\max \quad \int_0^1 u(r(t))dF(t) \tag{3}$$

$$\text{subject to} \quad r \geq 0 \ \text{ integrable on } [0,1], \tag{4}$$

$$S(t;r) \geq 0 \ \ \forall t \in [0,1], \tag{5}$$

$$H(t;r) \geq 0 \ \ \forall t \in (0,1]. \tag{12}$$

Problem (3)-(5), (12) differs from classical optimal control problems in several respects. In particular, the constraints (12) involve averages over the future path of r together with point values. This feature implies that the payout at one single point in time is restricted by the path of all future payouts. On the other hand, by the resource constraint (5), the path of past payouts obviously restricts future payouts. Finally, and technically most importantly, the incentive constraints (12) render the problem non-convex.

It is therefore impossible to apply standard optimal control techniques, such as in Scheinkmann and Weiss (1986) or Guesnerie and Laffont (1984), to the problem. However, the problem still has a separable structure that makes it amenable to local techniques. Intuitively, the designer of an incentive-compatible deposit contract faces the following problem. At any given time $\tau \in (0,1)$, the incentive-compatibility constraint (12) tends to push the value $r(\tau)$ down in order to make withdrawal and reinvestment unattractive. On the other hand, from the perspective of earlier $t < \tau$, future payouts $r(\tau)$ should be high in order to make sticking to the collective investment scheme as attractive as possible. The optimal contract has to balance these demands all along the path r.

The problem clearly has a recursive structure: given the maturity of the assets in place, at any point in time, t, the objective function and constraints are forward looking and depend on the past only through the state variable $S(t)$. But the problem has even more structure. To see this, suppose that the incentive compatibility constraint (12) is slack at one point t_0. It is simple to see that, because of the strict concavity of u, any solution to (3)-(5), (12) cannot have jumps (i.e. must agree a.e. with a continuous function). Therefore, for a given solution r, $H(\cdot;r)$ is continuous, too. Because $H(t_0;r) > 0$ and $H(1;r) = 0$, there exists a $t_1 > t_0$ such that H is strictly positive on (t_0, t_1) and $H(t_1;r) = 0$. Intuitively, if the incentive compatibility at one point does not bind, it cannot bind locally around this point, but it must bind later, because at the very last moment at least, everybody will want to cash in.

This simple observation implies that on (t_0, t_1) an optimal incentive-compatible contract solves an optimization problem which is only constrained by the amounts of assets in place at time t_0 and t_1, not by incentive considerations. To make this precise, suppose that r is a solution to our problem and that $S(t_0;r) = S_0$ and $S(t_1;r) = S_1$. Then, on $[t_0, t_1]$ x must solve the unconstrained problem as in $(3) - (5)$, modified such that the asset stock as given by (2) is depleted from S_0 to S_1 instead of from 1 to 0. Hence, on $[t_0, t_1]$, r solves

$$\begin{aligned} & \max \quad \int_{t_0}^{t_1} u(r(\tau))\mathrm{d}F(\tau) \\ \text{subject to} \quad & r \geq 0 \text{ integrable on } [t_0, t_1], \\ & \int_{t_0}^{t_1} \frac{r(\tau)}{R(\tau)}\mathrm{d}F(\tau) = \frac{S_0}{R(t_0)} - \frac{S_1}{R(t_1)}. \end{aligned}$$

Using Proposition 1, we know therefore that r must have the form

$$\begin{aligned} r(t) =& q\Big(\frac{C}{R(t)}\Big) \\ =& C^{-\frac{1}{\gamma}} R(t)^{\frac{1}{\gamma}} \end{aligned}$$

$\mathrm{d}F$-a.e. on $[t_0, t_1]$ for some constant $C > 0$. Hence, the problem has a "local optimality property": solutions to the local unconstrained problems differ from the first-best only by a factor (which depends on the local data).

This result allows us to write out H explicitly on (t_0, t_1) (where it is strictly positive):

$$H(t;r) = \frac{1}{\gamma - 1} C^{\frac{1-\gamma}{\gamma}} G(t), \tag{13}$$

where

$$G(t) = \int_t^1 \big[R(t)^{\frac{1-\gamma}{\gamma}} R(\tau - t)^{1-\gamma} - R(\tau)^{\frac{1-\gamma}{\gamma}}\big]\mathrm{d}F(\tau).$$

Note that G does not depend on the constant C and therefore not on t_0 and t_1 (this is where the specific form of the utility function, (10), is crucial). Furthermore, since G is differentiable, (13) implies that on all intervals on which H is positive, it is differentiable, and

$$H'(t;r) = \frac{1}{\gamma - 1} C^{\frac{1-\gamma}{\gamma}} G'(t), \tag{14}$$

This property and the local optimality property of the overall problem make G a useful tool in characterizing the solution.

In fact, as the next proposition will show, the set of time-points at which the incentive constraints (12) bind is completely determined by G. This set does not simply consist of those t with $G(t) \leq 0$, as a casual inspection of (13) might suggest.[11)] To see how this set can be constructed from G, note first that $G(1) = 0$. Hence, if there is an interval $(T, 1)$ on which $G > 0$, incentive compatibility will not bind on this interval (locally optimal intertemporal risk sharing). However, by the recursive nature of the problem, incentive compatibility will bind to the left of T as briefly as possible. The largest possible value of $t < T$ where (going backwards) incentive compatibility can cease to bind again is the next t with $G'(t) = 0$. To the left of this value a period of non-binding incentive compatibility is possible by (14) and will be optimal by the recursive nature of the problem. And so on, until one arrives at $t = 0$.

This informal argument also shows that if $G'(t) \geq 0$ for all t, then the incentive compatibility constraints must bind everywhere on $[0, 1]$.

To make the above reasoning precise, define the following subset of the set of local minima of G:

$$\begin{aligned} \mathcal{B} := \big\{ b \in [0,1]; G(t) \geq G(b) \ \ \forall t > b \ \text{ and} \\ \text{if } \ b > 0 : \ \ \exists \epsilon > 0 : G(t) > G(b) \ \ \forall t \in (b - \epsilon, b) \big\}. \end{aligned} \tag{15}$$

Loosely speaking, the elements of $\mathcal{B}$ are "right-looking weak absolute minima" and "left-looking strong local minima" of G. The upper part of Figure 1 provides an illustration (where $\mathcal{B}$ has six elements). If there is an interval $(T, 1)$ on which $G > 0$ (the case considered informally above), then $\max \mathcal{B} = 1$. If $G'(t) \geq 0$ for all $t \in [0, 1]$, then $\mathcal{B} = \{0\}$. Clearly, $\mathcal{B}$ is not empty. Also, if card $\mathcal{B} = \infty$, $\mathcal{B}$ attains its infimum.

– Figure 1 about here –

For each $b \in \mathcal{B}$ one can define a unique $a(b) \leq b$ as follows:

$$a(b) := \begin{cases} 0 & \text{if } b = \min \mathcal{B}, \\ \max\{t < b; G(t) = G(b)\} & \text{else.} \end{cases}$$

$(a(b), b)$ is the largest interval (α, b) on which $G(t) > G(b)$. Note that $a(b) = b$ if and only if $b = 0$. Figure 1 provides an illustration of this construction.

These definitions allow a full characterization of the incentive compatibility problem.

Proposition 2: The second-best problem (3)-(5), (12) has a (dF-a.e.) unique solution $\bar{r}$. $\bar{r}$ is piecewise continuously differentiable, and $H(t; \bar{r}) > 0$ if and only if $t \in \bigcup_{b \in \mathcal{B}} (a(b), b)$.

Proof: In the appendix.

The proof follows the logic outlined above, but cannot be phrased directly in recursive terms, because one cannot rule out the possibility that the set $\mathcal{B}$ is infinite (and that, therefore, the recursion gets stuck). The idea of the procedure can be described informally by means of Figure 1, which superimposes the graphs of G and H: going backward from any $b \in \mathcal{B}$, $\bar{r}$ is locally first-best as long as possible, namely until $a(b)$, then incentive compatibility binds as briefly as possible, namely until the next $b \in \mathcal{B}$, etc.

This solution procedure is reminiscent of the method used by Guesnerie and Laffont (1984) in the context of an adverse-selection contracting problem with a continuum of types. In their model, the second-best, $\bar{l}$, is obtained from the first-best, $l^\star$, by replacing different segments of $l^\star$ by constants, in a way that respects the continuity of $\bar{l}$. In particular, their solution is also piecewise differentiable, just as in our case. In fact, the "cut-and-paste" procedure of Figure 1 resembles Figure 5 in Guesnerie and Laffont (1984). The main difference between the two problems is that theirs is a standard optimal-control problem, which can be characterized by the maximum principle, a tool that is not available in the present case.

Despite the non-convexity of the problem, Proposition 2 shows that its solution is unique. The solution is characterized by alternating phases of binding and non-binding incentive compatibility (i.e. local risk sharing possibilities), whose duration is determined by the interplay between technical productivity (R), intertemporal risk aversion (γ), and the distribution of consumption shocks (F). In general, this relationship will be quite complex. Figure 2 provides a simple example in which the distribution of consumption shocks exhibits two peak phases (around $t = .3$ and $t = 1$) and where the incentive compatibility constraint binds before and at the end of the first peak, and nowhere else.[12)]

– Figure 2 about here –

In view of such "scattered" local risk sharing possibilities one may wonder whether there are conditions for the solution to have a simple structure. Two such conditions are particularly interesting, depending on what parameters one wants to focus on. The first highlights the interplay of technology and intertemporal risk aversion.

Proposition 3: If $\gamma g(0) \geq g(1)$, then $H(t;\bar{r}) \equiv 0$ for all $t \in [0,1]$. If $\gamma g(0) < g(1)$, then there is a $T < 1$ such that $H(t;\bar{r}) > 0$ on $(T,1)$.

Proof: We have

$$G'(t) = \frac{\gamma - 1}{\gamma} R(t)^{\frac{1-\gamma}{\gamma}} \int_t^1 R(\tau - t)^{1-\gamma} \big[\gamma g(\tau - t) - g(t)\big] dF(\tau).$$

If $\gamma g(0) \geq g(1)$

$$\gamma g(\tau - t) - g(t) \geq \gamma g(0) - g(1) \geq 0,$$

hence, $G'(t) \geq 0 \ \forall t \in [0,1]$, which implies $\mathcal{B} = \{0\}$. The result now follows from Proposition 2.

If $\gamma g(0) < g(1)$ the continuity of g implies that there exists a $t_0 < 1$ such that $\gamma g(1-t) < g(t) \ \ \forall t \in (t_0, 1]$. Since F is strictly increasing at $t = 1$,

$$G'(t) \leq \frac{\gamma - 1}{\gamma} R(t)^{\frac{1-\gamma}{\gamma}} \big[\gamma g(1-t) - g(t)\big] \int_t^1 R(\tau - t)^{1-\gamma} dF(\tau) < 0$$

for $t \in (t_0, 1)$. Hence, max $\mathcal{B} = 1$, and the result follows again from Proposition 2.

■

If $\gamma g(0) \geq g(1)$ there is no scope to relax the incentive compatibility constraint anywhere, regardless of the distribution of consumption shocks.[13)] The value of leaving the assets in place, and hence the value of collective investment, is so small as compared to γ that any attempt to equate the weighted marginal utilities, $u'(r(t))R(t)$, over some time period would induce withdrawal and private reinvestment. Therefore, the optimal payout path is exclusively determined by incentive-compatibility considerations.

Only if $g(1) > \gamma g(0)$ incentive compatibility ceases to be binding everywhere. Quite naturally, it does so at the end of the time horizon, when the collective asset has matured to yield the highest rates of return. However, as seen above, local risk sharing possibilities can arise also in other periods of the life of the collective asset. The following proposition, which focuses solely on the distribution of consumption shocks, provides a condition under which such scattered risk sharing possibilities cannot occur.

Proposition 4: If f is differentiable and log-concave then there is a $T \leq 1$ such that $\{t; H(t;\bar{r}) > 0\} = (T,1)$.

Proof: In the appendix.

Many density functions are log-concave, on $[0, 1]$, for example, the uniform, the truncated normal, exponential and the parabolic densities nt^{n-1} and $n(1-t)^{n-1}$, $n \geq 1$. Log-concavity is weaker than concavity, but slightly stronger than quasi-concavity (see, e.g., Caplin and Nalebuff (1991)).[14] In the present context, log-concavity does not render the problem convex, but imposing this specific form of single-peakedness gives enough regularity to the distribution of consumption shocks to obtain an optimal payout path consisting of at most two regimes. During the first phase, on $[0, T]$, payout is determined solely by incentive compatibility (plus a boundary condition, of course); during the second phase – which may be non-existent according to Proposition 3 – incentive compatibility does not bind, and the optimal degree of intertemporal risk sharing can be achieved.

Although the construction of the scheme $\bar{r}$ has eliminated the incentives to withdraw and reinvest individually, in general, a deposit contract offering the interest path $\bar{r}$ still provides some liquidity to households. Therefore, it is still vulnerable to expectation based bank-runs as identified by Diamond and Dybvig (1983). Hence, also in this case, a second-best optimal mechanism must include a suspension-of-convertibility clause in order to eliminate inefficient equilibria. Such an arrangement – a deposit contract offering $\bar{r}$ on demand, together with suspension of convertibility if withdrawal demands exceed the rate of $f(t)$ – achieves second-best liquidity provision as the unique Nash equilibrium.[15]

IV) Interpretation

In a first-best world, deposit contracts simply flatten the aggregate consumption path over time in response to intertemporal risk aversion of households. The introduction of incentive compatibility changes the analysis and assessment of deposit contracts in several ways.

Before turning to the differences, it is useful to point out what the present second-best results have in common with the traditional picture. First, one easily verifies that for back-loaded aggregate consumption distributions, such as $f(t) = 2t$, flattening individual consumption paths uses up less aggregate resources, hence leads to higher payout paths, than for front-loaded aggregate consumption distributions, such as $f(t) = 2(1-t)$. If we interpret back-loadedness of aggregate consumption as higher aggregate patience, then this implies that higher aggregate patience allows higher aggregate consumption. This is as in Diamond and Dybvig (1983).

A similarly intuitive picture emerges for the role of γ, which can be interpreted as a measure of individual impatience: the higher the individual's intertemporal risk aversion, the higher its demand for individual consumption at early dates. It is easy to verify that higher γs imply flatter consumption paths, just as standard arguments suggest.

Yet, a closer inspection of the structure identified in Propositions 2, 3, and 4 reveals some important differences between first-best and second-best arrangements. First, and quite surprisingly, as noted in Section II, the distribution of payouts over time in the first-best is independent of irregularities of the distribution of consumption shocks – only the curve's overall position depends on F, via the resource constraint (7).

In the second-best, however, the distribution of payouts responds strongly to the shape of F. In particular, if f has several peaks, local intertemporal risk sharing possibilities around these peaks will typically be restricted by incentive compatibility (see Figure 2 for an example). The extent to which this happens depends on the shape of R and on γ. However, whenever this happens, the transformation function of deposit contracts is restricted in periods where this is socially least desirable, namely in periods of high withdrawal demand.

A second, more fundamental, set of observations relates the effectiveness of deposit contracts to the degree of irreversibility in the production technology. Assume for simplicity that f is log-concave. Differentiating the identity $H(t;\bar{r}) \equiv 0$ on $[0,T]$ yields (equation $(A1)$ in the appendix):

$$\bar{r}'(t) = \bar{r}(t) \int_t^1 R(\tau - t)^{-\gamma+1} g(\tau - t) dF(\tau) \;\Big/\; \int_t^1 R(\tau - t)^{-\gamma+1} dF(\tau), \tag{16}$$

which implies

$$g(0) \leq \frac{\bar{r}'(t)}{\bar{r}(t)} \leq g(1-t). \tag{17}$$

Inequality (17) puts a bound on how much maturity transformation can be achieved, regardless of its social value. To illustrate the point consider the extreme case of a constant growth rate, $g(t) = a \quad \forall t \in [0,1]$. The (first-best) demand for maturity transformation as expressed by (9), $r^{\star\prime}(t)/r^\star(t) = a/\gamma$, requires the optimal consumption path to be the flatter the greater the households' intertemporal risk aversion, with constant consumption in the limit for $\gamma \to \infty$. However, by Proposition 3 and (17), $\bar{r}'(t)/\bar{r}(t) = a$ on $[0,1]$. Hence, $\bar{r} = R$, and only autarky is incentive compatible, regardless of γ.

To put this result into perspective, for $g(0) > 0$ we can interpret $g(1)/g(0)$ as an index of the irreversibility of the investment opportunity. Since g is increasing, this index is not smaller than 1. If $g(1)/g(0) = 1$, we have $g(t) \equiv a$, hence $R(\tau) = R(\tau - t)R(t)$ for $\tau > t$, and there is no irreversibility: a sequence of short-term investments is as good as one long-term investment.[16)] The larger $g(1)/g(0)$, the larger the overall irreversibility of the investment opportunity, and the larger the relative disadvantage of short-term investments.

If $g(1)/g(0) > 1$, Propositions 2 and 3 imply that for smaller values of γ, optimal liquidity provision is possible on an end-interval $[T, 1]$. However, as noted above, this is precisely the range of γ for which only little liquidity provision is first-best optimal, anyway. If γ grows larger, i.e. intertemporal risk aversion grows, and liquidity provision becomes more important, two effects are operating. First, the range over which optimal liquidity provision is possible shrinks, until it vanishes completely when $g(1)/g(0) \leq \gamma$ (Proposition 3). And second, over the complementary range in which incentive compatibility binds, liquidity provision is restricted by (17), and therefore, if $g(0) > 0$, bounded away from the first-best. In this double sense one can say that incentive compatibility restricts liquidity transformation by deposit contracts just when it would be socially most useful.

As a final point, equation (16) permits an interesting observation on the term structure of the interest rate path $\{\bar{r}(t)\}$. (16) implies that on $[0, T]$

$$\frac{\bar{r}'(t)}{\bar{r}(t)} > \frac{g(t)}{\gamma} \iff G'(t) > 0.$$

Note that by (9), $\bar{r}'(t)/\bar{r}(t) = g(t)/\gamma$ on $[T, 1]$, and that $G'(T) > 0$. Hence, the interest rate schedule compatible with the given investment opportunity is relatively steep in the short run and relatively flat in the long run, with a kink occurring at $t = T$.

V Conclusion

The model of dynamic liquidity provision developed in this paper has exposed a relatively general continuous-time incentive-contracting problem. This problem is to design a payment path that provides no incentive to deviate at any date, when the payoff from deviating is determined by the instantaneous payment, and the reward for not deviating by future payments. Hence, for incentive reasons, at any date the present payment should be small and future ones large. Yet, as soon as the present is past and the future is present, large instantaneous payments themselves turn into an incentive problem, which can only be solved by even larger payments later on. However, a policy of "buying time" by simply always increasing payments is not only ruled out by resource constraints, but also by optimizing considerations, which require that the payout path responds optimally to intertemporal consumption needs. Hence, all along the payout path these feedbacks between present and future payments must be balanced carefully.[17)]

The conclusions to be drawn from the analysis for the actual scope for liquidity transformation through demand deposits depend on the interpretation of the productivity function R. If one believes that the aggregate of a given cohort of investment projects is relatively easily reversible, be it because of high aggregation levels, of substitution effects between inflowing and outflowing capital or the nature of investment itself, then the analysis of this paper has shown liquidity provision through deposit contracts to be severely limited, if not impossible. In this view, deposit contracts can provide no more liquidity than the real investment opportunities. Moreover, the gap between demand for liquidity and the scope for liquidity provision through banks may be substantial given the relatively high empirical estimates of individuals' relative risk aversion.

Appendix

Proof of Proposition 2:

I) Existence: By the incentive constraint (12), $r(t)$ is bounded globally from above if it is bounded locally at $t = 1$, which is straightforward to show. Therefore, we can assume that $r(t) \in [0, K]$, with K large. Hence, the problem is compact. The integrand in (3) and the right-hand side of (1) are Carathéodory functions (measurable in t, continuous in (S, r)). Hence, compactness existence theorems for variational problems such as Theorem 11.4.ii in Cesari (1983, p. 388) apply.

Throughout the remainder of the proof I assume that F is strictly increasing, i.e. that the measure dF has full support on $[0, 1]$. The general case then follows by approximation. Furthermore, uniqueness will be understood as up to changes on sets of Lebesgue measure 0. Let $\bar{r}$ be a solution to problem (3) – (5), (12).

II) The following property is helpful for characterizing the set $\{t; H(t; \bar{r}) > 0\}$:

Lemma 1: If there exists $t_0 \in [0, 1)$ such that $H(t_0; \bar{r}) > 0$, then there exists a $\hat{t} > t_0$ and a number $C > 0$ such that $\bar{r}(t) = q(\frac{C}{R(t)})$ on $[t_0, \hat{t}]$ and $H(\hat{t}; \bar{r}) = 0$.

Proof: By the strict concavity of u, $\bar{r}$ must be continuous. Therefore, $H(\cdot; \bar{r})$ is continuous. Since $H(1; \bar{r}) = 0$, there is an interval $[t_0, \hat{t})$ on which $H(t; \bar{r}) > 0$ and $H(\hat{t}; \bar{r}) = 0$. Take any $\beta \in (t_0, \hat{t})$. As in Proposition 1, the local first-best problem

$$\begin{aligned} &\max && \int_{t_0}^{\beta} u(r(\tau))dF(\tau) \\ &\text{subject to} && r \geq 0 \ \text{ integrable on } [t_0, \beta], \\ & && \int_{t_0}^{\beta} \frac{r(\tau)}{R(\tau)} dF(\tau) = \frac{S(t_0; \bar{r})}{R(t_0)} - \frac{S(\beta; \bar{r})}{R(\beta)}, \end{aligned}$$

has a unique solution $r_{t_0\beta}(t) = q(\frac{C}{R(t)})$. Since increasing $\int_{t_0}^{\beta} u(\bar{r}(\tau))dF(\tau)$ increases $H(t; \bar{r})$ on $[0, t_0]$, and since $H(t; \bar{r}) > 0$ on $[t_0, \beta]$, $\bar{r}|_{[t_0,\beta]} = r_{t_0\beta}$. Because β was arbitrary the claim follows.

III) Characterization of $\{t; H(t; \bar{r}) > 0\}$:[18)]

i) Suppose there exists $b \in \mathcal{B}$, $b < 1$, such that $H(b; \bar{r}) > 0$. By Lemma 1, let $\hat{t} > b$ be such that $\bar{r}(t) = q(\frac{C}{R(t)})$ on $[b, \hat{t}]$ and $H(\hat{t}; \bar{r}) = 0$. (13) and the Fundamental Theorem of Calculus imply that $H(t; \bar{r}) = \frac{1}{\gamma - 1} C^{\frac{\gamma-1}{\gamma}} G(t) + K$ on $[b, \hat{t}]$, with K constant. Since $G(t) \geq G(b)\ \ \forall t \geq b$ by the definition of $\mathcal{B}$, $H(\hat{t}; \bar{r}) \geq H(b; \bar{r}) > 0$, a contradiction. Hence, $H(b; \bar{r}) = 0\ \ \forall b \in \mathcal{B}$. The continuity of H further implies that $H(t; \bar{r}) = 0$ for all $t \in \mathrm{cl}\mathcal{B}$.

ii) Let $b \in \mathcal{B}$, $b > 0$. Then, $a(b) < b$. Suppose there is no $C > 0$ such that $\bar{r}(t) = q(\frac{C}{R(t)})$ on $[a(b), b]$, i.e. $\bar{r}$ is not locally first-best on $[a(b), b]$. Replace $\bar{r}$ by $q(\frac{C}{R(\cdot)})$ on $[a(b), b]$, with C determined by the feasibility condition

$$\int_{a(b)}^{b} \big(q(\frac{C}{R(\tau)})/R(\tau)\big) dF(\tau) = \frac{S(a(b); \bar{r})}{R(a(b))} - \frac{S(b; \bar{r})}{R(b)},$$

to obtain a strict improvement $\hat{r}$ on $[0, 1]$. Since $H(t; \bar{r}) \geq 0$ for all $t \in [0, a(b)]$, $H(t; \hat{r}) \geq 0$ on $[0, a(b)]$. By (14), the Fundamental Theorem of Calculus, and the definition of $a(b)$, the fact that $H(a(b); \hat{r}) > 0$ implies that $H(t; \hat{r}) > 0$ on $[a(b), b]$. Hence, $\hat{r}$ satisfies (11), hence is admissible, a contradiction. It follows that $\bar{r}(t) = q(\frac{C}{R(t)})$ on $[a(b), b]$.

By the same argument, since $H(b; \bar{r}) = 0$ by *i*) and $\bar{r}$ is continuous, (14) implies that $H(t; \bar{r}) > 0$ on $(a(b), b)$ and that $H(a(b); \bar{r}) = 0$ if $a(b) > 0$. By continuity, $H(a; \bar{r}) = 0$ for all $a \in \mathrm{cl}\{a(b) > 0; b \in \mathcal{B}\}$.

iii) Consider $a \in \mathrm{cl}\{a(b); b \in \mathcal{B}\}$, $0 < a \notin \mathcal{B}$. Let $b'(a) := \sup\{\beta \in \mathcal{B}; \beta < a\}$ be the next smaller element to a in cl $\mathcal{B}$ (because $a > 0$ the set over which the sup is taken is not empty). Note that $b'(a) \leq a$.

By the definition of $\mathcal{B}$ and $a(b)$, $G' \geq 0$ on $[b'(a), a]$. By *ii*), we have $H(a; \bar{r}) = 0$. Now suppose that there is $t_0 \in [b'(a), a]$ such that $H(t_0; \bar{r}) > 0$. By Lemma 1, there is an interval $[t_0, \hat{t}] \subset [t_0, a]$ on which $\bar{r}(t) = q(\frac{C}{R(t)})$ and $H(\hat{t}; \bar{r}) = 0$. Hence, by (14), $H'(t; \bar{r}) = \frac{1}{\gamma - 1} C^{\frac{\gamma-1}{\gamma}} G'(t) \geq 0$ on $(t_0, \hat{t})$. However, this is incompatible with $H(t_0; \bar{r}) > 0 = H(\hat{t}; \bar{r})$. Hence, $H(t; \bar{r}) \equiv 0$ on $[b'(a), a]$.

iv) Steps *i*) to *iii*) provide a complete characterization.

Proof: If $G' \geq 0$ on $[0, 1]$, then $\mathcal{B} = \{0\}$, and the same argument as in *iii*) shows that $H(t; \bar{r}) \equiv 0$ on $[0, 1]$.

If G is not monotonic on $[0, 1]$, then $\max \mathcal{B} > 0$. Take any $x \in [0, \max \mathcal{B}]$, $x \notin \mathrm{cl}\mathcal{B}$. If $x < \min \mathcal{B}$, *ii*) applies. If $x > \min \mathcal{B}$, let

$$u(x) := \inf\{b \in \mathcal{B}; b > x\}$$
$$l(x) := \sup\{b \in \mathcal{B}; b < x\}.$$

If $u(x) \in \mathcal{B}$, $[l(x), u(x)] = [l(x), a(u(x))] \cup [a(u(x)), u(x)]$, and either *ii*) or *iii*) applies.

If $u(x) \notin \mathcal{B}$ then necessarily $u(x) \in \mathrm{cl}\{a(b); b \in \mathcal{B}\}$, and *iii*) applies.

IV) Uniqueness:

The above characterization implies that for each interval $[a(b), b]$, $b \in \mathcal{B}$, there is a unique constant $C_b > 0$ such that $\bar{r}(t) = q(\frac{C_b}{R(t)})$ on these intervals.

Furthermore, on each interval $[b,a]$ on which $H \equiv 0$, differentiating yields

$$\bar{r}'(t) = \bar{r}(t) \int_t^1 R(\tau - t)^{-\gamma+1} g(\tau - t) dF(\tau) \Big/ \int_t^1 R(\tau - t)^{-\gamma+1} dF(\tau). \qquad (A1)$$

The right hand side of $(A1)$ is continuous in (t, r) (if $a = 1$, continuity at $t = 1$ follows by l'Hôpital's rule) and Lipschitz in r. Hence, for each initial value $\bar{r}(a)$ there is exactly one solution of $(A1)$.

Hence, if $\mathcal{B}$ is finite, $\bar{r}(1)$ determines $\bar{r}$ uniquely by a finite recursion. The resource constraint $S(1; \bar{r}) = 0$ then pins down $\bar{r}(1)$.

If $\mathcal{B}$ is infinite, consider the following approximation of $\bar{r}$.

Let $t_1, \ldots, t_K$ be the points of accumation of $\mathcal{B}$. Take a sequence $\epsilon_n \searrow 0$, ϵ_1 sufficiently small. For any n, construct $\bar{r}_n$ recursively by letting $\bar{r}_n(t) = q(\frac{C_b^n}{R(t)})$ on $[a(b), b]$ if $b \in \mathcal{B}$ and $[a(b), b] \cap \bigcup_{k=1,\ldots,K} (t_k - \epsilon_n, t_k + \epsilon_n) = \emptyset$, and by letting $H(t; \bar{r}_n) \equiv 0$ otherwise. The recursion is finite, therefore the resource constraint $S(1; \bar{r}_n) = 0$ determines $\bar{r}_n$ uniquely on $[0, 1]$. Since $\bar{r}_n \to \bar{r}$ uniformly, $\bar{r}$ is unique.

$\blacksquare$

Proof of Proposition 4: We have to show that G cannot have interior local minima. Letting

$$k(\tau, t) := R(t)^{\frac{1-\gamma}{\gamma}} R(\tau - t)^{1-\gamma} - R(\tau)^{\frac{1-\gamma}{\gamma}}$$

on $\{(\tau, t); 0 \le t \le \tau \le 1\}$, we can write

$$G(t) = \int_t^1 k(\tau, t) \mathrm{d}F(\tau).$$

By straightforward computation,

$$\frac{\gamma}{\gamma - 1} \big[k_\tau(\tau, t) + k_t(\tau, t)\big] \;=\; -g(t) k(\tau, t) + \big[g(\tau) - g(t)\big] R(\tau)^{\frac{1-\gamma}{\gamma}}.$$

Therefore, and since $k(t, t) = 0$,

$$\begin{aligned} G'(t) &= \int_t^1 k_t(\tau, t) \mathrm{d}F(\tau) \\ &= \frac{\gamma - 1}{\gamma} \int_t^1 \big[g(\tau) - g(t)\big] R(\tau)^{\frac{1-\gamma}{\gamma}} \mathrm{d}F(\tau) - \frac{\gamma - 1}{\gamma} g(t) G(t) - \int_t^1 k_\tau(\tau, t) \mathrm{d}F(\tau). \end{aligned}$$

Differentiating yields

$$
\begin{aligned}
G''(t) = &-\frac{\gamma-1}{\gamma}\int_t^1 g'(t)R(\tau)^{\frac{1-\gamma}{\gamma}}\mathrm{d}F(\tau) - \frac{\gamma-1}{\gamma}g'(t)G(t) - \frac{\gamma-1}{\gamma}g(t)G'(t) \\
&+ \frac{\gamma-1}{\gamma}R(t)^{\frac{1-\gamma}{\gamma}}\big[g(t)-\gamma g(0)\big]f(t) - \int_t^1 k_{t\tau}(\tau,t)\mathrm{d}F(\tau).
\end{aligned} \tag{A2}
$$

Now suppose that $G'(t) = 0$. Hence, the third term on the right hand side of $(A2)$ is zero. Consider the other terms in $(A2)$ in turn. First,

$$
\int_t^1 g'(t)R(\tau)^{\frac{1-\gamma}{\gamma}}\mathrm{d}F(\tau) + g'(t)G(t) = g'(t)\int_t^1 R(t)^{\frac{1-\gamma}{\gamma}}R(\tau-t)^{1-\gamma}\mathrm{d}F(\tau).
$$

Next, integrating by parts,

$$
-\int_t^1 k_{t\tau}(\tau,t)\mathrm{d}F(\tau) = -k_t(1,t)f(1) + k_t(t,t)f(t) + \int_t^1 k_t(\tau,t)f'(\tau)\mathrm{d}\tau. \tag{A3}
$$

The second term on the right hand side of $(A3)$ cancels with the fourth term of $(A2)$. To obtain an upper bound for the third term in $(A3)$ note that for each $t \in [0,1]$

$$
k_t(\tau,t) = \frac{\gamma-1}{\gamma}R(t)^{\frac{1-\gamma}{\gamma}}R(\tau-t)^{1-\gamma}\big[\gamma g(\tau-t) - g(t)\big]
$$

has at most one zero $\tau_0(t)$. Consequently, there exists a unique $m(t) \in [t,1]$ such that, for all $\tau < 1$, $k_t(\tau,t) \geq 0$ if and only if $\tau \in [m(t),1]$. Hence, we can split the integral into two parts to exploit the monotonicity of f'/f:

$$
\begin{aligned}
\int_t^1 k_t(\tau,t)f'(\tau)\mathrm{d}\tau \;&=\int_t^{m(t)} \underbrace{k_t(\tau,t)}_{<0}\frac{f'(\tau)}{f(\tau)}\mathrm{d}F(\tau) \;+\; \int_{m(t)}^1 \underbrace{k_t(\tau,t)}_{\geq 0}\frac{f'(\tau)}{f(\tau)}\mathrm{d}F(\tau) \\
&\leq \frac{f'(m(t))}{f(m(t))}\int_t^{m(t)} k_t(\tau,t)\mathrm{d}F(\tau) \;+\; \frac{f'(m(t))}{f(m(t))}\int_{m(t)}^1 k_t(\tau,t)\mathrm{d}F(\tau) \\
&= \frac{f'(m(t))}{f(m(t))}G'(t) \\
&= 0.
\end{aligned}
$$

Collecting terms, $(A2)$ and the above imply

$$
G''(t) \leq -\frac{\gamma-1}{\gamma}g'(t)\int_t^1 R(t)^{\frac{1-\gamma}{\gamma}}R(\tau-t)^{1-\gamma}\mathrm{d}F(\tau) - k_t(1,t)f(1). \tag{A4}
$$

Since $G'(t) = 0$, necessarily $m(t) < 1$ and $k_t(1,t) > 0$. Hence, if $g' > 0$ on $(0,1)$, $(A4)$ implies $G''(t) < 0$, and G has no interior minima. If $g'(t) = 0$ for some $t \in (0,1)$ the claim follows by approximation.

∎

Footnotes:

1) For examples of such models see Engineer (1989) and von Thadden (1999).

2) The assumption of aggregate certainty corresponds to the setup of the first part of Diamond and Dybvig's (1983) paper. For treatments of the problem with aggregate uncertainty, which are rarer and considered more difficult, see Wallace (1990) and Green and Lin (1999).

3) As I have discussed elsewhere (von Thadden, 1999), this phenomenon is a new incarnation of the old insight that "banks are useless in the Arrow-Debreu world" (Freixas and Rochet, 1997).

4) Note that in this case investors may very well have high demand for liquidity. The point is that investor moral hazard makes it impossible to supply liquidity. This is why I rather use the term "liquidity" than "maturity" transformation.

5) It is not difficult to incorporate transaction costs into the analysis. The main consequence of transaction costs would be a loosening of the households' incentive compatibility constraint, (12).

6) See section IV for further discussion.

7) See von Thadden (1998) for details of the proof.

8) Usually, the literature confounds these features. The only exception I know is Haubrich and King (1990) who disentangle risk aversion and intertemporal substitution.

9) It is possible to generalise much of the arguments below to the case of non-constant intertemporal relative risk aversion, although – as far as I can see – at a considerable expense. Apart from the fact that the value added of such extra generality is small anyhow, there is a more important argument that justifies restriction (10). As we shall see, even with the restriction to constant intertemporal relative risk aversion, the problem has fairly little general structure, and hence there is a need to find more, rather than less structural constraints. This is done in Propositions 3 and 4.

10) Strictly speaking, everybody attempts to withdraw "as soon after $t = 0$ as possible", since withdrawal at $t = 0$ is not possible. This trivial open-set problem causes non-existence of equilibrium rather than a bank-run equilibrium.

11) Note that if G were non-negative on all of $[0, 1]$ then, by (12), $H(t; r^\star) \geq 0 \ \ \forall t \in [0, 1]$, which is not true in general, as seen earlier.

12) The example has $g(t) = 2t$ and $\gamma = 2$ and has been produced with Mathematica. Note that the G in Figure 2 has the same structure as the (more complicated) G in Figure 1. Special thanks to Lars Stole for telling me how to get the Mathematica ImplicitFunctions out of their curly brackets.

13) This case, for which a simple direct proof is available, has been discussed in von Thadden (1998).

14) Caplin and Nalebuff (1991) provide a discussion of log-concavity in the more general context of ρ-concavity and give several economic applications. Heckman and Honoré (1990) discuss log-concavity from an econometric perspective in the context of the Roy model.

15) This is because of the Nash assumption that individual depositors cannot coordinate to withdraw and re-invest collectively. If ex-post coordination were possible, it could be possible to obtain some insurance even after premature withdrawal. However, such a deviant coalition would again be vulnerable to deviation by a sub-coalition, etc. ad infinitum, which makes this process complicated. But, as discussed in the introduction, coordination by depositors is of little concern in our context, since the role of banks is precisely to create intertemporal risk sharing possibilities for small and isolated individuals.

16) This case shows the conceptual difference between maturity transformation and liquidity transformation. A bank policy of holding only short-term liabilities and interrupting production whenever needed causes no deadweight loss in production here, hence, there is no need for maturity transformation. Yet, if $\gamma > 1$ there is demand for liquidity transformation.

17) A structurally similar problem has been considered by Hart and Moore (1994) in the context of debt renegotiation. There, an investor and a firm face an ongoing bargaining problem over a continuous repayment flow to the investor. Obviously, renegotiation is not an issue in the present model.

18) As mentioned earlier, the proof simplifies if one assumes that G has only a finite number of zeroes (similar to Guesnerie and Laffont (1984)). In this case, the argument is direct and can proceed by recursion.

References

Bencivenga, V.R. and B.D. Smith (1991), "Financial Intermediation and Endogenous Growth", *Review of Economic Studies* 58, 195–209.

Bhattacharya, S. and D. Gale (1987), "Preference Shocks, Liquidity, and Central Bank Policy", in: W.A. Barnett, K.J. Singleton (eds.), *New Approaches to Monetary Economics*, Cambridge: Cambridge University Press, 69–88.

Bryant, J. (1980), "A Model of Reserves, Bank Runs, and Deposit Insurance", *Journal of Banking and Finance* 4, 335–344.

Caplin, A. and B. Nalebuff (1991), "Aggregation and Social Choice: A Mean Voter Theorem", *Econometrica* 59, 1–23.

Cesari, L. (1983), *Optimization – Theory and Applications*, New York: Springer.

Diamond, D.W. (1997), "Liquidity, Banks, and Markets", *Journal of Political Economy* 105, 928–956.

Diamond, D.W. and Ph.H. Dybvig (1983), "Bank Runs, Deposit Insurance, and Liquidity", *Journal of Political Economy* 91, 401–419.

Engineer, M. (1989), "Bank Runs and the Suspension of Deposit Convertibility", *Journal of Monetary Economics* 24, 443–454.

Freixas, X. and J.C. Rochet (1997), *Microeconomics of Banking*, Cambridge (MA), London (UK): MIT Press.

Green, E.J. and P. Lin (1999), "Implementing Efficient Allocations in a Model of Financial Intermediation", manuscript, Federal Reserve Bank of Minneapolis.

Guesnerie, R. and J.J. Laffont (1984), "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm", *Journal of Public Economics* 25, 329–369.

Hart, O.D. and J. Moore (1994), "A Theory of Debt Based on The Inalienability of Human Capital", *Quarterly Journal of Economics* CIX, 840–879.

Haubrich, J. and R. King (1990), "Banking and Insurance", *Journal of Monetary Economics* 26, 361–386.

He, H. and H.F. Pagès (1993), "Labor Income, Borrowing Constraints, and Equilibrium Asset Prices", *Economic Theory* 3, 663–696.

Heckman, J.J. and B.E. Honoré (1990), "The Empirical Content of the Roy Model", *Econometrica* 58, 1121–1149.

Jacklin, Ch.J. (1987), "Demand Deposits, Trading Restrictions, and Risk Sharing", in: E.C. Prescott, N. Wallace (eds.), *Contractual Arrangements for Intertemporal Trade*, Minneapolis: University of Minnesota Press, 1987, 26–47.

Judd, K.L. (1985), "The Law of Large Numbers with a Continuum of IID Random Variables", *Journal of Economic Theory* 35, 19–25.

Merton, Robert C. (1990), *Continuous-Time Finance*, Blackwell: Cambridge, MA, and Oxford, UK.

Postlewaite, A. and X. Vives (1987), "Bank Runs as an Equilibrium Phenomenon", *Journal of Political Economy* 95, 485–491.

Qi, J. (1994), "Bank Liquidity and Stability in an Overlapping Generations Model", *Review of Financial Studies* 7, 389–417.

von Thadden, E.L. (1998), "Intermediated versus Direct Investment: Optimal Liquidity Provision and Dynamic Incentive Compatibility", *Journal of Financial Intermediation* 7, 177–197.

von Thadden, E.L. (1999), "Liquidity Creation through Banks and Markets: Multiple Insurance and Limited Market Access", *European Economic Review* 43, 991–1006.

Wallace, N. (1988), "Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model with Sequential Service Taken Seriously", *Quarterly Review, Federal Reserve Bank of Minneapolis*, Fall 1988, 3–16.

Wallace, N. (1990), "A Banking Model in Which Partial Suspension Is Best", *Quarterly Review, Federal Reserve Bank of Minneapolis*, Fall 1990, 11–23.

Wallace, N. (1996), "Narrow Banking Meets the Diamond-Dybvig Model", *Quarterly Review, Federal Reserve Bank of Minneapolis* 20, 3–13.