Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 177 (2018) 70–100
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

A theory of intermediated investment with hyperbolic discounting investors ☆

Feng Gao [a], Alex Xi He [b], Ping He [a,*]

[a] *Tsinghua University, School of Economics and Management, Beijing 100084, China*
[b] *Massachusetts Institute of Technology, United States*

Received 29 January 2016; final version received 10 May 2018; accepted 25 May 2018
Available online 4 June 2018

Abstract

Financial intermediaries may reduce welfare losses caused by hyperbolic discounting investors, who may liquidate their investment prematurely when the liquidation cost is low. In a competitive equilibrium, sophisticated investors are offered contracts with perfect commitment, and first best results are achieved; naïve investors are attracted by contracts that offer seemingly attractive returns in the long run but introduce discontinuous penalties for early withdrawal. If the investor types are private information, naïve investors withdraw early and cross-subsidize sophisticated investors. When a secondary market for long-term contracts opens for trading, financial intermediaries are compelled to offer contracts that have more flexible withdrawal options with linear schemes, and the welfare of naïve investors is improved. Arbitrage-free linear contracts allow for a unique term structure for interest rates that includes a premium for naïveté. Solvency requirements may limit competition for contracts and result in positive profits; banks that have capital are able to compete more aggressively, which improves investor welfare.

JEL classification: D1; D91; G2

Keywords: Investment; Financial intermediary; Hyperbolic discounting; Term structure

☆ The authors would like to thank the seminar participants at Tsinghua University and Beijing University. Feng Gao would like to thank the financial support from Tsinghua University Initiative Scientific Research Program (Grant No. 2015Z21063) and National Natural Science Foundation of China (Grant No. 71671101); Ping He would like to thank the financial support from Tsinghua University Initiative Scientific Research Program (Grant No. 2012Z04104 and 2015Z21065).

* Corresponding author.
E-mail address: heping@sem.tsinghua.edu.cn (P. He).

https://doi.org/10.1016/j.jet.2018.05.012

1. Introduction

Fisher (1930) regards investment as not an end in itself, but rather a process for distributing consumption over time. A number of studies suggest that agents have a time-inconsistent desire for immediate gratification and are often naïve about its consequences when making intertemporal decisions (Laibson, 1997; Frederick et al., 2002). Time inconsistent preferences often imply that consumers constrain their own future choices and generate non-desirable results.[1] In this study, we analyze investors' investment behavior through an intermediary when the investors have time-inconsistent preferences.

Our model has three dates. At date-0, there is zero time discounting between date-1 consumption and date-2 consumption. At date-1, there is positive time discounting between date-1 consumption and date-2 consumption, and this preference generates a desire for immediate gratification that creates liquidity needs. Sophistication is defined as the extent to which investors anticipate their preference changes. An investor can invest in an illiquid long-term asset that yields low returns if the investment lasts for a single period, but high returns if the investment lasts for two periods. In the autarky case, a sophisticated investor may under-invest at date-0 to avoid liquidation at date-1; a naïve investor may liquidate an investment at date-1 and receive a lower return to satisfy the desire for immediate gratification.

We focus on the role of financial intermediaries in offering financial products against early liquidation. We demonstrate that, in a competitive equilibrium,[2] financial intermediaries provide perfect commitment for sophisticated agents. However, if agents are even slightly naïve, financial intermediaries offer contracts that have seemingly attractive returns in the long run, but introduce a discontinuous penalty for early withdrawal.[3] We show that this discontinuity has adverse welfare consequences. Unsophisticated agents, even if they are nearly entirely sophisticated and are only slightly unable to predict their future behavior, fail to resist the temptation of immediate gratification, and are charged with penalties for early withdrawals.[4]

When a secondary market opens for trading long-term contracts, investors who experience a preference shock can exchange their future contract payment for present consumption. We demonstrate that this case is equivalent to the case where financial intermediaries offer a more flexible linear liquidation scheme instead of a contract with discontinuous early withdrawal penalties. Therefore, if contracts are restricted to a linear scheme, consumers have greater flexibility to transfer their consumption between two dates at a pre-specified ratio. Linear intervention

[1] These implications include under-saving (Laibson, 1998; Diamond and Köszegi, 2003; Salanie and Treich, 2006), over-borrowing (Heidhues and Köszegi, 2010), and the use of a commitment device for saving (Basu, 2011).

[2] A competitive equilibrium is defined as a set of contracts from which competing financial intermediaries have no incentive to deviate, which is similar to the equilibrium described in Heidhues and Köszegi (2010).

[3] Early withdrawal penalties are common in CD contracts and pension funds such as IRAs and 401(k) plans (*HelloWallet*, 2013). A commonly articulated justification for early withdrawal penalties is to provide a commitment device for time-inconsistent depositors that will prevent them from making impulsive withdrawals (Laibson, 1997; Ashraf et al., 2003). However, empirical studies demonstrate that early withdrawals from CD accounts occur at an economically significant level (Gilkeson et al., 1999) and suggest that certain depositors may have time inconsistent preferences that they are not fully aware of.

[4] There are also other explanations for nonlinear deposit contracts in the literature. Lin (1996) demonstrates that when investors have random discount factors, the optimal incentive-compatible risk-sharing contract has a convex structure, i.e., the interest rate is higher when there are fewer earlier withdrawals and a larger deposit balance. Ambrus and Egorov (2012) demonstrate that withdrawal penalties (money burning) are optimal when consumers face severe and rare negative liquidity shocks.

prevents financial intermediaries from imposing large discontinuous penalties on naïve agents, and improves their welfare.

We demonstrate that a linear contract implies a unique term structure of interest rates in our three-date model. Under certain conditions, the more naïve investors there are in the population, the lower is the implied short-term interest rate, and the higher the long-term interest rate, which suggests the presence of a term premium for naïveté. Intuitively, financial intermediaries balance their losses from sophisticated investors with their gains from naïve investors. When there are more naïve agents, financial intermediaries can afford greater losses from sophisticated investors, and will extract more effectively from each naïve agent by offering more attractive "decoy" long-term consumption, which leads to a higher term premium. Therefore, our study identifies a new source for term premiums.[5]

Our model also has implications for the role of the solvency requirement of financial intermediaries. If a financial intermediary must disclose its financial records at the end of each period, it can no longer offer an unrealistically high interest rate over a long period of time to deceive naïve agents, because the "decoy" contracts are not feasible if we impose a solvency requirement. In equilibrium, financial intermediaries that are subject to solvency restrictions may earn positive profits by exploiting naïve investors' failure to predict their own behavior over time.[6] We show that owning some capital allows financial intermediaries to compete more aggressively. In addition, we analyze a case in which two financial intermediaries with different capital costs compete imperfectly, and we demonstrate that the financial intermediary with lower capital cost offers contracts that include higher expected utility to investors, and has a larger market share.

Prior studies have shown that a competitive market offers commitment devices to sophisticated agents, and exploitative contracts to naïve agents with time inconsistencies.[7] Our results most closely align with Heidhues and Köszegi (2010), in which contracts and welfare are discontinuous for fully sophisticated investors, and even slightly naïve consumers will switch from their preferred repayments and are penalized ex post. Certain results of our investment model could be reinterpreted using the framework of the credit model in Heidhues and Köszegi (2010), with fixed unit endowments and investment returns. We emphasize the role of financial intermediaries and discuss related issues, such as the role of solvency requirements and capital. In addition, our setting focuses more on intertemporal allocation and makes new predictions regarding the term structure of interest rates.

[5] Prior studies suggest that the normal pattern of the yield curve is upward sloping (Ang et al., 2006). One commonly accepted explanation for term premium is the liquidity preference theory, i.e., risk averse investors prefer short-term maturity and require a premium to engage in long term lending. Numerous studies analyze the slope of the term structure. The term spread can predict future economic activity (Estrella and Hardouvelis, 1991) and explain the entire cross-section of bond yields (Duffie and Kan, 1996). Recent studies suggest that news shocks about future total factor productivity and the endogenous response of monetary policy are key drivers of the unpredictable movements in the term spread (Kurmann and Otrok, 2013).

[6] In a different setting, Heidhues et al. (2012) discuss the possibility of positive profits in a seeming competitive equilibrium while price floors exist. The solvency requirement and price floor limit the competition and result in non-zero profits.

[7] There are also many related studies with non-competitive markets. DellaVigna and Malmendier (2004) demonstrate that a monopolistic firm offers a two-part tariff to partially naïve consumers, such that the per-usage price falls below the firm's marginal cost, in the case of investment goods, and rises above the firm's marginal cost in the case of leisure goods. These contract features adversely affect consumer welfare only when consumers are naïve. Eliaz and Spiegler (2006) analyze a two-period model in which a firm screens agents by the probability of their attachment to each state ex ante or, more specifically, their sophistication. The optimal menu provides a perfect commitment device for relatively sophisticated agents, and exploitative contracts that involve speculation for relatively naïve agents.

Other bounded rationalities are also explored in the literature, which yield some similar results. Gabaix and Laibson (2006) analyze a model in which myopic consumers do not understand their contracts and are exploited by marketing schemes that conceal high-priced add-ons. Our model shares certain features of their model, because naïve investors do not foresee their preference changes. However, naïveté in our model and myopia in Gabaix and Laibson (2006) represent two different types of bounded rationality, which apply to different contracting environments. One important implication of the myopia model is that, in equilibrium, firms hide information that could affect myopic consumers' welfare; however, in our naïveté model, financial intermediaries compete transparently, without hiding information. In addition, our study focuses on the role of financial intermediaries as delegated investors instead of firms who offer products directly.[8]

Finally, our model shares some similar features to Diamond and Dybvig (1983) in which banks provide insurance against preference shocks. This study analyzes the role of financial intermediaries when preference shocks are generated by time-inconsistent preferences. However, there are at least two primary differences: First, although the Diamond–Dybvig model includes uncertainties regarding whether an agent is patient or impatient ex ante, preference shocks that are generated by hyperbolic discounting are a certainty. The change of utility function is a shock only for unsophisticated agents who fail to predict their future utility. Second, a financial intermediary's role is to offer protection against early liquidation in this study rather than to provide risk sharing among agents with different preferences.

The remainder of the paper is organized as follows. Section 2 introduces our basic three-date model that includes time-inconsistent preferences and we characterize the first best case and the autarky case. Section 3 characterizes competitive equilibrium with financial intermediaries and derives both an unrestricted nonlinear deposit contract and a restricted linear contract in competitive equilibrium. In Section 4, we analyze the implied term structure for linear contracts. Section 5 provides the results of the extended models with a solvency requirement. Section 6 concludes with our final statements. All proofs are provided in the Appendix.

2. Basic model

The economy has three dates ($t = 0, 1, 2$) and a single homogeneous good. There is a continuum of agents with measure 1. Each agent possesses one unit of good at date 0 and the good is consumed at dates 1 and 2. The good can be stored from one date to the next or invested at $t = 0$ in a long-run technology that returns $R > 1$ units at $t = 2$ and liquidation is costly: an agent obtains L (≤ 1) units for each unit of investment liquidated at $t = 1$.

The agents have time-inconsistent preferences. Self 0's utility is represented by $u(c_1) + u(c_2)$, where $c_1 \geq 0$ and $c_2 \geq 0$ represent consumption at date 1 and date 2, respectively. Self 1 maximizes $u(c_1) + \beta u(c_2)$, where $0 \leq \beta \leq 1$ denotes the hyperbolic discount factor. The per-date utility function $u(\cdot)$ is strictly concave and twice differentiable, with $u(0)$ bounded and $u'(0) = \infty$.

In alignment with O'Donoghue and Rabin (2001), we assume that self 0 believes with certainty that self 1 will maximize $u(c_1) + \hat{\beta} u(c_2)$, where $\beta \leq \hat{\beta} \leq 1$. The parameters $(\beta, \hat{\beta})$ reflect

[8] Similar to Gabaix and Laibson (2006), Gottlieb and Smetters (2014) assume that consumers in the insurance market do not consider liquidity shocks ("background risk"). As a consequence, insurers profit from policyholders who lapse, but incur losses from policyholders who retain their policies to term or death. Premiums are initially expensive and decrease over time; insurers strongly oppose secondary markets despite the improvement in policyholders' welfare.

self 0's beliefs about his/her time inconsistency; therefore, an agent is naïve if the agent's preference turns out to be $u(c_1) + \beta u(c_2)$ with $\beta < \hat{\beta}$ at date 1 and an agent is sophisticated if his/her preference change is fully expected, i.e., $\beta = \hat{\beta}$.

We measure welfare using self 0's long-run preferences in alignment with studies regarding time inconsistency (DellaVigna and Malmendier, 2004; O'Donoghue and Rabin, 2006). The reasons for using this measurement are similar to Heidhues and Köszegi (2010). First, in reality, time inconsistency occurs over many periods; therefore, weighting each period equally would be more reasonable only if the model included more than two periods. Second, the desire for immediate gratification is often a mistake and does not reflect true welfare; therefore, measuring welfare by self 1's or self 2's utility may not be appropriate.

2.1. The first best allocation

We first consider the first best allocation in our model. For this allocation, there is no liquidation at date 1 and the welfare-maximizing allocation is determined by the following calculation:

$$\begin{aligned} &\max_{c_1, c_2} u(c_1) + u(c_2) \\ &\text{s.t. } c_2 = R(1 - c_1) \end{aligned} \tag{1}$$

We can easily verify that the first best allocation satisfies $-u'(c_1^{fb}) + u'(c_2^{fb})R = 0$, where $c_2^{fb} = R(1 - c_1^{fb})$.

Because the first best allocation does not depend on the degree of time inconsistency (β) or the sophistication ($\hat{\beta}$) of an agent, a social planner can maximize welfare by investing on behalf of all agents and offering each agent c_1^{fb} and c_2^{fb} at date 1 and date 2. In the following sections, we consider agents' welfare when agents make their own investments and when the model includes competitive profit-maximizing financial intermediaries.

2.2. The autarky case

For comparison, we derive autarky outcomes without considering financial intermediaries. At date 0, agents solve the following maximization problem:

$$\begin{aligned} &\max_{I} u(c_1) + u(c_2) \\ &\text{s.t. } c_1 = 1 - I + \hat{x}(I)L \\ &\quad\;\; c_2 = R\big(I - \hat{x}(I)\big) \end{aligned} \tag{2}$$

where I represents the investment at date 0 and $\hat{x}(I)$ represents the optimal liquidation unit that is determined by the following optimization problem:

$$\begin{aligned} &\max_{\hat{x} \geq 0} u(c_1) + \hat{\beta} u(c_2) \\ &\text{s.t. } c_1 = 1 - I + \hat{x}L \\ &\quad\;\; c_2 = R(I - \hat{x}) \end{aligned} \tag{3}$$

Lemma 1. *At date 0, an agent chooses an investment level with the expectation that there will be no liquidation at date 1 (i.e., $\hat{x} = 0$).*

Proof. See Appendix. □

Intuitively, a strictly positive $\hat{x}$ cannot be optimal because the agent can always improve his expected utility by decreasing I and $\hat{x}$.

At date 1, given I, an agent chooses the optimal liquidation based on the realized time preference and the optimization problem at date 1 can be calculated as follows:

$$\begin{aligned} &\max_{x\geq 0} u(c_1) + \beta u(c_2) \\ &\text{s.t. } c_1 = 1 - I + xL \\ &\qquad c_2 = R(I - x) \end{aligned} \tag{4}$$

A sophisticated agent who correctly predicts his/her preference change will not liquidate any investments. However, a naïve agent may liquidate a portion of his/her investment at date 1. The following proposition summarizes the primary results for the autarky investment decision and liquidation decision.

Proposition 1. *For the autarky case, if $L \leq \hat{\beta}$, the initial investment at date 0 is identical to the first best investment; if $L > \hat{\beta}$, the initial investment at date 0 satisfies $\frac{u'(c_1)}{u'(c_2)} = \frac{u'(1-I)}{u'(IR)} = \frac{\hat{\beta}R}{L}$ and is less than the first best investment. Sophisticated agents will not liquidate at date 1, but naïve agents will liquidate a portion of the initial investment if $L > \beta$.*

Proof. See Appendix. □

Intuitively, when the liquidation cost is low, if the first best investment level were chosen at date 0, the agents would expect that they would liquidate at date 1, which is not optimal and they will choose to invest less than the first best level at date 0 to avoid liquidation at date 1. Concurrently, when liquidation is very costly, naïve agents will not liquidate at date 1 even if they become less patient; however, if the liquidation cost is low, naïve agents will liquidate at date 1 once their preference shock is realized.

Proposition 1 indicates that sophisticated agents and naïve agents will choose a lower investment level than the first best case if $L > \hat{\beta}$; naïve agents will further deviate from the first best case by liquidating a portion of their initial investment if $L > \beta$. Therefore, when the liquidation cost is low, neither sophisticated nor naïve agents can achieve the first best outcome. We will demonstrate in Section 3 how financial intermediaries can improve the welfare for both types of agents in a competitive equilibrium.

2.3. Heterogeneous agents in a trading market

We now consider a market where the agents can trade claims on date 2 consumption at date 1 and seek to determine if the market can help the agents improve their welfare when the autarky outcomes are not first best. We assume that there are two types of agents with different levels of sophistication. At date 0, both types of agents anticipate that their preference at date 1 will be $u(c_1) + \hat{\beta}u(c_2)$, with $\hat{\beta} \leq 1$. Among all agents, a fraction λ of them are naïve with the realized discount factor $\beta_1 = \beta < \hat{\beta}$ and the remaining $1 - \lambda$ agents are sophisticated with the realized discount factor $\beta_2 = \hat{\beta}$.

Proposition 2. *Trading does not occur between naïve agents and sophisticated agents if $L \geq \hat{\beta}$ or $L \leq \beta$. If $\beta < L < \hat{\beta}$, naïve agents will exchange date 2 consumption for date 1 consumption and the investment is not liquidated.*

Proof. See Appendix. □

Proposition 2 indicates that when the liquidation cost is relatively small ($L \geq \hat{\beta}$), despite naïve agents' liquidity needs, sophisticated agents' consumption plan is already intertemporally optimal and sophisticated agents will not trade unless the price of the date 2 consumption is very low, under which, however, naïve agents would prefer to liquidate their own investments. When the liquidation cost is relatively high ($L \leq \beta$), both naïve and sophisticated agents achieve the same allocation, which is first best and they will not trade with each other. When the liquidation cost satisfies $\beta < L < \hat{\beta}$, naïve agents, who have an incentive to liquidate, will trade with sophisticated agents and liquidation is avoided. In this case, the realized consumptions are identical to the case where agents invest through competitive financial intermediaries and trade their contracts in a secondary market, which is discussed in Section 3.2.

In Section 3, we introduce competitive financial intermediaries into the model to study whether financial intermediaries can improve agents' welfare. Competitiveness among financial intermediaries guarantees that all surplus is acquired by the economic agents rather than financial intermediaries. When financial intermediaries know agent types, we find that financial intermediaries improve agents' welfare. When the liquidation cost is low, sophisticated agents profit more by investing through a financial intermediary than by making their own investments because a financial intermediary can help them commit against their desire for immediate gratification by limiting their choices or charging an early withdrawal penalty. Concurrently, the welfare of naïve agents improves and they are attracted by contracts with seemingly high long-term payoffs at date 0 but eventually suffer from high early withdrawal penalties; however, financial intermediaries do not liquidate the investment. When a secondary market for long-term contracts opens for trading, which is equivalent to the case that financial intermediaries offer more options for early withdrawals on a linear scheme, naïve agents can obtain a higher welfare than for the autarky case. When both naïve and sophisticated agents exist and financial intermediaries cannot differentiate between them, an early withdrawal penalty acts as a commitment device for sophisticated agents but exploits agents' inaccurate expectations regarding their future behavior.

3. Investment through competitive financial intermediaries

3.1. Competitive equilibrium without secondary market

In this subsection, we assume that there are a continuum of competitive financial intermediaries with measure 1. We consider a competitive market in which agents sign deposit contracts with risk-neutral and profit-maximizing financial intermediaries and agents cannot trade their deposits contracts in a secondary market. Financial intermediaries possess the same long-run technology that is described above. Agents invest all their endowment in financial intermediaries at date 0 and sign unrestricted nonlinear contracts on their consumption schedule.[9] The contracts

[9] It is without loss of generality to not allow an agent to invest. The agent will choose not to invest on his/her own even if he/she is allowed to do so as the "decoy" consumption plan offered by financial intermediaries is more attractive in equilibrium.

are exclusive; once an agent signs a contract with a financial intermediary, that agent cannot interact with other financial intermediaries. To clarify, financial intermediaries offer consumption options in exchange for the entire amount the agent can invest and the agent cannot split the investment between different intermediaries.[10] In Section 3.2 we introduce a secondary market for contracts and demonstrate that equilibrium contracts possess a linear structure.

In alignment with Heidhues and Köszegi (2010), we define competitive equilibrium as the following. Formally, because only two possible hyperbolic discounting factors exist, β and $\hat{\beta}$, financial intermediaries will offer two consumption options $C = \{(c_1(s), c_2(s))\}_{s\in\{\beta,\hat{\beta}\}}$ to each agent at date 0, where $c_1 \geq 0$ and $c_2 \geq 0$ are consumptions at dates 1 and 2, respectively. We refer to the maps as follows $(c_1(\cdot), c_2(\cdot)) : \beta, \hat{\beta} \to R_+$ are incentive compatible if agents with hyperbolic discounting factor s prefer $(c_1(s), c_2(s))$, i.e., $u(c_1(s)) + su(c_2(s)) \geq u(c_1) + su(c_2)$, for all $(c_1, c_2) \in C$. Therefore, an agent of type $(\hat{\beta}, \beta)$ believes at date 0 that he/she would choose $(c_1(\hat{\beta}), c_2(\hat{\beta}))$ at date 0 and prefers $(c_1(\beta), c_2(\beta))$ at date 1.

Definition 1. A competitive equilibrium includes a set of consumption options $C = \{(c_1(s), c_2(s))\}_{s\in\{\beta,\hat{\beta}\}}$ that are offered by financial intermediaries and an implied incentive compatible map, $(c_1(\cdot), c_2(\cdot))$, that satisfies the following properties:

1. Zero profit: for each financial intermediary C yields zero expected profits.
2. No profitable deviation: there exists no contract C' with an incentive compatible map, $(c_1'(\cdot), c_2'(\cdot))$, such that $u(c_1'(s)) + su(c_2'(s)) \geq u(c_1(s)) + su(c_2(s))$ for all $s \in \{\beta, \hat{\beta}\}$ with strict inequality for $s = \hat{\beta}$, and C' yields positive profits.
3. Non-redundancy: for each consumption option, $(c_{1j}, c_{2j}) \in C$, there is a corresponding type $(\hat{\beta}, \beta)$ such that either $(c_{1j}, c_{2j}) = (c_1(\beta), c_2(\beta))$ or $(c_{1j}, c_{2j}) = (c_1(\hat{\beta}), c_2(\hat{\beta}))$.

In the above definition, the first two conditions indicate that by offering these contracts, financial intermediaries can do no better and they make zero expected profits. The third condition indicates that all consumption options are relevant in that they affect the expectations or behaviors of agents. Because of the non-redundancy condition, many options are excluded from the competitive-equilibrium contracts; specifically, non-sophisticated agents can only change their consumption options by paying a large early withdrawal penalty as discussed below.

We now characterize certain features of competitive equilibrium that do not include a secondary market for trading deposit contracts. We consider a case where only one type $(\hat{\beta}, \beta)$ exists. This case is equivalent to the full information case, where financial intermediaries are aware of each agent's type.

First, we consider a case where all agents are sophisticated (i.e., $\beta = \hat{\beta}$). Because a sophisticated agent would correctly predict his/her choice in date 1, only his/her chosen consumption option is relevant in both periods. We assume that the contract offered by a financial intermediary only includes one consumption option that the agent can choose. Financial intermediaries compete to the extent that there is zero profit, i.e., each financial intermediary solves the following problem:

[10] If we allow an agent to split his/her investment between different intermediaries, it is equivalent to allowing the agent to partially liquidate his/her contract.

$$\begin{aligned} &\max_{c_1,c_2} R(1-c_1)-c_2 \\ &\text{s.t. } u(c_1)+u(c_2)\geq \underline{u} \quad (PC) \end{aligned} \tag{5}$$

where $\underline{u}$ ensures that the maximum profit is zero.

When all agents are naïve (i.e., $\beta < \hat{\beta}$), agents do not accurately predict their discount factor and the consumption option they will choose at date 1. Financial intermediaries offer these agents a former "decoy" consumption option $(\hat{c}_1, \hat{c}_2)$ that self 0 expects to choose and a latter "chosen" consumption option (c_1, c_2) that self 1 chooses. Each financial intermediary solves the following profit maximization problem:

$$\begin{aligned} &\max_{c_1,c_2,\hat{c}_1,\hat{c}_2} R(1-c_1)-c_2 \\ &\quad u(\hat{c}_1)+u(\hat{c}_2)\geq \underline{u} \quad (PC) \\ &\text{s.t. } u(\hat{c}_1)+\hat{\beta}u(\hat{c}_2)\geq u(c_1)+\hat{\beta}u(c_2) \quad (PCC) \\ &\quad u(c_1)+\beta u(c_2)\geq u(\hat{c}_1)+\beta u(\hat{c}_2) \quad (IC) \end{aligned} \tag{6}$$

where $\underline{u}$ ensures that the maximum profit is zero.

The participation constraint (PC) is based on an incorrectly forecasted future choice of self 0. The perceived-choice constraint (PCC) implies that at date 0, agents believe that they would prefer the decoy option because of their belief, $\hat{\beta}$. The standard incentive-compatibility constraint (IC) implies that agents actually choose another option at date 1. PC and IC are binding and PCC in (6) is equivalent to $\hat{c}_1 \leq c_1$ or $\hat{c}_2 \geq c_2$ with $\beta < \hat{\beta}$. Intuitively, if self 1 is indifferent between two consumption options, then self 0, who overestimates his/her future patience, predicts that he/she will prefer the option that provides more consumption at a later date.

Proposition 3. *In a competitive equilibrium without a secondary market: (i) If only sophisticated agents exist in the economy, the competitive equilibrium outcome with financial intermediaries yields the first best results and financial intermediaries will offer a contract with full commitment at date 0 without an option to liquidate at date 1. (ii) If only naïve agents exist in the economy, the competitive-equilibrium contract includes two consumption options* $(\hat{c}_1, \hat{c}_2)$ *and* (c_1, c_2)*. At date 0, agents choose* $(\hat{c}_1, \hat{c}_2)$*, which satisfies* $\hat{c}_1 = 0$, $\hat{c}_2 > 0$ *and* $u(\hat{c}_1) + u(\hat{c}_2) = \underline{u}$*; at date 1, agents choose* (c_1, c_2)*, which satisfies* $u(c_1) + \beta u(c_2) = u(\hat{c}_1) + \beta u(\hat{c}_2)$, $\frac{u'(c_1)}{u'(c_2)} = \beta R$ *and* $R(1 - c_1) = c_2$.

Proof. See Appendix. □

The above proposition indicates that sophisticated agents rationally choose the contract that will help them commit to the first best consumption. Compared to the autarky model, sophisticated agents profit more in terms of the date-0 utility because financial intermediaries help them commit to the first best allocation.

For naïve agents, the contract offers an option for them to consume less in the short run, but introduces an option to withdraw in advance for a significant penalty. Agents maximize their utility by selecting the "decoy" consumption option, but financial intermediaries' profits are determined by the "chosen" consumption option; a discontinuity always exists at full sophistication as in Heidhues and Köszegi (2010): all non-sophisticated agents including near-sophisticated agents with β very close to $\hat{\beta}$, receive discretely different contracts than fully sophisticated agents who obtain the first best outcome.

The properties of the equilibrium contracts closely resemble certain important features of CD contracts that are offered by banks. For example, the penalty of early withdrawal from a one-year CD is often three months' interest. CDs promise a favorable return when the investment cannot be liquidated for a fixed period, but if the depositor withdraws even one dollar prior to maturity, a substantial penalty would be charged and investing in CDs becomes extremely unprofitable. Similarly, pension funds also include penalties for early withdrawals.

The following corollary contains the results related to welfare.

Corollary 1. *In a competitive equilibrium without a secondary market: (i) If only sophisticated agents exist in the economy, sophisticated agents obtain the first best utility. (ii) If only naïve agents exist in the economy, naïve agents obtain the same realized date-0 utility as in the autarky case if the liquidation cost is 0 with* $L = 1$*. If the liquidation cost is not very high (i.e.,* $1 > L > \beta$*) and a naïve agent's relative risk aversion is sufficiently large, he/she obtains a higher realized date-0 utility than the autarky case; if the liquidation cost is high (i.e.,* $L < \beta$*) he/she obtains a lower realized date-0 utility than the autarky case, in which the first best allocation is achieved.*

Proof. See Appendix. □

The equilibrium contract offers an option for naïve agents to consume little in the short term but includes an option that allows an agent to withdraw early from the contract for a significant penalty. Because financial intermediaries design the contract to induce consumption behavior that self 0 does not expect, its goal with the chosen option is to maximize the gains from trade with self 1, which satisfies the IC constraint and caters to self 1's desire for immediate gratification. Therefore, regardless of how naïve an agent is, competitive financial intermediaries will help the agent achieve the optimal realized date-1 utility. This utility coincides with the utility the agent obtains in the autarky economy when there is no liquidation cost because the savings offered by financial intermediaries (in the case of intermediated investment) is now equivalent to costless liquidation by the agent (in the autarky case).

When the cost for liquidation is sufficiently large, an intermediated investment is worse than the autarky investment because the high liquidation cost efficiently prevents naïve agents from liquidating their investments and ensure that they commit to the first best consumption in the autarky case. When the cost for liquidation is small, competitive financial intermediaries will help naïve agents to obtain greater date-1 utility by avoiding liquidation costs, but concurrently hurt naïve agents in regards to the perspective of date-0 utility because of misaligned intertemporal consumption allocation. The proof indicates that higher relative risk aversion provides naïve agents with a greater incentive to liquidate; a financial intermediary can reduce their liquidation cost thereby improving their welfare. Large relative risk aversion is a sufficient condition for the results to hold. When the relative risk aversion is small, naïve agents' realized date-0 utility may be smaller than for the autarky case.

Next, we consider competitive equilibrium when two types of agents exist and financial intermediaries do not observe the types: two types of agents exist that have different levels of sophistication. Both agents anticipate $\hat{\beta}$ at date 0; a proportion λ of these agents are naïve with realized discount factor $\beta_1 = \beta < \hat{\beta}$ and the remaining agents, $1 - \lambda$, are sophisticated at a level of $\beta_2 = \hat{\beta}$. Because sophisticated and naïve agents have the same belief at date 0, they accept the same contract. Each financial intermediary's problem is calculated as follows:

$$
\begin{aligned}
&\max_{c_1,c_2,\hat{c}_1,\hat{c}_2} \lambda\big[R(1-c_1)-c_2\big]+(1-\lambda)\big[R(1-\hat{c}_1)-\hat{c}_2\big] \\
&\qquad u(\hat{c}_1)+u(\hat{c}_2)\geq \underline{u} \quad (PC) \\
&\text{s.t. } u(c_1)+\beta u(c_2)\geq u(\hat{c}_1)+\beta u(\hat{c}_2) \quad (IC_1) \\
&\qquad u(\hat{c}_1)+\hat{\beta} u(\hat{c}_2)\geq u(c_1)+\hat{\beta} u(c_2) \quad (IC_2)
\end{aligned}
\tag{7}
$$

Note that IC_2 in (7) includes the same expression as PCC in (6). Indeed, IC_2 in (7) serves two roles; one role is the incentive constraint for sophisticated agents and the other role is the perceived-choice constraint for naïve agents, which is identical for PCC in (6). Clearly, PC and IC_1 in (7) must bind in equilibrium and IC_2 in (7) is equivalent to $\hat{c}_1 \leq c_1$ or $\hat{c}_2 \geq c_2$ with $\beta < \hat{\beta}$.

Proposition 4. *If financial intermediaries cannot distinguish between sophisticated and naïve agents in the economy, in a competitive equilibrium without a secondary market, financial intermediaries offer two consumption options, $(\hat{c}_1, \hat{c}_2)$ and (c_1, c_2) that satisfy the following conditions:*

$$
\begin{aligned}
&\lambda\big[R(1-c_1)-c_2\big]+(1-\lambda)\big[R(1-\hat{c}_1)-\hat{c}_2\big]=0 \\
&u'(c_1)=\beta R u'(c_2) \\
&\frac{u'(\hat{c}_1)}{u'(\hat{c}_2)}=\left(1+\frac{\lambda(1-\beta)u'(\hat{c}_1)}{(1-\lambda)u'(c_1)}\right)R \\
&u(c_1)+\beta u(c_2)=u(\hat{c}_1)+\beta u(\hat{c}_2)
\end{aligned}
$$

Both types of agents choose the consumption option $(\hat{c}_1, \hat{c}_2)$ at date 0 and naïve agents change to the consumption option (c_1, c_2) at date 1.

Proof. See Appendix. □

The third condition in Proposition 4 indicates that sophisticated agents' consumption schedules are too back-loaded even from the long-term self 0's perspective. Sophisticated agents' consumption is distorted from the date-0 perspective to attract naïve agents, the more profitable type of agent from the perspective of financial intermediaries. When considering whether to allocate more of a sophisticated agent's consumption to date 2, financial intermediaries are confronted with a trade-off: an unbalanced consumption schedule $(\hat{c}_1, \hat{c}_2)$ with large $\hat{c}_2$ serves better as a "decoy" consumption option to seduce naïve agents and increases financial intermediaries' profits; conversely, the total consumption level for sophisticated agents is increased, which decreases financial intermediaries' profits. A higher proportion of naïve agents, λ, leads to more distortion in sophisticated agents' consumption.

The following corollary discusses the results regarding welfare.

Corollary 2. *If financial intermediaries cannot distinguish between sophisticated agents and naïve agents in the economy, financial intermediaries profit from naïve agents, but incur losses from sophisticated agents. The date-0 utility of sophisticated agents is higher than the first best level and increases in the proportion of naïve agents λ, but the realized date-0 utility of a naïve agent is less than in the autarky economy if liquidation costs are sufficiently small or sufficiently large.*

Proof. See Appendix. □

Financial intermediaries profit from naïve agents and incur losses from sophisticated agents; an increase in the proportion of naïve agents λ would results in positive profits when the contracts stay unchanged. Competition drives financial intermediaries to offer more attractive contracts at date 0, which drives up the date-0 utility for sophisticated agents, who correctly anticipate their preference change. However, the date-0 utility of a naïve agent might not be a monotonic function of the proportion of naïve agents. When more naïve agents exist in the economy, competition leads to more attractive "decoy" contracts for naïve agents at date 0, who suffer more at date 1; conversely, more naïve agents subsidize sophisticated agents and each naïve agent may suffer less. For moderate liquidation costs, whether financial intermediaries hurt naïve agents depends on the specific form of the utility function.

3.2. *Competitive equilibrium with a secondary market*

In the previous subsection, where a secondary market does not exist for contracts, financial intermediaries offer a contract that does not offer agents many options to choose from at date 1; they experience a preference shock that requires immediate gratification. This corresponds to a CD contract that includes an early withdrawal penalty. However, certain CDs, such as brokerage CDs, can be traded on a secondary market. For this type of CD, if an agent wants to liquidate the contract prior to maturity, the agent can sell a portion of the contract in a secondary market rather than paying the early withdrawal penalty. Therefore, the presence of a secondary market for a contract may change the optimal contract if we assume that all agents and financial intermediaries know that a secondary market exists for contracts.

In an economy that includes a secondary market that operates at date 1, each agent signs an anonymous contract with a financial intermediary at date 0. Intermediaries obtain goods from agents and make investments. At date 1, agents who are unsatisfied with their contracts exchange a portion of their contracts for other contracts that are available in the secondary market. The equilibrium for this economy requires a competitive equilibrium for financial intermediaries and a competitive equilibrium in the secondary market, which is defined below.

Definition 2. Competitive equilibrium in an economy that includes a secondary market is represented by a set of consumption options $C = \{(c_1(s), c_2(s))\}_{s\in\{\beta,\hat{\beta}\}}$ that are tradable at date 1 in a secondary market and corresponding prices that satisfy the following properties:

1. Zero profit: for each financial intermediary, C yields zero profits.
2. No profitable deviation: there does not exist a tradable contract C' that is strictly preferred by agents at date 0 and yields positive profits for financial intermediaries.
3. Market clearing: the prices for consumption options are priced such that the market is cleared and the utility for each agent is maximized.

Heidhues and Köszegi (2010) propose that borrowers are penalized less when they are allowed to choose from linear repayment schemes; the penalties in our model can be reduced by opening up a secondary market. Linear contracts that have more options for agents to choose from based on a linear scheme when date 1 arrives, are also prevalent in deposit contracting and pension funds. An example of a linear contract is a demand deposit, which allows depositors to withdraw any amount at a pre-specified interest rate. An example is an enhanced or "flex" CD, which has been recently introduced (Brooks, 1996; Cline and Brooks, 2004). This type of CD offers an option to withdraw early once without penalty. Although enhanced CDs were introduced to reduce liquidity or interest rate risks, to a certain extent they restrict CD contracts to linear structures by allowing depositors to withdraw early at a specified interest rate. Pension fund deposit contracts also have a linear structure. For example, when an investor withdraws earnings from a new Roth 401(k) or 403(b) account before the age of 59 1/2, the funds are subject to income tax and a 10 percent early withdrawal penalty; the 10 percent penalty resembles the linear contract in our model. A competitive equilibrium with restricted linear contracts is formally defined as follows.

Definition 3. A competitive equilibrium in the economy with restricted linear contracts is a linear scheme $\{\tilde{R}, T\}$ offered by the financial intermediaries, and an agent has the right to choose any non-negative consumption schedule (c_1, c_2) that satisfies $c_1 + c_2/\tilde{R} = T$ at date 0 and date 1 that satisfies the following properties:

1. Zero profit: for each financial intermediary, $\{\tilde{R}, T\}$ yields zero profits.
2. No profitable deviation: at date 0, all agents choose $(\hat{c}_1, \hat{c}_2)$ to optimize $u(\hat{c}_1) + \hat{\beta} u(\hat{c}_2)$ subject to $\hat{c}_1 + \hat{c}_2/\tilde{R} = T$ and at date 1 naïve agents switch to (c_1, c_2) to optimize $u(c_1) + \beta u(c_2)$ subject to $c_1 + c_2/\tilde{R} = T$; there are no other linear schemes $\{\tilde{R}', T'\}$ associated with the realized incentive compatible consumption choices, $(\hat{c}_1', \hat{c}_2')$ and (c_1', c_2'), of sophisticated agents and naïve agents, respectively, and $\{\tilde{R}', T'\}$ is strictly preferred by agents at date 0 and yields positive profits.

The following Proposition indicates that when naïve and sophisticated agents exist, the equilibrium outcome with unrestricted contracts that are tradable is equivalent to the outcome with restricted linear contracts. This result suggests that if a secondary market opens at date 1 to trade date-2 consumption, competitive financial intermediaries have to offer a linear contract with more liquidation options for agents to choose from.

Proposition 5. *If financial intermediaries cannot distinguish between sophisticated and naïve agents in the economy, and contracts offered by financial intermediaries can be traded on a secondary market at date 1, in the competitive equilibrium agents obtain the same allocations as when the contracts are restricted to be linear.*

Proof. See Appendix. □

Intuitively, when agents can trade date-2 consumption for date-1 consumption under a certain price, their choice set is a linear set. In equilibrium, supply equals demand and all investors maximize their date-1 utility along the linear set; therefore, tradable CDs are equivalent to restricted linear deposit contracts. Note that at date 1, naïve and sophisticated agents have different pricing for the date-2 consumption if they do not trade their unrestricted contract; naïve agents are more tempted to sell their date-2 consumption, but all agents prefer to liquidate a portion of their date-2 consumption through trading. As a consequence, sophisticated agents will obtain more date-2 consumption from naïve agents through trading in a secondary market and financial intermediaries will not gain from naïve agents as much as in the case without a secondary market because a portion of naïve agents' date-2 consumption is sold to sophisticated agents. Therefore, the optimal contract discussed in the previous section cannot be a result of the equilibrium outcome in the presence of a secondary market.

We now discuss certain properties of the competitive equilibrium with a secondary market. According to the Duality Principal, the optimization problem for the intermediary in competitive equilibrium with a secondary market, equivalently, with a linear restriction, can be expressed as follows:

$$\begin{aligned}
&\max_{T,\tilde{R}} \lambda\big[R(1-c_1)-c_2\big]+(1-\lambda)\big[R(1-\hat{c}_1)-\hat{c}_2\big] \\
&\quad u(\hat{c}_1)+u(\hat{c}_2)\geq \underline{u} \quad (PC) \\
&\text{s.t. } (c_1,c_2)=\text{argmax}_{x,y}\big\{u(x)+\beta u(y)|x+y/\tilde{R}\leq T\big\} \quad (IC_1) \\
&\quad (\hat{c}_1,\hat{c}_2)=\text{argmax}_{x,y}\big\{u(x)+\hat{\beta} u(y)|x+y/\tilde{R}\leq T\big\} \quad (IC_2)
\end{aligned} \tag{8}$$

Proposition 6 characterizes the optimal contract when investment contracts must be linear because of the presence of a secondary market.

Proposition 6. *When financial intermediaries cannot distinguish between sophisticated agents and naïve agents in the economy, in a competitive equilibrium with a secondary market to trade deposit contracts at date 1, the equilibrium linear scheme $(T,\ \tilde{R})$ satisfies the following conditions:*

$$\tilde{R}=\frac{u'(c_1)}{\beta u'(c_2)}=\frac{u'(\hat{c}_1)}{\hat{\beta} u'(\hat{c}_2)}$$

$$T=c_1+\frac{c_2}{\tilde{R}}=\hat{c}_1+\frac{\hat{c}_2}{\tilde{R}}$$

where (c_1,c_2) and $(\hat{c}_1,\hat{c}_2)$ represent the realized consumptions for naïve and sophisticated agents, respectively. If only sophisticated agents exist in the economy, the first best outcome is achieved and the competitive equilibrium contract is characterized by $\tilde{R}=\frac{R}{\beta}$ and $T=c_1^{fb}+\frac{\hat{\beta}c_2^{fb}}{R}$.

Proof. See Appendix.[11] □

It is optimal to offer a sophisticated agent a contract with an interest rate of $\tilde{R}=\frac{R}{\beta}$ that aligns self 1's interest with long-term welfare. This contract counteracts self 1's desire for immediate gratification and maximizes sophisticated agents' welfare. This result implies that both restricted linear contracts and unrestricted contracts yield the same equilibrium outcome when all agents are sophisticated.

We have demonstrated that sophisticated agents can achieve the first best consumption through intermediated investment when all agents are sophisticated. The results are complicated when a percentage of agents are naïve. However, restricted linear contracts prevent financial intermediaries from setting early withdrawal penalties too high; therefore, a slightly naïve agent only fails to accurately predict his/her future behavior by a small amount and can achieve nearly first best welfare.

Corollary 3. *If financial intermediaries cannot distinguish between sophisticated and naïve agents in the economy, when β is close to $\hat{\beta}$, in a competitive equilibrium with restricted linear*

[11] An analytical characterization does not exist for the equilibrium when only naive agents exist. In such cases, the first best outcome cannot be achieved.

contracts, naïve agents' date-0 utility is higher than for the autarky economy if the liquidation cost is small (and higher than the utility in the unrestricted contract case according to Corollary 2); however, a sophisticated agents' date-0 utility is higher than the first best level and all agents' realized date-0 utility increases with the proportion of naïve agents in the economy, λ. However, both sophisticated and naïve agents prefer competitive equilibrium outcomes with unrestricted contracts that offer agents fewer options.

Proof. See Appendix. □

Corollary 3 indicates that at date 0, all agents strictly prefer unrestricted contracting to restricted linear contracting. Sophisticated agents are worse off when the linear restriction is imposed because linear contracts eliminate discontinuity in welfare and reduce the cross-subsidy from naïve agents to sophisticated agents. Notably, naïve agents prefer unrestricted contracting to restricted linear contracting although their welfare is improved with the linear restriction; naïve agents would oppose opening a secondary market or being offered more options for early withdrawal. This result is in contrast with the study by Gottlieb and Smetters (2014) in which consumers in the insurance market are not affected by liquidity shock ("background risk") and insurers strongly oppose secondary markets that could improve policyholders' welfare.

If both sophisticated and naïve agents exist, restricting contracts to a linear structure does not Pareto-dominate unrestricted nonlinear contracts. Nevertheless, if naïve agents are sufficiently sophisticated, the benefit of this restriction for naïve agents outweighs the harm to sophisticated agents. Because this intervention decreases the distortion in repayment schedules for both types of investors, it may increase the population-weighted sum of welfare.

4. Term premium for restricted linear contracts

The restricted linear contract implies a term structure of interest rates in our three-date model. If the one-period and two-period interest rates are i_1 and i_2, respectively, then the repayment options under this term structure include all (c_1, c_2) that satisfy:

$$\frac{c_1}{1+i_1} + \frac{c_2}{(1+i_2)^2} = 1 \tag{9}$$

A linear contract $c_1 + \frac{c_2}{\tilde{R}} = T$ assumes the term structure as follows:

$$\begin{cases} 1+i_1 = T \\ 1+i_2 = \sqrt{\tilde{R}T} \end{cases} \tag{10}$$

First, we note that according to Proposition 6, when all agents are sophisticated, we obtain $1 + i_1 = c_1^{fb} + \frac{\hat{\beta} c_2^{fb}}{R}$ and $(1+i_2)^2 = \frac{c_1^{fb} R}{\hat{\beta}} + c_2^{fb}$; therefore, when agents become less patient ($\hat{\beta}$ is smaller), as the first best consumptions are independent of $\hat{\beta}$, financial intermediaries would offer a term structure with a steeper slope (a lower one-period interest rate and a higher two-period interest rate) to persuade agents to commit to the first best allocation.

If both sophisticated and naïve agents exist in the economy, financial intermediaries make profits from naïve agents to subsidize the losses from sophisticated agents. Compared to an economy without a secondary market, a secondary market increases the value of liquidated assets at date 1 and each naïve agent provides fewer subsidies. When more naïve agents exist in the economy, there are more liquidation needs at date 1 and the secondary market price of the consumption at date 2 is less than date 1, which is equivalent to a steeper term structure that is implied by a linear contract.

When contracts are restricted to a linear form, naïve agents will liquidate according to the linear scheme. A steeper term structure results in a decreased liquidation value for the date 2 consumption and each naïve agent pays more for each unit of liquidation. When the liquidation is more costly, each naïve agent would liquidate less of his/her date 2 consumption; however, the subsidy to sophisticated agents is higher for each unit of date 2 consumption that is liquidated by a naïve agent. Therefore, when more naïve agents exist in the economy, a steeper term structure yields a greater total amount of subsidy and a higher perceived utility (which is also the realized utility for sophisticated agents). This may not be feasible with a flatter term structure because naïve agents will liquidate more at date 1 because of a lower cost for liquidation, a financial intermediary must reduce its initial investment at date 0 and in doing so, may not be able to afford a high payment to sophisticated agents. This indicates that a term premium for naïveté, which increases with the proportion of naïve agents. The following proposition confirms the above intuition.

Proposition 7. *If financial intermediaries cannot distinguish between sophisticated and naïve agents in the economy, when $\hat{\beta}$ is close to 1 in a competitive equilibrium, the linear contract is featured with a term structure (i_1, i_2) where i_1, i_2 represent one-period and two-period interest rates with $i_1 < 0$, $i_2 > 0$ and the term premium $i_2 - i_1$ increases in λ.*

Proof. See Appendix. □

To understand the condition on $\hat{\beta}$ (close to 1) that is required for the results to hold, we note that in addition to the cross-subsidy effect associated with a steeper term structure discussed above, there is an intertemporal consumption allocation effect, which is the distortion on the perceived consumptions at date 0 because of the steeper term structure. When $\hat{\beta}$ is close to 1, this intertemporal consumption allocation effect is negligible. When $\hat{\beta}$ is small (close to β), the intertemporal consumption allocation effect becomes substantial. This suggests that the slope of the term structure may not increase with the proportion of naïve agents because a very steep term structure results in biased and undesirable intertemporal (perceived) consumption allocation. Therefore, when more naïve agents exist in the economy and there is little perceived time inconsistency (large $\hat{\beta}$), a steeper term structure yields a higher perceived utility for all agents and is the outcome of competitive equilibrium.

Notably, the long-term interest rate is higher than the short-term interest rate, which is consistent with the typical upward sloping shape of the yield curve, but the negative nominal short-term interest rate is seldom empirically observed. However, the interest rates are in real terms in our model. The negative real short-term interest rate is due to the assumptions of zero short-term investment returns and no time discounting between date 0 and date 1.

To examine the robustness of the conclusion of Proposition 7, we extend the model to a four-date (three-period) model. Similar to the basic model, each agent is endowed with one unit of good at date 0 and the good is consumed at dates 1, 2 and 3. The good can be stored from one date to the next or invested at date 0 in a long-run technology that returns $R > 1$ units at date 3 and early liquidation at date 1 or 2 has a cost and only returns L units.

The restricted linear contract in this case also implies a term structure of interest rates. Denote the k-period interest rate as i_k, the repayment options for this term structure include all positive (c_1, c_2, c_3) that satisfy $\frac{c_1}{1+i_1} + \frac{c_2}{(1+i_2)^2} + \frac{c_3}{(1+i_3)^3} = 1$.

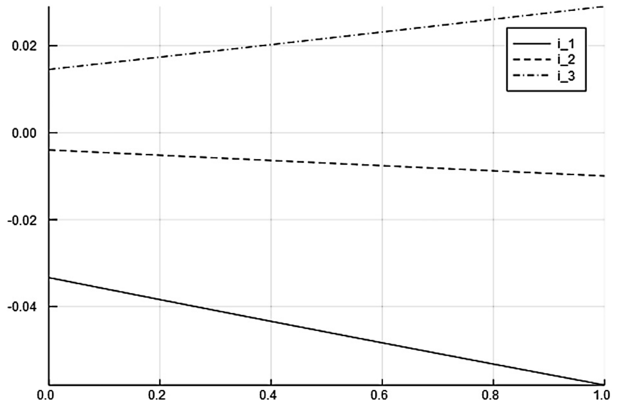

Fig. 1. Term structure and the proportion of naïve agents.

Similar to the two-period game, when all agents are sophisticated, financial intermediaries offer lower interest rates for the first two periods and a higher three-period interest rate to help investors commit to the first best allocation.

Lemma 2. *In a three-period model, if all agents are sophisticated in competitive equilibrium, a linear contract is featured with a term structure* (i_1, i_2, i_3) *where* i_1, i_2, *and* i_3 *are one-period, two-period and three-period interest rates with* $i_1 < 0$, $i_2 < 0$, *and* $i_3 > 0$, *respectively. Specifically, the interest rates should satisfy:*

$$1 + i_1 = c_1^{fb}\left(1 + \frac{1+i_1}{(1+i_2)^2}\right) + \left(1 - 2c_1^{fb}\right)\hat{\beta}\frac{1+i_1}{(1+i_2)^2} < 1$$

$$(1 + i_1)^2 = c_2^{fb}\left(1 + \frac{(1+i_2)^2}{1+i_1}\right) + \left(1 - 2c_2^{fb}\right)\hat{\beta} < 1$$

$$(1 + i_3)^3 = \frac{R}{2\hat{\beta}}\left(1 + \frac{(1+i_2)^2}{1+i_1}\right)\left(1 - \frac{c_3^{fb}}{R}\right) + c_3^{fb} > 1$$

Proof. See Appendix. □

The above lemma indicates that interest rates for one-period and two-period are negative and the three-period interest rate is positive when all agents are sophisticated.

When some agents are naïve, if the interest rates are initially set as if all agents are sophisticated (and financial intermediaries makes zero profits), financial intermediaries has an incentive to lower the interest rates for the first two periods and increase the three-period interest rate. This makes the consumptions more backloaded to seduce naïve agents, who subsidize sophisticated agents. Therefore, more naïve agents would lead to lower interest rates for the first two periods and a higher three-period interest rate. The following numerical exercise confirms this intuition and demonstrates that the term premium increases with the proportion of naïve agents.

We assume $R = 1.1$ and $u(c) = \frac{1-\gamma}{\gamma}(\frac{ac}{1-\gamma} + b)^\gamma - \frac{1-\gamma}{\gamma}b^\gamma$, where $a = 1, b = 0.1, \gamma = -2$, so that $u(0) = 0$ and $u'(0) = 1000$. The hyperbolic discount rate β and $\hat{\beta}$ of agents are set to be $\beta = 0.9$ and $\hat{\beta} = 0.95$.

Consistent with our previous results, Fig. 1 indicates that for the parameter value in the numerical example, one-period and two-period interest rates are both negative and decreasing relative to the proportion of naive agents; only the three-period interest rate is positive and increasing relative to the proportion of naive agents.

5. Solvency requirement of financial intermediaries

Although financial intermediaries are solvent at date 1 and date 2, we do not require financial intermediaries to be solvent at date 0 if no agents prematurely withdraw and all agents adhere to the "decoy" contract i.e., $\hat{c}_1 + \hat{c}_2/R \leq 1$. In this section we consider a case in which financial intermediaries must disclose the amount of investment at date 0 and satisfy all agents' needs according to the "decoy" contract. For example, the equilibrium unrestricted nonlinear contract for naïve agents characterized in Proposition 3 is not feasible with a solvency requirement because the expected consumption schedule does not align with financial intermediaries' budget and although it would not be realized in equilibrium, financial intermediaries cannot remain solvent ex-ante if all agents remain with the expected "decoy" consumption option.

5.1. Equilibrium with exogenous capital

We reconsider financial intermediaries' problem when all agents are naïve and solvency requirement is imposed at date 0. Financial intermediaries are required to own initial capital K to satisfy the budget constraint of the expected "decoy" consumption option. We focus on linear contracts although the primary results remain valid for unrestricted nonlinear contracts.

We first consider a case with exogenous capital K and we define the equilibrium with the solvency requirement and linear contract as follows.

Definition 4. A competitive equilibrium with restricted linear contracts and solvency requirement is a linear scheme $\{\tilde{R}, T\}$ that is offered by financial intermediaries and an agent has the right to choose any non-negative consumption schedule (c_1, c_2) that satisfies $c_1 + \frac{c_2}{\tilde{R}} = T$. The equilibrium has the following properties:

1. Solvency requirement: At date 0, all agents choose $(\hat{c}_1, \hat{c}_2)$ to optimize $u(\hat{c}_1) + \hat{\beta}u(\hat{c}_2)$ subject to $\hat{c}_1 + \frac{\hat{c}_2}{\tilde{R}} = T$; at date 1 naïve agents switch to (c_1, c_2) to optimize $u(c_1) + \beta u(c_2)$ subject to $c_1 + \frac{c_2}{\tilde{R}} = T$. Both (c_1, c_2) and $(\hat{c}_1, \hat{c}_2)$ satisfy the solvency requirement, i.e., $c_1 + c_2/R \leq 1 + K$ and $\hat{c}_1 + \frac{\hat{c}_2}{R} \leq 1 + K$.[12]
2. No deviation: there exists no other linear scheme $\{\tilde{R}', T'\}$ that is associated with the realized incentive compatible consumption choices, $(\hat{c}_1', \hat{c}_2')$ and (c_1', c_2'), for sophisticated and naïve agents, respectively; and $\{\tilde{R}', T'\}$ is strictly preferred by agents at date 0 and yields higher profits.

[12] It is natural to assume that only the bundles (c_1, c_2) and $(\hat{c}_1, \hat{c}_2)$ satisfy the solvency requirement. We do not require that all the bundles that an agent is allowed to choose from the linear scheme should satisfy the solvency requirement, i.e., we do not consider the solvency issue for the off-equilibrium choices. This is consistent with the fact that the regulator would only check what are shown on financial intermediaries' balance sheet.

With limited capital, the equilibrium is no longer the competitive equilibrium that is defined in Section 3, and specifically, financial intermediaries may make non-zero profits. We may interpret this as a special case of an imperfect competition equilibrium, which will be further discussed below with more general assumptions.

In equilibrium, financial intermediaries attempt to maximize their profits under the solvency constraint:

$$\begin{aligned} &\max_{T,\tilde{R}} \lambda\big[R(1-c_1)-c_2\big]+(1-\lambda)\big[R(1-\hat{c}_1)-\hat{c}_2\big] \\ &\quad u(\hat{c}_1)+u(\hat{c}_2)\geq \underline{u} \quad (PC) \\ &\text{s.t. } (c_1,c_2)=\text{argmax}_{x,y}\big\{u(x)+\beta u(y)|x+y/\tilde{R}\leq T\big\} \quad (IC_1) \\ &\quad (\hat{c}_1,\hat{c}_2)=\text{argmax}_{x,y}\big\{u(x)+\hat{\beta} u(y)|x+y/\tilde{R}\leq T\big\} \quad (IC_2) \\ &\quad \max\{c_1+c_2/R,\hat{c}_1+\hat{c}_2/R\}\leq 1+K \quad (SC) \end{aligned} \tag{11}$$

The last condition is the "solvency condition", which requires that both of the repayment options lie within the budget constraint. Because of the cross-subsidy effect, the SC condition can be further simplified as $\hat{c}_1+\hat{c}_2/R\leq 1+K$. If K is large enough, the SC condition is always satisfied and the equilibrium is not affected by the solvency requirement. However, if K is small, the SC is always binding. In this case, financial intermediaries may receive positive profits and there is no deviation that can earn higher profits because ex-ante naïve agents are only concerned with the "decoy" repayment option; however, the "decoy" repayment option is constrained by the solvency condition. The regulation that requires financial intermediaries to remain solvent hurts naïve agents by restricting financial intermediaries' repayment options.

5.2. *Equilibrium with endogenous capital and homogeneous cost of capital*

We assume that there are a continuum of competitive financial intermediaries with measure 1, and the total cost of capital for each financial intermediary is $C(K)=\gamma K$, where $\gamma>0$ is the unit cost of capital. We also assume γ is small so that the intermediary will prefer a positive capital. The assumption of a constant unit cost of capital simplifies our analysis. A financial intermediary is solving the following problem:

$$\begin{aligned} &\max_{K,T,\widetilde{R}} \lambda\big[R(1-c_1)-c_2\big]+(1-\lambda)\big[R(1-\hat{c}_1)-\hat{c}_2\big]-\gamma K \\ &\text{s.t. } u(\hat{c}_1)+u(\hat{c}_2)\geq \underline{u} \quad (PC) \\ &\quad (c_1,c_2)=\text{argmax}_{x,y}\big\{u(x)+\beta u(y)\mid x+y/\widetilde{R}\leq T\big\} \quad (IC_1) \\ &\quad (\hat{c}_1,\hat{c}_2)=\text{argmax}_{x,y}\big\{u(x)+\hat{\beta} u(y)\mid x+y/\widetilde{R}\leq T\big\} \quad (IC_2) \\ &\quad \max\left\{c_1+\frac{c_2}{R},\hat{c}_1+\frac{\hat{c}_2}{R}\right\}\leq 1+K \quad (SC) \end{aligned} \tag{12}$$

Clearly, SC is binding because otherwise a financial intermediary can always reduce capital to tighten the constraint. In equilibrium, capital K is positive because the decoy repayment option is beyond the budget constraint. Competition among financial intermediaries eliminates the possibility of strictly positive profits, and all agents are worse off than in competitive equilibrium without solvency requirement due to the additional cost of capital. We summarize these results in Proposition 8.

Proposition 8. *If financial intermediaries cannot distinguish between sophisticated and naïve agents in the economy and there is a solvency requirement, there exists a competitive equilibrium in which financial intermediaries' gain from the deposit will exactly cover the cost of the capital. Both naïve and sophisticated agents receive lower date-0 utility than in the case without the solvency requirement and their date-0 utility is lower when the unit cost of capital is higher.*

Proof. See Appendix. □

Financial intermediaries compete to increase capital and contract payments until they make zero profit. An economy without a solvency requirement is equivalent to a situation where financial intermediaries can acquire capital at zero cost; therefore, when capital acquirement involves a strict positive unit cost, the total resources allocated to agents decrease and the social welfare measured by self 0' utility decreases.

5.3. Equilibrium with endogenous capital and heterogeneous cost of capital

Next, we consider a case where the intermediaries have different costs of capital. For simplicity, assume two financial intermediaries exist, whose costs of capital are $C_1(K) = \gamma_1 K$ and $C_2(K) = \gamma_2 K$, respectively. We assume that $0 < \gamma_1 < \gamma_2$ and γ_2 is small.

We assume a variation of Hotelling-type imperfect competition exists among financial intermediaries. Specifically, there is a linear city of length 1. Agents are uniformly distributed with density 1 along this line. Two financial intermediaries are located at the extremes of the city: financial intermediary 1 is at $x = 0$ and financial intermediary 2 is at $x = 1$. Agents incur a transportation cost t (in utility term) per unit of length. Thus, if financial intermediary 1 offers a contract that yields $\underline{u}_1$ units of self-0 utility for agents and financial intermediary 2 offers a contract that yields $\underline{u}_2$ units of self-0 utility for agents, then the agents in the interval $[0, x_1]$ will go to financial intermediary 1 and the agents in $[x_2, 1]$ will go to financial intermediary 2, where $x_1 = x_2 = (\underline{u}_1 - \underline{u}_2 + t)/2t$ if t is relatively small and $x_1 = (\underline{u}_1 - \underline{u}_0)/t$, $x_2 = (\underline{u}_0 - \underline{u}_2 + t)/t$ if t is large ($\underline{u}_0$ is the date-0 utility in the autarky case without financial intermediary).[13]

We will focus on the case with a reasonable small transportation cost such that financial intermediaries are competing for all agents in the economy, while the case with large transportation cost t or sufficiently small t is close to the monopoly case. When t is large, each intermediary only attracts the agents nearby and some agents will not use the service of either financial intermediary. When t is very small, the intermediary with lower cost capital will get 100% market share.[14] In both cases there will be no competition essentially.

In the economy with a reasonable small transportation cost, given $(\underline{u}_1, \underline{u}_2)$, the market share of intermediary i is $w_i = (\underline{u}_i - \underline{u}_j + t)/2t$. In a Nash equilibrium, financial intermediary i chooses $(K_i, T_i, \widetilde{R}_i)$ so as to maximize its profit given the strategy $(K_j, T_j, \widetilde{R}_j)$ of its rival:

[13] We only consider the case that the value of $(\underline{u}_1, \underline{u}_2)$ satisfies $\underline{u}_1 > \underline{u}_0$, $\underline{u}_2 > \underline{u}_0$ and $|\underline{u}_1 - \underline{u}_2| \leq t$. It is easy to see that in equilibrium $|\underline{u}_1 - \underline{u}_2| \leq t$; otherwise the intermediary that offers higher utility would make more profits by offering a contract with lower utility.

[14] This will not be the case if we have a convex cost of capital $C(K)$ that satisfies $C(0) = 0$, $C'(0) = 0$, $C'(K) > 0$, $C''(K) > 0$, in which case the low cost intermediary will never be able to get 100% market share as the capital become increasing costly when it attract more clients.

$$\begin{aligned}
&\max_{K_i,T_i,\widetilde{R}_i} w_i\left\{\lambda\left[R(1-c_{1i})-c_{2i}\right]+(1-\lambda)\left[R(1-\hat{c}_{1i})-\hat{c}_{2i}\right]\right\}-\gamma_i K_i \\
&\text{s.t. } u(\hat{c}_{1i})+u(\hat{c}_{2i})\geq \underline{u}_0+tw_i \quad (PC) \\
&\quad (c_{1i},c_{2i})=\text{argmax}_{x,y}\left\{u(x)+\beta u(y)\mid x+y/\widetilde{R}\leq T\right\} \quad (IC_1) \\
&\quad (\hat{c}_{1i},\hat{c}_{2i})=\text{argmax}_{x,y}\left\{u(x)+\hat{\beta} u(y)\mid x+y/\widetilde{R}\leq T\right\} \quad (IC_2) \\
&\quad \max\left\{w_i\left(c_{1i}+\frac{c_{2i}}{R}\right), w_i\left(\hat{c}_{1i}+\frac{\hat{c}_{2i}}{R}\right)\right\}\leq w_i+K \quad (SC)
\end{aligned} \tag{13}$$

where $w_i = (\underline{u}_i - \underline{u}_j + t)/2t$.

The following proposition summarizes some characteristics of this equilibrium.

Proposition 9. *In a reasonable low-transportation-cost Hotelling type economy with two financial intermediaries competing with each other subject to a solvency requirement, if the financial intermediaries cannot distinguish between sophisticated and naïve agents, a Nash equilibrium with endogenous capital and heterogeneous cost of capital has the following features:*

(i) *Each financial intermediary makes strictly positive profit.*
(ii) *The agents who deposit in the financial intermediary with a lower cost of capital have higher welfare measured by self 0's utility, and the financial intermediary with a lower cost of capital produces more capital, makes more profits and obtains a larger market share.*

Proof. See Appendix. □

6. Conclusion

Investors are not entirely rational when making investment decisions and financial intermediaries can help these investors to overcome their behavioral bias. In this study, we explore a case when investors have time inconsistent preferences, which are well documented in prior studies. Our model offers a new explanation of early withdrawal penalties in the investment contracts of financial intermediaries when investors are time inconsistent. In addition, our study identifies simple welfare-improving interventions. For example, a secondary market that allows long-term deposit contracts to be traded could improve naïve agents' welfare. However, although these interventions increase social welfare, they may not be accepted by agents, who believe that they were making rational decisions. Therefore, it is uncertain whether there exist modifications of this intervention that the agents would prefer.

For simplification, we consider a model with three dates for the majority of our analyses. An interesting direction for future research would be to extend this model to an infinite horizon with overlapping generations and consider the term structure of interest rates. One model prediction proposes that the term premium should be higher in an economy that has more naïve agents, which may be empirically tested by comparing the term premiums in countries with different education levels or by comparing term premiums in developing and developed economies. Future research may consider these issues.

To isolate the commitment problem of time-inconsistent consumers, we assume that liquidity needs are entirely generated by irrational time-inconsistent preferences and all agents differ only in their beliefs (or degree of sophistication). Future studies could consider a more general model that combines time inconsistency and real liquidity shocks, as in Amador et al. (2006), to provide a more complete characterization of financial intermediaries' role in providing risk sharing and commitment in a competitive equilibrium. Future studies could also investigate whether bank runs would occur in equilibrium in the generalized model.

Appendix A

Proof for Lemma 1. The Lagrange condition for (3) implies $\hat{x} = 0$ or $\hat{x} > 0$ with $\hat{x}' = \frac{u''(c_1)L+\hat{\beta}R^2u''(c_2)}{u''(c_1)L^2+\hat{\beta}R^2u''(c_2)}$. We know from (2) that $\frac{d(u(c_1)+u(c_2))}{dI} = u'(c_1)(L\hat{x}'-1)+u'(c_2)R(1-\hat{x}')$, and a positive $\hat{x}$ always leads to $\frac{d(u(c_1)+u(c_2))}{dI} < 0$, which cannot be true. Thus, a strictly positive $\hat{x}$ cannot be optimal. □

Proof for Proposition 1. Substituting $\hat{x} = 0$, the first best investment can be derived from the first order condition for (2), which suggests that $\frac{u'(c_1)}{u'(c_2)} = R$. At the same time, $\hat{x} = 0$ requires that the first order condition in (3) satisfies that $\frac{u'(c_1)}{u'(c_2)} \le \frac{\hat{\beta}R}{L}$, which is satisfied at the first best investment when $L \le \hat{\beta}$. When $L > \hat{\beta}$, the first best investment is not optimal in the autarky case, and we have $\frac{u'(c_1)}{u'(c_2)} = \frac{u'(1-I)}{u'(IR)} = \frac{\hat{\beta}R}{L}$. For the liquidation decision at date 1, if an agent is naïve with realized β smaller than $\hat{\beta}$, the first order conditions for (2) and (4) imply that $x = 0$ only if $\frac{u'(c_1)}{u'(c_2)} = \frac{u'(1-I)}{u'(IR)} = R \le \frac{\beta R}{L}$, which is true when $L \le \beta$. □

Proof of Proposition 2. At date 1, the price of consumption at date 2 must be between L/R and 1. If the price of date 2 consumption is smaller than L/R, an agent will prefer to liquidate his/her own investment instead of selling it; if the price of date 2 consumption is bigger than 1, an agent will prefer to sell all his/her investment and store the proceeds. At date 1, for all agents, before they trade, their consumption plan is $(\hat{c}_1^{at}, \hat{c}_2^{at})$, the autarky consumption plan calculated in Section 2.2. Given date 2 consumption price, p, a naïve agent is choosing trading amount Δ_n and liquidation amount L_n to solve the following optimization problem:

$$\max_{\Delta_n, L_n} u\big(\hat{c}_1^{at} + p\Delta_n + L_n L\big) + \beta u\big(\hat{c}_2^{at} - \Delta_n - RL_n\big),$$

and a sophisticated agent is choosing trading amount Δ_s and liquidation amount L_s to solve the following optimization problem:

$$\max_{\Delta_s, L_s} u\big(\hat{c}_1^{at} - p\Delta_s + L_s L\big) + \hat{\beta} u\big(\hat{c}_2^{at} + \Delta_s - RL_s\big).$$

We will only look at the equilibrium in which sophisticated agents buy the future consumption from naïve agents, who have the liquidity need, that is, $\Delta_n \ge 0$ and $\Delta_s \ge 0$. It is easy to check the other type of equilibrium does not exist.

For a sophisticated agent, it is easy to check for any $p \in [L/R, 1]$, he/she will choose $L_s = 0$. Taking the first order condition with respect to Δ_s gives us:

$$-pu'\big(\hat{c}_1^{at} - p\Delta_s\big) + \hat{\beta}u'\big(\hat{c}_2^{at} + \Delta_s\big) = 0,$$

which implies: $\frac{\hat{\beta}}{p} = \frac{u'(\hat{c}_1^{at} - p\Delta_s)}{u'(\hat{c}_2^{at} + \Delta_s)} > \frac{u'(\hat{c}_1^{at})}{u'(\hat{c}_2^{at})} = \frac{\hat{\beta}R}{L}$, or, equivalently, $p < L/R$, a contradiction to $p \in [L/R, 1]$ if $L \ge \hat{\beta}$. Therefore, in equilibrium, we have $\Delta_s = 0$, which leads to $\Delta_n = 0$ due to market clearance condition.

When $L \le \beta$, all agents are of the first best and hence have no incentive to trade.

When $\beta < L < \hat{\beta}$, naïve agents also choose $L_n = 0$ due to the high selling price. Thus the first order condition suggests $\frac{\beta}{p} = \frac{u'(\hat{c}_1 + p\Delta_n)}{u'(\hat{c}_2 - \Delta_n)}$. Therefore, it is possible to find a price $p \in [\frac{L}{R}, 1]$ to clear the market and no agents liquidate their assets. □

Proof of Proposition 3. For sophisticated agents, with the Lagrange multiplier, μ, the first order conditions for (5) are:

$$-R + \mu u'(c_1) = 0$$
$$-1 + \mu u'(c_2) = 0$$

Using the zero-profit condition and applying $I = 1 - c_1$, we can see they are equivalent to the first order conditions for the first best optimization defined in (1). Therefore, the solution gives the first best consumption (c_1^{fb}, c_2^{fb}).

For naïve agents, we first show that PC and IC in (6) must bind in equilibrium, and PCC in (6) is equivalent to $\hat{c}_1 \le c_1$ or $\hat{c}_2 \ge c_2$ with $\beta < \hat{\beta}$. If PC is not binding, then financial intermediaries can lower c_1 and $\hat{c}_1$ to increase the profit. If IC is not binding, then financial intermediaries could increase profits by lowering c_1. With IC binding, it is easy to check, when $\beta < \hat{\beta}$, PCC in (6) is equivalent to $\hat{c}_1 \le c_1$. We first derive the optimal solution for (6) without PCC and confirm that $\hat{c}_1 \le c_1$ is satisfied. The relaxed problem is:

$$\max_{c_1, c_2, \hat{c}_1, \hat{c}_2} R(1 - c_1) - c_2$$
$$u(\hat{c}_1) + u(\hat{c}_2) \ge \underline{u} \quad (PC)$$
$$\text{s.t. } u(c_1) + \beta u(c_2) \ge u(\hat{c}_1) + \beta u(\hat{c}_2) \quad (IC)$$

The optimal solution must have $\hat{c}_1 = 0$, otherwise since $\beta < 1$, financial intermediaries could decrease $u(\hat{c}_1)$ and increase $u(\hat{c}_2)$ by the same amount, keeping PC unchanged and relaxing the IC constraint, thus allowing them to lower c_1.

Using $\hat{c}_1 = 0$, we combine PC and IC constraints into:

$$u(c_1) + \beta u(c_2) = u(0) + \beta\big(\underline{u} - u(0)\big).$$

So the first order condition with respect to c_1 and c_2 gives $\frac{u'(c_1)}{u'(c_2)} = \beta R$. Since $u(c_1) + \beta u(c_2) = u(\hat{c}_1) + \beta u(\hat{c}_2)$, and $\hat{c}_1 = 0$, we have $\hat{c}_1 \le c_1$ and $\hat{c}_2 \ge c_2$, thus PCC holds. □

Proof of Corollary 1. The welfare result for sophisticated agents is the immediate implication of Proposition 3. For naïve agents, it is easy to check that naïve agents' consumptions (c_1, c_2) are the same with the autarky consumptions characterized in Section 2 when liquidation cost is 0, that is $L = 1$. When $L \le \beta$, the autarky consumption is identical to the first best consumption, and hence is better than the consumption with financial intermediary. When $1 > L > \beta$, the liquidation x solves $\frac{u'(1-I+xL)}{u'(IR-xR)} = \frac{\beta R}{L}$, thus $\frac{\partial x}{\partial L} = (u''(1 - I + xL)x + \frac{\beta R}{L^2} u'(IR - xR))/(-Lu''(1 - I + xL) - \frac{\beta R^2}{L} u''(IR - xR))$. When the relative risk aversion is sufficiently large, we always have $\frac{\partial x}{\partial L} < 0$. Therefore, $\frac{\partial (u(1-I+xL)+u(IR-xR))}{\partial L} = u'(1 - I + xL)(x + x'L) + u'(IR - xR)(-x'R) = u'(1 - I + xL)x + x'(u'(1 - I + xL)L - u'(IR - xR)R) > 0$, suggesting in the autarchy economy a lower L leads to a lower date-0 utility. Since the consumption plan with financial intermediary is identical to the autarky case with $L = 1$, which implies that the consumption plan with financial intermediary is better than the one in autarky economy for naive agents for $L < 1$. □

Proof of Proposition 4. The first condition is the zero-profit condition. It is easy to see that PC and IC_1 bind. First we ignore IC_2, and the first order conditions for (7) are:

$$-\lambda R+\mu_2 u'(c_1)=0$$
$$-\lambda+\mu_2\beta u'(c_2)=0$$
$$-(1-\lambda)R+\mu_1 u'(\hat{c}_1)-\mu_2 u'(\hat{c}_1)=0$$
$$-(1-\lambda)R+\mu_1 u'(\hat{c}_2)-\mu_2\beta u'(\hat{c}_2)=0$$

Eliminating μ_1 and μ_2, we can get:

$$u'(c_1)=\beta R u'(c_2)$$
$$\frac{u'(\hat{c}_1)}{u'(\hat{c}_2)}=\left(1+\frac{\lambda(1-\beta)u'(\hat{c}_1)}{(1-\lambda)u'(c_1)}\right)R$$

Finally, since $u(c_1)+\beta u(c_2)=u(\hat{c}_1)+\beta u(\hat{c}_2)$, and $\frac{u'(c_1)}{u'(c_2)}<\frac{u'(\hat{c}_1)}{u'(\hat{c}_2)}$, we have $c_2<\hat{c}_2$, thus verify that IC_2 holds. □

Proof of Corollary 2. We can write the profits of financial intermediaries from naïve agents and sophisticated ones as follows:

$$\pi_1=(1-c_1)R-c_2$$
$$\pi_2=(1-\hat{c}_1)R-\hat{c}_2$$

which implies:

$$\pi_1-\pi_2=(\hat{c}_1-c_1)R+\hat{c}_2-c_2$$

According to Mean Value Theorem to $u(c_1)+\beta u(c_2)=u(\hat{c}_1)+\beta u(\hat{c}_2)$, there exist $\hat{c}_1<c_1^\star<c_1$ and $c_2<c_2^\star<\hat{c}_2$, such that:

$$u'\left(c_1^\star\right)(\hat{c}_1-c_1)+\beta u'\left(c_2^\star\right)(\hat{c}_2-c_2)=0.$$

At the same time, we know $\beta R=\frac{u'(c_1)}{u'(c_2)}<\frac{u'(c_1^*)}{u'(c_2^*)}$ and $\hat{c}_1-c_1$ is negative, which implies:

$$0=u'\left(c_1^\star\right)(\hat{c}_1-c_1)+\beta u'\left(c_2^\star\right)(\hat{c}_2-c_2)<\beta u'\left(c_2^\star\right)\left[(\hat{c}_1-c_1)R+(\hat{c}_2-c_2)\right].$$

Therefore, we have $\pi_1-\pi_2>0$, which implies $\pi_1>0$ and $\pi_2<0$, that is, the financial intermediaries make positive profits from naïve agents, but incur losses by serving sophisticated agents.

In terms of welfare, for naïve agents, when the liquidation cost is zero, that is, $L=1$, the first order conditions under the unrestricted contract case for the competitive equilibrium with financial intermediaries and the autarky case always satisfy $u'(c_1)=\beta R u'(c_2)$. In the autarky case, we have $c_1R+c_2=R$, and in the unrestricted contract case for the competitive equilibrium with financial intermediaries, we have $c_1R+c_2<R$, thus the realized date-0 utility is lower in the unrestricted contract case. By continuity, naïve agents' realized date-0 utility will be lower than the autarky case if liquidation cost is sufficiently small. Noticing that when liquidation cost is sufficiently large, the realized date-0 utility in the autarky case is identical to the first best case, naïve agents' realized date-0 utility in the unrestricted contract case will be lower when liquidation cost is sufficiently large.

Let $\underline{u}(\lambda)$ be the perceived outside option in the competitive equilibrium when the proportion of naïve agents is λ, and this is also sophisticated agents' actual welfare. For $\lambda' > \lambda$, assume $\underline{u}(\lambda') \leq \underline{u}(\lambda)$. Since financial intermediaries make profits on naïve agents, if the proportion increases to λ', the equilibrium contract associated with λ is still feasible with all the participation and incentive constraints satisfied, and makes strictly positive profits, which is a contradiction. Therefore we must have $\underline{u}(\lambda') > \underline{u}(\lambda)$. When $\lambda = 0$, we know sophisticated agents get first best outcome. Therefore, for any $\lambda > 0$, sophisticated agents get a higher date-0 utility than in the first best case. □

Proof of Proposition 5. First, note that at date 1 the existence of a secondary market for contracts is equivalent to the existence of secondary market for date-2 consumption.

Now suppose financial intermediaries only offers one consumption option (c_{10}, c_{20}) at date 0, and contracts offered by financial intermediaries can be traded on a secondary market at date 1, we prove there exists a date-1 price of date-2 consumption, p, that clears the market. At date 1, an agents of type i maximizes $u(c_{10} - px_i) + \beta_i u(c_{20} + x_i)$, where x_i is his/her purchase of date-2 consumption. The FOC of a naïve agent is:

$$\frac{u'(c_{10} - px_1)}{u'(c_{20} + x_1)} = \frac{\beta}{p}.$$

FOC of a sophisticated agent is:

$$\frac{u'(c_{10} - px_2)}{u'(c_{20} + x_2)} = \frac{\hat{\beta}}{p}.$$

Market clearing requires $\lambda x_1 + (1 - \lambda)x_2 = 0$, so there exists a price p within the range $(\frac{\beta u'(c_{20})}{u'(c_{10})}, \hat{\beta}\frac{u'(c_{20})}{u'(c_{10})})$ so that market clears. An agent chooses (c_1, c_2) to maximize his/her date-1 utility along the line: $pc_1 + c_2 = pc_{10} + c_{20}$. Therefore, if contracts offered by financial intermediaries can be traded in a secondary market at date 1, offering a consumption option (c_{10}, c_{20}) is equivalent to offering a linear contract that is linked to that consumption option. Moreover, to maximize profits, financial intermediaries would choose exactly the same linear contract as the equilibrium contract under linear restriction.

Next, we prove that the above results still hold when financial intermediaries offer more than one consumption options. Suppose financial intermediaries offer two consumption options (c'_{10}, c'_{20}) and (c''_{10}, c''_{20}). We only need to prove that there exists a unique equilibrium (c_{10}, c_{20}, p) in which every agents maximizes his/her date-1 utility along the line $pc_1 + c_2 = pc_{10} + c_{20}$. Then we can show such equilibrium is equivalent to the competitive equilibrium with linear contract. Without loss of generality, assume $c'_{10} < c''_{10}$ and define $p = -\frac{c'_{20} - c''_{20}}{c'_{10} - c''_{10}}$. By the proof above, there exists a price p' such that market clears at the consumption option (c'_{10}, c'_{20}). If $p' \leq p$, (c'_{10}, c'_{20}) dominates (c''_{10}, c''_{20}), and (c'_{10}, c'_{20}, p) is an equilibrium. Similarly, there exists a p'' such that market clears at (c''_{10}, c''_{20}), and $(c''_{10}, c''_{20}, p'')$ is an equilibrium if $p'' \geq p$.

Now we consider the case when $p' > p$ and $p'' < p$. By substitution effects, under price p, there is excessive supply of date 2 consumption at the consumption option (c'_{10}, c'_{20}), and there is excessive demand for date 2 consumption at the consumption option (c''_{10}, c''_{20}). So there exists an equilibrium (c_{10}, c_{20}, p), where (c_{10}, c_{20}) is a point on the line segment connecting (c'_{10}, c'_{20}) and (c''_{10}, c''_{20}), which means that in equilibrium, agents choose a mixture of (c'_{10}, c'_{20}) and (c''_{10}, c''_{20}). When financial intermediaries offer more than two repayment options, we can also prove in similar ways that such an equilibrium exists.

Finally, we show that if all other financial intermediaries offer the optimal linear contract, a financial intermediary will not deviate and provide another contract. Otherwise, the new contract must be more attractive for sophisticated agents, in other words, to subsidy sophisticated agents more. Then at date 1 naïve agents will liquidate their contracts in the secondary market, and hence the intermediary cannot profit from naïve agents and take a loss. □

Proof of Proposition 6. Since (IC_1) and (IC_2) in (8) are equivalent to $\frac{u'(c_1)}{\beta u'(c_2)} = \frac{u'(\hat{c}_1)}{\hat{\beta} u'(\hat{c}_2)} = \tilde{R}$ and $c_1\tilde{R} + c_2 = \hat{c}_1\tilde{R} + \hat{c}_2 = T\tilde{R}$. $(T, \tilde{R})$ can be recovered from (c_1, c_2) and $(\hat{c}_1, \hat{c}_2)$, which are the solution of the following optimization problem which is equivalent to (8):

$$\max_{c_1, c_2, \hat{c}_1, \hat{c}_2} u(\hat{c}_1) + u(\hat{c}_2)$$

$$\frac{u'(c_1)}{\beta u'(c_2)} = \frac{u'(\hat{c}_1)}{\hat{\beta} u'(\hat{c}_2)} = \tilde{R}$$

$$\text{s.t. } c_1\tilde{R} + c_2 = \hat{c}_1\tilde{R} + \hat{c}_2 = T\tilde{R}$$

$$\lambda\big[R(1-c_1) - c_2\big] + (1-\lambda)\big[R(1-\hat{c}_1) - \hat{c}_2\big] = 0.$$

When there are only sophisticated agents, let $\beta_1 = \beta_2 = \hat{\beta}$, $c_1 = \hat{c}_1$ and $c_2 = \hat{c}_2$, we must have $\frac{u'(\hat{c}_1)}{u'(\hat{c}_2)} = R$ by solving the above optimization problem. Setting $\tilde{R} = R/\hat{\beta}$ and $T = c_1^{fb} + c_2^{fb}/\tilde{R}$, and agents can achieve the first best results in the equilibrium. □

Proof of Corollary 3. First, we know that the perceived utility $\underline{u}$ in competitive equilibrium in unrestricted market is higher than that in restricted market because any feasible contract in restricted market is also feasible in unrestricted market.

To prove that restricted market raises welfare for sufficiently sophisticated consumers, we notice that, given $\hat{\beta}$ and λ, the optimal contract $(\tilde{R}, T)$ is continuous in β. Thus, as $\beta \to \hat{\beta}$, both types' consumption schedules approach the welfare-optimal one in long-term restricted market, that is, the first-best solutions. While in unrestricted market, the consumption schedule of naïve agents approaches to $u'(c_1)/u'(c_2) = \beta$, thus the allocation in unrestricted market is inferior.

Finally, when λ increases to $\lambda + \delta$, the contract for λ is always feasible. Hence the optimal contract for $\lambda + \delta$ should make sophisticated agents' date-0 utility higher, suggesting in the restricted linear market sophisticated agents' date-0 utility is increasing in λ. It also implies that the restricted linear market is better than the autarky economy and even the first best case for sophisticated agents. Since β is close to $\hat{\beta}$, naïve agents' consumptions in the restricted linear market are close to sophisticated agents' consumption, suggesting that the restricted linear market is better than the autarky economy for naïve agents. □

Proof of Proposition 7. From Proposition 6, we know that $u'(c_1)/\beta u'(c_2) = u'(\hat{c}_1)/\hat{\beta} u'(\hat{c}_2) = \tilde{R}$ and $T = c_1 + \frac{c_2}{\tilde{R}} = \hat{c}_1 + \frac{\hat{c}_2}{\tilde{R}}$, which imply:

$$\frac{\partial c_1}{\partial T} = \frac{\beta\tilde{R}u''(c_2)}{\beta\tilde{R}u''(c_2) + u''(c_1)}, \qquad \frac{\partial \hat{c}_1}{\partial T} = \frac{\hat{\beta}\tilde{R}u''(\hat{c}_2)}{\hat{\beta}\tilde{R}u''(\hat{c}_2) + u''(\hat{c}_1)}$$

$$\frac{\partial c_1}{\partial \tilde{R}} = \frac{\beta(u''(c_2)c_2 + u'(c_2))}{\beta\tilde{R}u''(c_2) + u''(c_1)}, \qquad \frac{\partial \hat{c}_1}{\partial \tilde{R}} = \frac{\hat{\beta}(u''(\hat{c}_2)\hat{c}_2 + u'(\hat{c}_2))}{\hat{\beta}\tilde{R}u''(\hat{c}_2) + u''(\hat{c}_1)}$$

Let financial intermediaries slightly increase $\tilde{R}$ and decrease T such that $\hat{c}_1$ decreases and $\hat{c}_2$ increases satisfying $u(\hat{c}_1)+u(\hat{c}_2)=u(\hat{c}_1-\hat{\delta}_1)+u(\hat{c}_2+\hat{\delta}_2)$. With $\hat{\beta}<1$, we must have $u(\hat{c}_1)+\hat{\beta}u(\hat{c}_2)>u(\hat{c}_1-\hat{\delta}_1)+\hat{\beta}u(\hat{c}_2+\hat{\delta}_2)$, suggesting $\hat{c}_1-\hat{\delta}_1+\frac{\hat{c}_2+\hat{\delta}_2}{\tilde{R}}<T$, otherwise agents would have chosen $(\hat{c}_1,\hat{c}_2)$ instead of $(\hat{c}_1-\hat{\delta}_1,\hat{c}_2+\hat{\delta}_2)$. Similarly, naïve agents' consumption satisfies the condition $c_1-\delta_1+\frac{c_2+\delta_2}{\tilde{R}}<T$. Therefore, we must have $\lambda(-\delta_1+\frac{\delta_2}{\tilde{R}})+(1-\lambda)(-\hat{\delta}_1+\frac{\hat{\delta}_2}{\tilde{R}})<0$. Thus we have the equilibrium $\tilde{R}$ is larger than R, otherwise, we would have $\lambda(-\delta_1+\frac{\delta_2}{R})+(1-\lambda)(-\hat{\delta}_1+\frac{\hat{\delta}_2}{R})<0$ and we can increase financial intermediaries' profit without lowering self 0's utility by increasing $\tilde{R}$.

To show $i_1<0$, $i_2>0$, we just need to show $T<1$. The zero-profit condition $\lambda(c_1R+c_2)+(1-\lambda)(\hat{c}_1R+\hat{c}_2)=R$ implies $R=(\lambda c_1+(1-\lambda)\hat{c}_1)(R-\tilde{R})+T\tilde{R}$. Hence $\tilde{R}>R$ implies $T<1$, that is, $i_1<0$, and $T\tilde{R}>R$, that is, $i_2>0$.

When λ increases, to maintain $\lambda(Rc_1+c_2)+(1-\lambda)(R\hat{c}_1+\hat{c}_2)=R$, we must have

$$\lambda\big(\Delta(Rc_1+c_2)\big)+(1-\lambda)\big(\Delta(R\hat{c}_1+\hat{c}_2)\big)=\Delta\lambda\big((R\hat{c}_1+\hat{c}_2)-(Rc_1+c_2)\big)>0.$$

Let $d_2=\frac{1}{T\tilde{R}}$. $d_1=\frac{1}{T}$, it is trivial that the linear contract can also be characterized by (d_1,d_2). Given the contract (d_1,d_2), the optimal consumption (c_1,c_2) satisfies $c_1d_1+c_2d_2=1$ and $\frac{u'(c_1)}{u'(c_2)}=\beta\frac{d_1}{d_2}$, $(\hat{c}_1,\hat{c}_2)$ satisfies $\hat{c}_1d_1+\hat{c}_2d_2=1$ and $\frac{u'(\hat{c}_1)}{u'(\hat{c}_2)}=\hat{\beta}\frac{d_1}{d_2}$. Simple algebra leads to $\frac{\partial c_1}{\partial d_1}=\frac{\beta u'(c_2)d_2-\beta u''(c_2)c_1d_1}{u''(c_1)d_2^2+\beta u''(c_2)d_1^2}<0$, $\frac{\partial c_1}{\partial d_2}=\frac{-u'(c_1)d_2-\beta u''(c_2)c_2d_1}{u''(c_1)d_2^2+\beta u''(c_2)d_1^2}$, $\frac{\partial c_2}{\partial d_1}=\frac{-\beta u'(c_2)d_1-u''(c_1)c_1d_2}{u''(c_1)d_2^2+\beta u''(c_2)d_1^2}$, and $\frac{\partial c_2}{\partial d_2}=\frac{u'(c_1)d_1-u''(c_1)c_2d_2}{u''(c_1)d_2^2+\beta u''(c_2)d_1^2}<0$.

Now assume the contract (d_1,d_2) changes to $(d_1+d_1\epsilon,d_2+d_2\epsilon+\eta)$ when λ increases to $\lambda+\Delta\lambda$. As $\Delta u(\hat{c}_1)+\Delta u(\hat{c}_2)=\frac{u'(\hat{c}_2)}{d_2}(\hat{\beta}d_1\Delta\hat{c}_1+d_2\Delta\hat{c}_2)$ and $\hat{\beta}d_1\Delta\hat{c}_1+d_2\Delta\hat{c}_2=\frac{u'(\hat{c}_1)d_1d_2(1-\hat{\beta})\eta-(u''(\hat{c}_1)d_2^2+\hat{\beta}^2u''(\hat{c}_2)d_1^2)(\epsilon+\hat{c}_2\eta)}{u''(\hat{c}_1)d_2^2+\hat{\beta}u''(\hat{c}_2)d_1^2}$, the optimization problem becomes:

$$\max_{\eta,\epsilon}\frac{u'(\hat{c}_1)}{d_2}\cdot\frac{u'(\hat{c}_1)d_1d_2(1-\hat{\beta})\eta-(u''(\hat{c}_1)d_2^2+\hat{\beta}^2u''(\hat{c}_2)d_1^2)(\epsilon+\hat{c}_2\eta)}{u''(\hat{c}_1)d_2^2+\hat{\beta}u''(\hat{c}_2)d_1^2}$$
$$\text{s.t. }\lambda\frac{(-R\beta u''(c_2)d_1-u''(c_1)d_2)(\epsilon+c_2\eta)+u'(c_1)d_2(\frac{d_1}{d_2}-R)\eta}{u''(c_1)d_2^2+\beta u''(c_2)d_1^2}$$
$$+(1-\lambda)\frac{(-R\hat{\beta}u''(\hat{c}_2)d_1-u''(\hat{c}_1)d_2)(\epsilon+\hat{c}_2\eta)+u'(\hat{c}_1)d_2(\frac{d_1}{d_2}-R)\eta}{u''(\hat{c}_1)d_2^2+\hat{\beta}u''(\hat{c}_2)d_1^2}$$
$$=\Delta\lambda\big((R\hat{c}_1+\hat{c}_2)-(Rc_1+c_2)\big)$$

Now we show that d_1 should be increasing in λ. This is equivalent to show that $\epsilon>0$ and $\eta<0$ are optimal for previous mentioned programming. Noticing that the programming has the format as $\max_{\eta,\epsilon}A\epsilon+B\eta$ subject to $a\epsilon+b\eta=c>0$, where $A,B,a,b<0$, $\epsilon>0$ is optimal if and only if $\frac{B}{A}>\frac{b}{a}$. In our programming, $\frac{B}{A}=\hat{c}_2+\frac{(d_1/d_2)(1-\hat{\beta})}{\hat{\beta}(d_1/d_2)A(\hat{c}_2)}$, $\frac{b}{a}$ is a weighted average of $\hat{c}_2+\frac{d_1/d_2-R}{A(\hat{c}_1)+RA(\hat{c}_2)}$ and $c_2+\frac{d_1/d_2-R}{A(c_1)+RA(c_2)}$, and the weight for $c_2+\frac{d_1/d_2-R}{A(c_1)+RA(c_2)}$ is increasing in λ. If $\frac{d_1}{d_2}$ is decreasing when λ is sufficiently small, $\hat{c}_2+\frac{(d_1/d_2)(1-\hat{\beta})}{A(\hat{c}_1)+\hat{\beta}(d_1/d_2)A(\hat{c}_2)}>\hat{c}_2+\frac{d_1/d_2-R}{A(\hat{c}_1)+RA(\hat{c}_2)}$. Since $\hat{c}_2>c_2$ and $\frac{d_1}{d_2}(1-\hat{\beta})$, $\frac{d_1}{d_2}-R$ is close to 0 when $\hat{\beta}$ is close to 1, we have $\hat{c}_2+\frac{(d_1/d_2)(1-\hat{\beta})}{A(\hat{c}_1)+\hat{\beta}(d_1/d_2)A(\hat{c}_2)}>$

$c_2 + \frac{d_1/d_2 - R}{A(c_1)+RA(c_2)}$. Therefore, $\frac{B}{A} > \frac{b}{a}$ when λ is sufficiently small and $\hat{\beta}$ is close to 1, suggesting $\epsilon > 0$ and $\eta < 0$, that is, $\frac{d_1}{d_2}$ is increasing in λ. It is a contradiction. Therefore, we have $\frac{d_1}{d_2}$ is increasing in λ when λ is sufficiently small. Similarly, we can prove that d_2 is decreasing in λ when λ is sufficiently small.

On the other hand, $\frac{d_1}{d_2}$ should be increasing in λ even when λ is relatively large. Otherwise, there is a λ such that $\frac{B}{A} = \frac{b}{a}$ and ϵ could be positive or negative. Since there is a term $\frac{u'(\hat{c}_2)}{d_2}$ in the objective function, decreasing d_2 and increasing d_1 lead to higher date-0 utility. Thus, d_1 is increasing in λ and d_2 is decreasing in λ, which imply that i_1 is decreasing in λ, i_2 is increasing in λ, and $i_2 - i_1$ is increasing in λ. $\square$

Proof of Lemma 2. In 3-period model, when all agents are sophisticated, the interest rates should satisfy:

$$\left(c_2^{fb}, c_3^{fb}\right) = \text{argmax}_{x,y}\left\{u(x) + \hat{\beta}u(y) : \frac{x}{(1+i_2)^2} + \frac{y}{(1+i_3)^3} = 1 - \frac{c_1^{fb}}{1+i_1}\right\}$$

$$c_1^{fb} = \text{argmax}_z\left\{u(z) + \hat{\beta}\left[u\left(\hat{c_2}(z)\right) + u\left(\hat{c_3}(z)\right)\right] : \frac{z}{1+i_1} + \frac{\hat{c_2}(z)}{(1+i_2)^2} + \frac{\hat{c_3}(z)}{(1+i_3)^3} = 1\right\}$$

The first order conditions suggest that:

$$\frac{c_1^{fb}}{1+i_1} + \frac{c_2^{fb}}{(1+i_2)^2} + \frac{c_3^{fb}}{(1+i_3)^3} = 1$$

$$\frac{u'(c_2^{fb})}{u'(c_3^{fb})} = \hat{\beta}\frac{(1+i_3)^3}{(1+i_2)^2}$$

$$\begin{aligned} &u'\left(c_1^{fb}\right) + \hat{\beta}u'\left(c_2^{fb}\right)\frac{-\hat{\beta}u''(c_3^{fb})(1+i_3)^6/[(1+i_1)(1+i_2)^2]}{u''(c_2^{fb}) + \hat{\beta}u''(c_3^{fb})(1+i_3)^6/(1+i_2)^4} \\ &\quad + \hat{\beta}u'\left(c_3^{fb}\right)\left(\frac{-1}{1+i_1} - \frac{1}{(1+i_2)^2}\frac{-\hat{\beta}u''(c_3^{fb})(1+i_3)^6/[(1+i_1)(1+i_2)^2]}{u''(c_2^{fb}) + \hat{\beta}u''(c_3^{fb})(1+i_3)^6/(1+i_2)^4}\right)(1+i_3)^3 \\ &= 0 \end{aligned}$$

The last two equations suggest that for any sophisticated agent, his/her consumption should satisfy:

$$\frac{u'(c_1)}{u'(c_2)} = \frac{(1+i_2)^2}{(1+i_1)}\left(1 - (1-\hat{\beta})\frac{\hat{\beta}u''(c_3)}{u''(c_2)((1+i_2)^2/(1+i_3)^3)^2 + \hat{\beta}u''(c_3)}\right)$$

Notice that $c_1^{fb} = c_2^{fb} = 0.5(1 - \frac{c_3^{fb}}{R})$ and $\frac{R}{\hat{\beta}} = \frac{(1+i_3)^3}{(1+i_2)^2}$, $\frac{(1+i_2)^2}{(1+i_1)} = \frac{\hat{\beta}u''(c_2^{fb})+R^2u''(c_3^{fb})}{\hat{\beta}u''(c_2^{fb})+\hat{\beta}R^2u''(c_3^{fb})}$, and, by simple algebra, we have

$$1+i_1 = c_1^{fb}\left(1 + \frac{1+i_1}{(1+i_2)^2}\right) + \left(1 - 2c_1^{fb}\right)\hat{\beta}\frac{1+i_1}{(1+i_2)^2} < 1$$

$$(1+i_2)^2 = c_2^{fb}\left(1 + \frac{(1+i_2)^2}{1+i_1}\right) + \left(1 - 2c_2^{fb}\right)\hat{\beta} < 1$$

$$(1+i_3)^3 = 0.5\frac{R}{\hat{\beta}}\left(1 + \frac{(1+i_2)^2}{1+i_1}\right)\left(1 - \frac{c_3^{fb}}{R}\right) + c_3^{fb} > 1 \quad \square$$

Proof of Proposition 8. First we prove that financial intermediaries make zero profits in equilibrium. Suppose financial intermediaries make non-zero profits. Then all financial intermediaries of the same type can deviate to slightly increase K and make SC less tight which will increase agents' utility and attract all agents, which means there exists a profitable deviation. So each financial intermediary must have zero profit in equilibrium.

We can also show the existence of the competitive equilibrium with zero profits, that is, there exist a positive K such that the maximal profit in problem (12) is zero. Obviously, the maximum of problem (12) is positive when $K = 0$ as we discussed in section 5.1, and the maximum of problem (12) is negative when K is sufficiently large due to the boundedness of profits. According to Berger's maximum theorem (Aliprantis and Border, 2007), the maximum value of problem (12) is a continuous function of K, therefore there exist a K such that the optimal profit of financial intermediaries is zero.

Then we prove that both naïve agents and sophisticated agents get lower date-0 utility than in the case without the solvency requirement. We showed that $R < \widetilde{R}$ and $\hat{c}_1 + \hat{c}_2/R > 1$ in Proposition 7, which implies that if SC is satisfied we must have $K > 0$. Since profits are non-negative, we have $\lambda(c_1 + c_2/R) + (1-\lambda)(\hat{c}_1 + \hat{c}_2/R) < 1 - \gamma K/R < 1$. Because the budget constraint is tighter, the date-0 utility $u(\hat{c}_1) + u(\hat{c}_2)$ must be lower than in the case without SC; otherwise a financial intermediary can deviate by increasing T by a small amount and relax PC while making positive profits.

Finally we show agents' date-0 utility is lower when the unit cost of capital is higher. The equilibrium choice of capital is to balance the gain from relaxing the solvency constraint and the cost of capital. So if $\gamma_0 < \gamma$, then at the equilibrium capital associated a lower unit cost of γ_0, now the marginal cost already exceeds the marginal benefit when the true unit cost of capital is γ, so financial intermediaries will lower the capital level and finally the equilibrium capital is also lowered, which leads to a lower agents' date-0 utility. □

Proof of the Proposition 9. With a linear cost of the capital, the choice of an intermediary could be divided into two steps. In the first step, the intermediary calculates the optimal $(T, \widetilde{R}, K)$ to given date-0 utility $\underline{u}$ and unit cost of capital γ, and we denote the optimal contract as a correspondence $CT(\underline{u}, \gamma)$, and denote the maximal profit under the optimal contract $CT(\underline{u}, \gamma)$ as $PF(\underline{u}, \gamma)$. In the second step, the intermediary chooses the optimal date-0 utility $\underline{u}$ to maximize $PF(\underline{u}, \gamma)$ given the function $CT(\underline{u}, \gamma)$.

The optimization problem for the first step is as the following:

$$\max_{K,T,\widetilde{R}} \lambda\big[R(1-c_1) - c_2\big] + (1-\lambda)\big[R(1-\hat{c}_1) - \hat{c}_2\big] - \gamma K$$

$$\text{s.t. } u(\hat{c}_1) + u(\hat{c}_2) \geq \underline{u} \quad (PC)$$

$$(c_1, c_2) = \text{argmax}_{x,y}\big\{u(x) + \beta u(y) \mid x + y/\widetilde{R} \leq T\big\} \quad (IC_1)$$

$$(\widetilde{c}_1, \widetilde{c}_2) = \text{argmax}_{x,y}\big\{u(x) + \widetilde{\beta} u(y) \mid x + y/\widetilde{R} \leq T\big\} \quad (IC_2)$$

$$\max\left\{\left(c_1 + \frac{c_2}{R}\right), \left(\hat{c}_1 + \frac{\hat{c}_2}{R}\right)\right\} \leq 1 + K \quad (SC)$$

And it is easy to show the maximal value of this programming $PF(\underline{u}, \gamma)$ is a decreasing function of $\underline{u}$ and γ because larger $\underline{u}$ or γ essentially leads to smaller feasible set. The absolute marginal change of $PF(\underline{u}, \gamma)$ with respect to $\underline{u}$ is an increasing function of γ for a similar reason.

The optimization problem for the second step for intermediary i given the other intermediary's strategy could be stated as the following:

$$\max_{\underline{u}_i} PF(\underline{u}_i, \gamma_i)\frac{\underline{u}_i - \underline{u}_j + t}{2t}$$

We denote the optimal response function of intermediary i as $\underline{u}_i(\underline{u}_j, \gamma_i)$. We could show that $\underline{u}_i(\underline{u}_j, \gamma_i)$ is a decreasing function of γ_i because the benefits of obtaining more market share dominates the loss due to increasing contract's welfare for lower γ_i owing to the properties of $PF(\underline{u}, \gamma)$.

We first prove there exist a Nash equilibrium. It is equivalent to the existence of solution $(\underline{u}_1^*, \underline{u}_2^*)$ of the following equations.

$$\underline{u}_1^* = \text{argmax}_{\underline{u}_1}\ PF(\underline{u}_1, \gamma_1)\frac{\underline{u}_1 - \underline{u}_2^* + t}{2t}$$

$$\underline{u}_2^* = \text{argmax}_{\underline{u}_2}\ PF(\underline{u}_2, \gamma_2)\frac{\underline{u}_2 - \underline{u}_1^* + t}{2t}$$

That is, there should be an intersection between curves $\underline{u}_1(\underline{u}_2, \gamma_1)$ and $\underline{u}_2(\underline{u}_1, \gamma_2)$.

Noticing that $\underline{u}_i(\underline{u}_0, \gamma_i) = \underline{u}_0 + t$ when γ_i is small because it is always optimal to obtain more market shares when the unit cost of capital is small. Therefore we must have $\underline{u}_i(\underline{u}_0 + t, \gamma_i) > \underline{u}_0$ when γ_i is small. This implies that there is $\underline{u}_j^0$ such that $\underline{u}_j^0 > \underline{u}_0 + t$ and $\underline{u}_i(\underline{u}_j^0, \gamma_i) = \underline{u}_j^0 - t + \epsilon$ for sufficiently small ϵ. Due to the boundedness of the feasible date-0 utility, there also exists $\underline{u}_\gamma$ such that $\underline{u}_i(\underline{u}_\gamma, \gamma) = \underline{u}_\gamma$, i.e., $\underline{u}_\gamma$ is the equilibrium date-0 utility when the two intermediaries have same unit cost of capital γ. Obviously, $\underline{u}_{\gamma_2} < \underline{u}_{\gamma_1}$.

We thus show that the curve $\underline{u}_1(\underline{u}_2, \gamma_1)$ passes through points $(\underline{u}_{\gamma_1}, \underline{u}_{\gamma_1})$ and $(\underline{u}_0 + t, \underline{u}_0)$, and the curve $\underline{u}_2(\underline{u}_1, \gamma_2)$ passes through points $(\underline{u}_{\gamma_2}, \underline{u}_{\gamma_2})$ and $(\underline{u}_1^0, \underline{u}_1^0 - t + \epsilon)$. $\underline{u}_{\gamma_2} < \underline{u}_{\gamma_1}$ and $\underline{u}_1^0 > \underline{u}_0 + t$ suggest there must be an intersection between curves $\underline{u}_1(\underline{u}_2, \gamma_1)$ and $\underline{u}_2(\underline{u}_1, \gamma_2)$, and the intersection point $(\underline{u}_1^*, \underline{u}_2^*)$ satisfies $\underline{u}_1^* > \underline{u}_2^*$. We thus demonstrate the existence of the Nash equilibrium and in this equilibrium intermediary 1 provides a contract with higher date-0 utility. As a consequence, intermediary 1 also obtains a larger market share.

We now show when both intermediaries make strictly positive profits by contradiction. Without loss of generality, we may assume intermediary 1 makes zero profit and attracts a $\alpha(> 0)$ portion of depositors in equilibrium. This implies that the profit from each investor provided by intermediary 1 is equal to the average cost of capital, i.e., $PF_1 = \gamma_1$, where PF_1 is profit from a representative depositor for intermediary 1. Given the strategy of intermediary 2, if intermediary 1 changes the contract such that the new contract's date-0 utility is $\underline{u}_1^0 - \epsilon$, where $\underline{u}_1^0$ is the date-0 utility associated with the original contract and ϵ is sufficiently small. Then intermediary 1 will lose market shares $\frac{\epsilon}{2t}$ but makes small profits from each investor, suggesting the total profits intermediary 1 earns is strictly positive. This is a contradiction. As a consequence, in equilibrium both intermediaries make strictly positive profits.

Finally, we show intermediary 1 makes more profits in the equilibrium. We may write the total profits of intermediary 1 as αPF_1 and the profits of intermediary 2 as $(1 - \alpha)PF_2$. Suppose that intermediary 1 makes less profits than intermediary 2, that is, $\alpha PF_1 < (1 - \alpha)PF_2$. According to above result, we know $\alpha > 0.5$ and $PF_1 < PF_2$. Now let intermediary 1 adopt the strategy of intermediary 2. It is easy to know that the profits of intermediary 2 would be $0.5PF_2$ and the profits of intermediary 1 would be greater than $0.5PF_2$ due to the lower cost of intermediary 1. Thus intermediary 1 should deviate from its original strategy because $0.5PF_2 > \alpha PF_1$. It is a contradiction. Therefore, intermediary 1 must make more profits than intermediary 2 in the equilibrium. □

References

Aliprantis, Charalambos D., Border, Kim C., 2007. Infinite Dimensional Analysis: A Hitchhiker's Guide, 3rd ed. Springer.

Amador, Manuel, Werning, Iván, Angeletos, George-Marios, 2006. Commitment vs. flexibility. Econometrica 74, 365–396.

Ambrus, Attila, Egorov, Georgy, 2012. Commitment-flexibility trade-off and withdrawal penalties. Unpublished manuscript.

Ang, Andrew, Piazzesi, Monika, Wei, Min, 2006. What does the yield curve tell us about GDP growth? J. Econom. 131, 359–403.

Ashraf, Nava, Gons, Nathalie, Karlan, Dean S., Yin, Wesley, 2003. A Review of Commitment Savings Products in Developing Countries. Asian Development Bank Economic and Research Department Working Paper No. 45.

Basu, Karna, 2011. Hyperbolic discounting and the sustainability of rotational savings arrangements. Am. Econ. J. Microecon. 3, 143–171.

Brooks, Robert, 1996. Computing yields on enhanced CDs. Financial Services Review 5, 31–42.

Cline, Brandon N., Brooks, Robert, 2004. Embedded options in enhanced certificates of deposit. Financial Services Review 13, 19–32.

DellaVigna, Stefano, Malmendier, Ulrike, 2004. Contract design and self-control: theory and evidence. Q. J. Econ. 119, 353–402.

Diamond, Douglas W., Dybvig, Philip H., 1983. Bank runs, deposit insurance and liquidity. J. Polit. Econ. 91, 401–419.

Diamond, Peter, Köszegi, Botond, 2003. Hyperbolic discounting and retirement. J. Public Econ. 87, 1839–1872.

Duffie, Darrell, Kan, Rui, 1996. A yield-factor model of interest rates. Math. Finance 6, 379–406.

Eliaz, Kfir, Spiegler, Ran, 2006. Contracting with diversely naïve agents. Rev. Econ. Stud. 73, 689–714.

Estrella, Arturo, Hardouvelis, Gikas A., 1991. The term structure as predictor of real economic activity. J. Finance 46, 555–576.

Frederick, Shane, Loewenstein, George, O'Donoghue, Ted, 2002. Time discounting and time preference: a critical review. J. Econ. Lit. 40, 351–401.

Fellowes, Matt, Willemin, Katy, 2013. The retirement breach in defined contribution plans: size, causes, and solutions. HelloWallet, 1–22.

Fisher, Irving, 1930. The Theory of Interest. Macmillan, New York.

Gabaix, Xavier, Laibson, David, 2006. Shrouded attributes, consumer myopia, and information suppression in competitive market. Q. J. Econ. 121, 505–540.

Gilkeson, James H., List, John A., Ruff, Craig K., 1999. Evidence of early withdrawal in time deposit portfolios. J. Financ. Serv. Res. 15, 103–122.

Gottlieb, Daniel, Smetters, Kent, 2014. Lapse-Based Insurance. Working paper.

Heidhues, Paul, Köszegi, Botond, 2010. Exploiting naïveté about self-control in the credit market. Am. Econ. Rev. 100, 2279–2303.

Heidhues, Paul, Köszegi, Botond, Murooka, Takeshi, 2012. Inferior Products and Profitable Deception. Working paper.

Kurmann, André, Otrok, Christopher, 2013. News shocks and the slope of the term structure of interest rates. Am. Econ. Rev. 103, 2612–2632.

Laibson, David, 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112, 443–477.

Laibson, David, 1998. Life-cycle consumption and hyperbolic discount functions. Eur. Econ. Rev., Pap. Proc. 42, 861–871.

Lin, Ping, 1996. Banking, incentive constraints, and demand deposit contracts with nonlinear returns. Econ. Theory 8, 27–39.

O'Donoghue, Ted, Rabin, Matthew, 2001. Choice and procrastination. Q. J. Econ. 116, 121–160.

O'Donoghue, Ted, Rabin, Matthew, 2006. Optimal sin taxes. J. Public Econ. 90, 1825–1849.

Salanie, Francois, Treich, Nicolas, 2006. Over-savings and hyperbolic discounting. Eur. Econ. Rev. 50, 1557–1570.