Economic Quarterly—Volume 99, Number 4—Fourth Quarter 2013—Pages 305–340

Pecuniary Externalities, Segregated Exchanges, and Market Liquidity in a Diamond-Dybvig Economy with Retrade

Borys Grochulski

Price changes affect economic agents primarily by altering their budget constraints. In many economic environments, however, price changes additionally impact the agents by altering other constraints agents face. Those additional ways in which prices affect agents, other than through budget constraints, are known as pecuniary externalities.[1] Examples of the additional constraints that can be affected by prices include incentive compatibility, participation, and collateral constraints.

Numerous recent macroeconomic studies have shown that pecuniary externalities can lead to market failure.[2] The intuition behind this failure is as follows. In standard Arrow-Debreu economies, where

The author would like to thank Tee Kilenthong, Sam Marshall, Wendy Morrison, Pierre Sarte, Felipe Schwartzman, and Ned Prescott for their helpful comments. The views expressed in this article are those of the author and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: borys.grochulski@rich.frb.org.

[1] The term pecuniary externality has been used more broadly than this definition. Viner (1932) uses it to describe the impact of a change in the price of an input on the production cost curve of a firm. Greenwald and Stiglitz (1986) and, more recently, Bianchi (2011) and Dávila et al. (2012) among others, use it in reference to the generic constrained-inefficiency of competitive equilibria in economies with exogenously imposed market incompleteness (studied in, e.g., Stiglitz [1982] and Geanakoplos and Polemarchakis [1986]). In the language of Prescott and Townsend (1984), the definition we use corresponds to prices having a direct impact on the agents' consumption possibility set, in addition to the budget constraint.

[2] See, e.g., Kehoe and Levine (1993), Golosov and Tsyvinski (2007), Lorenzoni (2008), and Di Tella (2014).

prices only affect budget constraints, equilibrium allocations are efficient. It is therefore impossible to alter equilibrium prices (perhaps by imposing taxes) and obtain a Pareto improvement (i.e., make an agent better off without making someone else worse off). An increase in the price of good x, for example, will relax budget constraints of some agents, loosely speaking the sellers of x, making them better off, but it will tighten budget constraints of others, the buyers of x, making this group worse off. The equilibrium price of good x cannot therefore be improved upon in Pareto sense.

The same may no longer hold true when prices affect not only budget but also some other constraints that can be tightened or relaxed for all agents simultaneously. If an increase in the price of x relaxes everyone's incentive compatibility constraint, for example, then not only the sellers of x but also the buyers of x can benefit from a higher price of x, as long as the relaxed incentive constraint helps them more than the tightened budget constraint hurts them. The benevolent social planner—a stand-in concept we use to calculate optimal allocations—will take this effect into account. In a market economy, however, agents take prices as independent of their individual actions. By ignoring the general equilibrium impact of their actions on prices, agents also ignore the indirect effect they have on how tight their own incentive constraints are. The planner's and the agents' costs-benefit calculus are thus different, which leads to suboptimal equilibrium outcomes.

By relaxing a constraint that all agents face, a high price of good x has in the preceding example a positive "external" effect similar to, e.g., a clean environment or a good public highway system. Agents' inability to coordinate on a sufficiently high price for good x in equilibrium is therefore similar to the failure to internalize an external effect, which has led to the name pecuniary externality.

In this article, we discuss the pecuniary externality that leads to underprovision of liquidity in the banking model of Diamond and Dybvig (1983) (hereafter, DD). We introduce the DD economy in Section 1. In this economy, agents have access to two assets: a short-term, liquid asset with net return normalized to zero and a long-term, illiquid asset with positive net return $\hat{R} - 1 > 0$. Agents face random liquidity shocks: They may become impatient, i.e., find themselves having to consume before the illiquid asset matures, or remain patient, in which case they can postpone consumption until the illiquid asset pays off. By investing a part of their initial endowment/wealth in the low-yielding liquid asset, agents purchase insurance against the liquidity shock.

In Section 2, we derive the efficient allocation of liquidity in this economy, i.e., the optimal levels of investment in the two assets along with the resulting amounts of consumption for the agents who do and do not experience the need for liquidity. At the optimum, the liquidity shock is partially insured: The impatient agents are able to capture a part of the return on the long-term asset despite the fact that they have to consume before this asset matures.

There are several variants of the DD model in the literature. The variant we consider follows closely Jacklin (1987) and Farhi, Golosov, and Tsyvinski (2009). It has been designed to focus on market provision of liquidity and not on the possibility of bank runs.[3] In particular, we assume that liquidity shocks are agents' private information, but we do not assume a sequential service constraint: Trade can be organized after all agents have received their realizations of the liquidity shock. To study pecuniary externalities, we follow Farhi, Golosov, and Tsyvinski (2009) in giving the agents access to an anonymous, hidden market in which they can borrow and lend at the market-determined gross rate of return R. As this rate of return (the price of credit) affects incentive compatibility constraints, it gives rise to a pecuniary externality. This pecuniary externality makes competitive equilibria inefficient.

To show this inefficiency, we analyze in Section 3 a simple model of trade with incomplete markets. In this model, agents invest directly in the two assets ex ante and trade the long-term asset for cash ex post, i.e., after they find out their liquidity needs. Diamond and Dybvig (1983) showed that competitive equilibrium in this simple, incomplete-markets model is inefficient. In this model, a no-arbitrage condition determines how the return on the long-term asset is allocated in equilibrium: The whole net return $\hat{R} - 1$ is captured by the patient agents, leaving the impatient agents with zero net return on their investment, which is too low relative to the optimal allocation. In this incomplete-markets equilibrium, thus, agents do not obtain sufficient liquidity insurance.

This inefficiency prevails even when markets for state-contingent contracts are introduced. Jacklin (1987) and Farhi, Golosov, and Tsyvinski (2009) show that when agents can borrow and lend privately in a hidden retrade market, liquidity is underprovided in competitive equilibrium with complete markets and fully state-contingent contracts (or banks). The inefficiency is caused by a pecuniary externality that, as we mentioned, enters the model through the agents' incentive compatibility constraints that depend on the retrade interest rate R. In equilibrium, this interest rate is too high, which, by arbitrage, forces the secondary-market price for the long-term asset to be too low. The impatient agents, thus, re-sell their holdings of the long-term asset

[3] See Ennis and Keister (2010) for a review of the literature on bank runs in the DD model.

in the secondary market for too little. As in the incomplete-markets model, they are unable to capture any part of the long-run net return $\hat{R} - 1$, which again is inefficient. We review this result in detail in Section 4.

As is the case with standard externalities like pollution, the market failure caused by the pecuniary externality creates a role for government intervention. Farhi, Golosov, and Tsyvinski (2009) consider direct government intervention imposing a minimum requirement on the level of liquid investment. They show that this intervention decreases the retrade interest rate R and increases the return on the initial investment in the liquid asset. This allows the impatient agents to capture a part of $\hat{R}$ and eliminates the effect of the pecuniary externality.[4]

If the extent of an externality can be costlessly and verifiably quantified, the problem of excessive externality can also be addressed with a more decentralized approach that can be implemented through the so-called cap-and-trade mechanism. An explicit assignment of property rights over the extent of the externality lets markets for these rights emerge. In these markets, agents face prices for generating the externality, which makes them take into account the full impact of the externality and thus restores the efficiency of the equilibrium outcome.[5] Pollution is a textbook example of a negative external effect. Currently, emission of greenhouse gasses is regulated through the cap-and-trade mechanism in many countries.[6]

In a recent article, Kilenthong and Townsend (2011) (hereafter, KT) study a market solution to the pecuniary externality problem analogous to cap-and-trade.[7] In addition to a class of moral hazard environments, they consider a DD economy with retrade.[8] In their model, the impact of one's liquidity demand on the retrade interest rate is priced, which results in efficient ex ante investment, sufficient liquidity, and an optimal amount of retrade in competitive equilibrium. Clearly, this approach is interesting because it implies no need for direct government intervention into markets. Similar to the cap-and-trade

[4] In this article, we do not present details of the implementation of this intervention. The interested reader is referred directly to Farhi, Golosov, and Tsyvinski (2009).

[5] See Chapter 11 of Mas-Colell, Whinston, and Green (1995).

[6] The first and to-date largest implementation of this mechanism is the European Union Emission Trading Scheme; see Ellerman and Buchner (2007).

[7] Bisin and Gottardi (2006) use a similar approach in the Rothschild-Stiglitz adverse selection economy.

[8] Kilenthong and Townsend (2014a) study the model with segregated exchanges in a class of environments with collateral constraints. Kilenthong and Townsend (2014b) extend the analysis of segregated exchanges to a generalized framework nesting collateral and liquidity constraints, incentive constraints with retrade, and exogenously incomplete markets.

mechanism, this approach requires that agents' activities generating the externality—in this case retrade—be observable. We discuss the KT model in Section 5.

In the KT market model, retrade is allowed but only within access-controlled ex-post markets called segregated exchanges. Agents are admitted to membership in an exchange upon payment of an entry fee. The size of the entry fee depends on the composition of the agent's investment portfolio. The defining characteristic of a segregated exchange is the price at which agents expect to be able to (re)trade the long-term asset ex post. In equilibrium, these expectations must be correct. This market structure is free of pecuniary externalities because agents can no longer take retrade prices as independent of their actions. The portfolio-contingent exchange entry fee, similar to the price for greenhouse gas emissions in the cap-and-trade mechanism, creates an explicit connection between the investment decisions an agent makes ex ante and the price at which he is able to trade ex post. Consequently, equilibrium with segregated exchanges does not suffer from the problem of underprovision of liquidity, and the market outcome is efficient.

Our exposition of the KT mechanism in Section 5 extends the exposition in Kilenthong and Townsend (2011). We explicitly solve for equilibrium entry fees associated with each segregated exchange and show how with these prices the agent's ex ante utility maximization problem becomes aligned with the planner's problem of maximization of ex ante welfare.

In Section 6, we conclude the article with a discussion of the question of whether the possibility of retrade in the DD model implies the need for government intervention. The literature we review makes it clear that the answer depends on the agents' ability to commit themselves to restrict retrade to access-controlled venues with priced entry. This means that retrade itself does not imply the existence of a pecuniary externality requiring government intervention, only hidden retrade without commitment does. Which of these two kinds of retrade possibilities financial firms face in reality is an important empirical question.

The Appendix contains proofs of two auxiliary results and a precise definition of the incomplete-markets equilibrium studied in Section 3. Table 1 summarizes the frictions and outcomes associated with all allocation mechanisms we discuss in this article.

1. A DIAMOND-DYBVIG ECONOMY WITH RETRADE

The version of the Diamond-Dybvig economy that we consider here is close to those studied in Jacklin (1987); Allen and Gale (2004); Farhi, Golosov, and Tsyvinski (2009); and Kilenthong and Townsend (2011). There is a continuum of ex ante identical agents. There are three dates: $t = 0, 1, 2$. There is a single consumption good at each date. Each agent is endowed with resources e at date 0. These resources can be invested in two available technologies/assets. The short-term asset pays the return of 1 unit of the consumption good at date 1 per unit of resources invested at date 0. We will often refer to this asset as the cash asset. The long-term asset pays nothing at date 1 and $\hat{R} > 1$ at date 2 per unit invested at date 0. Note that the long-term asset is technologically illiquid at date 1, i.e., it cannot be physically turned into the consumption good.

Agents do not consume at date 0. Their preferences over consumption at dates 1 and 2 are represented by a DD utility function

$$u(c_1 + \theta c_2),$$

where $\theta \in \{0, 1\}$ is an idiosyncratic shock with $\Pr\{\theta = 0\} = \pi > 0$. Note that if $\theta = 0$, the agent is extremely impatient: He only values consumption at date 1. The standard interpretation of this shock is that with $\theta = 0$ the agent experiences at date 1 a critical need for liquidity. If $\theta = 1$, however, the agent is extremely patient: He is in fact indifferent to the timing of consumption between dates 1 and 2.[9] We follow DD in assuming that relative risk aversion is larger than 1, i.e., $-cu''(c)/u'(c) > 1$ for all c. As we will see, this assumption implies that the impatient agents will be allocated consumption with present value larger than the value of their initial endowment e.

A consumption allocation c consists of $\{c_1(0), c_2(0), c_1(1), c_2(1)\}$, where $c_t(\theta) \geq 0$ denotes date-t consumption for an agent with shock θ. Associated with allocation c are initial asset investment $s \geq 0$ in the liquid asset and $x \geq 0$ in the illiquid asset. To ensure that resources at date 1 and 2 are sufficient to provide consumption as specified in c, initial investment (s, x) associated with allocation c must satisfy

$$s \geq \pi c_1(0) + (1 - \pi) c_1(1), \tag{1}$$

[9] Note that with these preferences the DD economy violates standard smoothness and convexity assumptions. In particular, the shadow interest rate (i.e., the rate at which an agent is willing to refrain from borrowing or saving) is plus infinity for the impatient type and one for the patient type regardless of the allocation of consumption.

and

$$\hat{R}x \geq \pi c_2(0) + (1 - \pi)c_2(1). \tag{2}$$

The amounts s and x that can be invested in the two technologies are constrained by the amount e of resources available at date 0:

$$s + x \leq e. \tag{3}$$

Substituting (1) and (2) into (3), we can express the economy's aggregate resource constraint in terms of just the consumption allocation c:

$$\pi \left(c_1(0) + \frac{c_2(0)}{\hat{R}} \right) + (1 - \pi) \left(c_1(1) + \frac{c_2(1)}{\hat{R}} \right) \leq e. \tag{4}$$

Allocation c gives an agent an expected utility value of

$$\mathbb{E}[u(c_1 + \theta c_2)] = \pi u(c_1(0)) + (1 - \pi)u(c_1(1) + c_2(1)). \tag{5}$$

Since all agents are ex ante identical, the expected utility of the representative agent measures total utility, or social welfare, attained in this economy.

We follow DD in assuming that realizations of θ are private information. That is, given an allocation $c = \{c_1(0), c_2(0), c_1(1), c_2(1)\}$, an agent can obtain either $\{c_1(0), c_2(0)\}$ or $\{c_1(1), c_2(1)\}$ depending on what realization of θ he reports.

In addition, we follow Farhi, Golosov, and Tsyvinski (2009) and Kilenthong and Townsend (2011) in assuming that individual final consumption is also private and that agents have access to a hidden retrade market where they can lend and borrow from one another "behind the back" of the planner, i.e., with all trades in this market being hidden from everyone but the parties directly involved. More precisely, at date 1 agents have access to a perfectly competitive market for one-period IOUs. Given an allocation $c = \{c_1(0), c_2(0), c_1(1), c_2(1)\}$, an agent reporting shock realization $\tilde{\theta}$ obtains the bundle $(c_1(\tilde{\theta}), c_2(\tilde{\theta}))$. But this bundle does not have to be his actual consumption. Rather, this bundle becomes his endowment of goods in the hidden retrade market. The agent's final consumption is determined by his retrade activity. At the hidden-market interest rate R, the agent can either save some of his $c_1(\tilde{\theta})$ for consumption at date 2, or borrow against $c_2(\tilde{\theta})$ for consumption at date 1. Specifically, given an allocation c and a gross interest rate R in the hidden retrade market, an agent of type θ selects a report $\tilde{\theta} \in \{0, 1\}$, IOU purchases b, and a final consumption bundle

$(\tilde{c}_1, \tilde{c}_2) \geq (0, 0)$ that solve

$$\begin{aligned} \tilde{V}(c, R; \theta) \quad = \quad & \max_{\tilde{\theta}, \tilde{c}_1, \tilde{c}_2, b} u(\tilde{c}_1 + \theta \tilde{c}_2) \\ & s.t. \\ & \tilde{c}_1 + b \leq c_1(\tilde{\theta}), \\ & \tilde{c}_2 \leq Rb + c_2(\tilde{\theta}). \end{aligned} \tag{6}$$

The value $\tilde{V}(c, R; \theta)$, thus, is determined by the agent's best strategy with respect to reporting his realization of the shock θ as well as saving/borrowing in the hidden market.

Allocation c is incentive compatible (IC) if agents prefer to reveal their type truthfully and not use the retrade market. That is, c is IC if it satisfies

$$u(c_1(\theta) + \theta c_2(\theta)) \geq \tilde{V}(c, R; \theta) \tag{7}$$

for both θ, with R being an equilibrium gross interest rate in the hidden retrade market.

2. OPTIMAL ALLOCATION

In this section, we first provide a result of DD characterizing the best allocation with no frictions (i.e., without private information or hidden retrade), which is often referred to as the first-best allocation. This allocation provides the highest social welfare among all allocations that are resource feasible, i.e., it maximizes (5) subject to (4). Next, we present a result of Farhi, Golosov, and Tsyvinski (2009) showing that the first-best allocation remains feasible even with the frictions of private θ and hidden retrade. The first-best allocation thus remains optimal in this environment, even with these two frictions present.

Optimal Allocation with no Frictions

Let us start out by noting that given the infinite impatience of the agents of type $\theta = 0$, it is never efficient in this economy to have the impatient types consume a positive amount at date 2. Likewise, given the complete patience of type $\theta = 1$ and $\hat{R} > 1$, it is never efficient to have the patient types consume a positive amount at date 1.

Lemma 1 *If $c = \{c_1(0), c_2(0), c_1(1), c_2(1)\}$ maximizes (5) subject to (4), then $c_2(0) = c_1(1) = 0$.*

Proof. In the Appendix. ■

Below, we will often write c_1 for $c_1(0)$ and c_2 for $c_2(1)$, silently assuming $c_2(0) = c_1(1) = 0$, and refer to (c_1, c_2) as an allocation. With these notational shortcuts, the social welfare function (5) can be written simply as

$$\pi u(c_1) + (1-\pi)u(c_2), \tag{8}$$

the aggregate resource constraint (4) as

$$\pi c_1 + (1-\pi)\frac{c_2}{\hat{R}} \leq e, \tag{9}$$

and first-best allocation can be defined as a maximizer of (8) subject to just (3), i.e., ignoring the incentive constraint (7). Further, from (1) and (2) we have $c_1 = \frac{s}{\pi}$ and $c_2 = \frac{x}{1-\pi}\hat{R}$. If no initial wealth is to be wasted, we must have $x = e - s$. We can thus express any resource-feasible allocation (c_1, c_2) as a function of the initial liquid investment s alone:

$$(c_1, c_2) = \left(\frac{s}{\pi}, \frac{e-s}{1-\pi}\hat{R}\right)$$

with $s \in [0, e]$. The social welfare function (8) can thus be written as

$$\pi u\left(\frac{s}{\pi}\right) + (1-\pi)u\left(\frac{e-s}{1-\pi}\hat{R}\right). \tag{10}$$

Denote this function by $W(s)$. The first-best planning problem is reduced here to finding a level of liquid investment s in $[0, e]$ that maximizes $W(s)$. Denote such a level by s^*. The corresponding level of illiquid investment is $x^* = e - s^*$ and the first-best optimal allocation is $(c_1^*, c_2^*) = \left(\frac{s^*}{\pi}, \frac{e-s^*}{1-\pi}\hat{R}\right)$.

Proposition 1 *(Diamond and Dybvig) The social welfare function $W(s)$ has a unique maximizer s^* in $[0, e]$. The maximizer satisfies*

$$\pi e < s^* < \pi\frac{\hat{R}}{\pi\hat{R} + 1 - \pi}e. \tag{11}$$

Proof. In the Appendix. ■

The two inequalities in (11) imply that the first-best consumption allocation (c_1^*, c_2^*) satisfies

$$e \quad < \quad c_1^* < \frac{\hat{R}e}{\pi\hat{R} + 1 - \pi}, \tag{12}$$

$$\hat{R}e \quad > \quad c_2^* > \frac{\hat{R}e}{\pi\hat{R} + 1 - \pi}. \tag{13}$$

The right inequalities above show that the first-best allocation does not provide full insurance, $c_1^* < \frac{\hat{R}e}{\pi\hat{R}+1-\pi} < c_2^*$. The reason for this is that first-period consumption is more expensive to provide than second-period consumption. At the full-insurance allocation

$$c_1 = c_2 = \frac{\hat{R}e}{\pi\hat{R} + 1 - \pi}, \tag{14}$$

marginal utility of consumption is the same at both dates, but by giving up $\varepsilon > 0$ units of consumption at date 1 the planner can deliver $\hat{R}\varepsilon > \varepsilon$ units of consumption at date 2. Such a reallocation would therefore increase overall expected welfare, and so full insurance is not optimal.

The left inequality in (11) implies that the first-best allocation gives a larger present value of consumption to impatient agents than to patient ones. Indeed, discounting consumption at date 1 and 2 at, respectively, the rate of return of the short- and long-term asset, and using the left inequalities in (12) and (13), shows

$$\frac{c_1^*}{1} > e > \frac{c_2^*}{\hat{R}}. \tag{15}$$

The optimality of this unequal allocation of the present value of consumption follows because relative risk aversion of the utility function $u(c)$ larger than 1 means that as consumption c increases, marginal utility of consumption $u'(c)$ drops fast (faster than $1/c$). Liquid investment $s = \pi e$ gives a final consumption allocation $(c_1, c_2) = (e, \hat{R}e)$, where the present value of both types' consumption is the same (and equal to the per capita initial endowment):

$$\frac{c_1}{1} = e = \frac{c_2}{\hat{R}}. \tag{16}$$

At this allocation, however, $c_2 = \hat{R}e > e = c_1$, so the marginal utility of c_2 is low and the marginal utility of c_1 is high. By increasing the liquid investment s at date 0 above $s = \pi e$, say by $\varepsilon > 0$, the planner gives up the return $\hat{R}\varepsilon$ but is able to increase consumption in the high marginal utility state, i.e., at date 1. On balance, this is an improvement because $u'(c_1)$ is sufficiently high relative to $u'(c_2)$ and $\hat{R}$ [that is, $\varepsilon u'(e) > \hat{R}\varepsilon u'(\hat{R}e)$].

Alternatively, we can express this intuition using the elasticity of substitution of the utility function u. With zero elasticity of substitution (Leontief preferences), the full insurance allocation (14) would be optimal. With unit elasticity of substitution (logarithmic preferences), the allocation (16) spending the same amount on each good would be optimal. Under the DD assumption of the elasticity of substitution larger than zero but smaller than one, it is optimal to make c_1 and c_2 closer to one another than under logarithmic preferences, but not go all the way to full insurance.

Optimal Allocation with Private Shocks and Retrade

Having characterized the optimal allocation in the first-best version of the DD environment, we now ask what the optimal allocation is with private information and a hidden retrade market, i.e., with the addition of the IC constraint (7).

With realizations of θ being private information and with agents having access to retrade, Farhi, Golosov, and Tsyvinski (2009) show that the first-best allocation is incentive compatible, i.e., remains feasible and thus optimal. This result is obtained as follows. The retrade interest rate R associated with the optimum (i.e., the shadow interest rate at the first-best), denoted by R^*, is

$$R^* = \frac{c_2^*}{c_1^*}. \tag{17}$$

First, let us check that with "endowments" (c_1^*, c_2^*), the interest rate $R = R^*$ is an equilibrium interest rate in the hidden market. Note that from $c_2^* > c_1^*$ we get $R^* > 1$ and from $\frac{c_1^*}{1} > e > \frac{c_2^*}{\hat{R}}$ we get $R^* < \hat{R}$, so $1 < R^* < \hat{R}$. Suppose the impatient types enter the hidden market with an endowment vector $(c_1^*, 0)$ and patient types enter with $(0, c_2^*)$. The impatient agent has no income at $t = 2$, so he cannot borrow in this hidden market (for there is nothing he could pay back with). Also, this agent wants to consume his income c_1^* irrespective of the interest rate. Thus, the impatient type's utility is maximized with the quantity of zero traded at the interest rate R^*. A patient agent could borrow against his date-2 endowment c_2^* and consume at date 1, but $R^* > 1$ implies he would not want to do it, as his marginal utility of consumption is the same at either date and he can consume only $\frac{c_2^*}{R^*} < c_2^*$ if he decides to use the hidden market and consume at date 1. This confirms that consumption (c_1^*, c_2^*) and interest rate R^* are an equilibrium in the retrade market (with zero quantity traded in equilibrium).

Now consider potential deviations in the revelation of θ combined with retrade. The first-best allocation is immune to these deviations because at the interest rate R^* the present value of each type's endowment is the same. Indeed, the impatient types could claim endowment $(0, c_2^*)$ and borrow against c_2^* in order to consume at date 1, but doing so would give them $\frac{c_2^*}{R^*} = c_1^*$ units of consumption, so there is no gain for them from doing so. As well, the patient types could claim endowment $(c_1^*, 0)$ and save at the market interest rate R^*. But doing so gives them final consumption $R^* c_1^* = c_2^*$ so, again, no gain. This confirms that the first-best allocation is incentive compatible in the model with private information and hidden retrade.

Note that although the possibility of hidden retrade does not change the optimal allocation, it does change the IC constraint. With just private information about the liquidity shock θ (without retrade), the IC constraint would be $c_2 \geq c_1$. The first-best allocation satisfies this constraint as a strict inequality simply because $c_2^* > c_1^*$. With the hidden retrade market, however, the IC constraint holds only as an equality because $\frac{c_2^*}{R^*} = c_1^*$.

Next, we move on to discuss market provision of liquidity in this environment.

3. COMPETITIVE EQUILIBRIUM WITH INCOMPLETE MARKETS

The remainder of this article is devoted to studying competitive equilibrium outcomes under three different market arrangements, and comparing these outcomes with the optimal allocation (c_1^*, c_2^*).

In this section, we discuss a simple incomplete-markets model of trade, in which agents invest directly in the two assets and subsequently trade them (i.e., there are no intermediaries, no state-contingent contracts). This natural model of trade is a point of departure for Diamond and Dybvig (1983). DD start their analysis of market provision of liquidity by considering this incomplete market structure. They conclude that the equilibrium level of liquidity is too low, i.e., there is a market failure. We briefly review this result in this section and move on to showing in the next section that with hidden retrade this conclusion generalizes to any market structure (even when state-contingent contracts and/or intermediaries are taken into consideration).

The simple market structure is as follows. At date 0, each agent invests directly in the two assets subject to $s + x \leq e$. At date 1, after agents find out their type θ, they trade the long-term asset for cash at a market-determined price p. In addition to the market for the long-term asset, agents have access at date 1 to a market for one-period IOUs.[10] A formal statement of the agents' optimization problem and competitive equilibrium in this economy is given in the Appendix. Note that this market structure is incomplete: There are no contracts for provision of consumption conditional on θ.

A simple arbitrage argument shows that in any equilibrium of this trading arrangement the date-1 cash price p of a unit of the long-term

[10] As we will see, however, the (hidden) IOU market will not be active here, nor imposing any binding constraints on the equilibrium allocation.

asset must be 1. This argument is as follows. The fact that a market for the long-term asset exists at date 1 makes the long-term asset de facto liquid and thus a perfect substitute, at date 0, for the short-term asset. The return from holding the long-term asset for one period, therefore, must be the same as the return from investing in the short-term asset. The date-1 price of the long-term asset must therefore be $p = 1$, or else there is an arbitrage.

Indeed, if $p > 1$, all agents want to invest their initial resources in the long-term asset only, as investing a unit of resources in that asset and selling it at date 1 yields p, while investing in the short-term asset yields 1. In this case, however, nobody has cash at date 1 and thus aggregate demand for the long-term asset is zero. This level of demand is inconsistent with the equilibrium price p being positive. Similarly, if $p < 1$, all agents want to invest exclusively in the short-term asset at date 0, as investing a unit of resources in the long-term asset is dominated by investing this unit in the short-term asset and then buying the long-term asset at date 1 at price $p < 1$. This, however, means that supply of the long-term asset at date 1 is zero while demand is positive, as the patient types are willing to buy at $p < 1$. Thus, $p < 1$ cannot be an equilibrium price, either.[11]

The only price p consistent with equilibrium, therefore, is $p = 1$. At this price, the return from holding the short- and the long-term asset from date 0 to date 1 is the same, so agents are indifferent between investments s and x. At date 1, the impatient agents want to sell their holdings x of the illiquid asset. With $p = 1$, the patient agents want to hold on to their x and spend their cash s to purchase additional units of the long-term asset, as the return on this investment, $\frac{\hat{R}}{p} = \hat{R}$, exceeds their required rate of return, 1. Aggregate supply of the long-term asset to the market at date 1 is therefore πx and the supply of cash is $(1 - \pi)s$. The market-clearing condition, thus, is

$$\pi x p = (1 - \pi)s,$$

where, by the arbitrage argument given above, $p = 1$. The date-0 budget constraint implies

$$x = e - s.$$

Solving the above two conditions, we obtain

$$s = \pi e, \; x = (1 - \pi)e. \tag{18}$$

[11] Strictly speaking, these corner investment strategies are not arbitrages because they are not self-financing. But they could be turned into arbitrages if agents could short the expensive asset at date 0.

This solution is unique, so there exists only one equilibrium. In equilibrium, consumption of the impatient types is $c_1 = s + px = \pi e + 1(1-\pi)e = e$, while the patient types consume $c_2 = \left(x + \frac{s}{p}\right)\hat{R} = \left((1-\pi)e + \frac{\pi e}{1}\right)\hat{R} = e\hat{R}$. Let us denote the unique equilibrium consumption bundle by $(\hat{c}_1, \hat{c}_2)$. We have just shown that

$$(\hat{c}_1, \hat{c}_2) = (e, \hat{R}e). \tag{19}$$

In the hidden retrade market, there is no active trade. The equilibrium retrade interest rate is $R = \hat{R}$. At this rate, agents choose not to alter their consumption allocation $(\hat{c}_1, \hat{c}_2)$ by either borrowing or lending. The hidden retrade market has no impact on the equilibrium outcome here because the (regular, "non-hidden") date-1 market for the long-term asset already offers a riskless return $\frac{\hat{R}}{p} = \hat{R} = R$. The hidden IOU retrade market is thus redundant.

A key property of the DD environment is that the equilibrium allocation of consumption, $(\hat{c}_1, \hat{c}_2)$, is inefficient. That is, this allocation yields lower ex ante welfare than the optimal allocation c^*. Clearly, the right inequalities in (12) and (13) tell us that $c_1^* > \hat{c}_1$ and $c_2^* < \hat{c}_2$. Since, by Proposition 1, the optimum (c_1^*, c_2^*) is a unique welfare maximizer, equilibrium allocation $(\hat{c}_1, \hat{c}_2)$ is indeed inefficient.

As we saw in Section 2, optimal allocation calls for a present-value transfer from the patient types to the impatient types. In equilibrium with incomplete markets, however, each agent consumes the worth of his own initial endowment, e, i.e., there are no present value transfers between types, and insurance markets are missing. Moreover, it is easy to see that an intervention by a benevolent planner/government can improve welfare without introducing any new markets. If the planner forces each agent to invest $(s, x) = (s^*, x^*)$ at date 0 and allows free trade at date 1, the market price for the long-term asset will be $p = p^*$, the retrade market rate will be $R = R^*$, and the equilibrium consumption allocation will be (c_1^*, c_2^*).[12]

In sum, the equilibrium investment in the liquid asset is too low relative to the optimum, $s = \pi e < s^*$, i.e., free trade leads to underprovision of liquidity.

[12] In the language of the incomplete-markets literature, equilibrium $(\hat{c}_1, \hat{c}_2)$ is constrained-inefficient.

4. COMPETITIVE EQUILIBRIUM WITH CONTINGENT CONTRACTS

In this section, we allow for state-contingent contracts. We review the following important result. Jacklin (1987) points out that when retrade is allowed, an arbitrage argument similar to the one used in the previous section implies that markets will underprovide liquidity, even when fully state-contingent contracts are allowed. With retrade, thus, the market failure shown in the previous section for the simple incomplete-markets model continues to hold for all feasible models of trade in the DD environment, including the intermediation economy of Diamond and Dybvig (1983).

Consider the following general model of trade with fully state-contingent contracts, direct investment, and retrade.[13] In addition to directly investing in the two assets, agents can contract with intermediaries and access the hidden IOU market. Intermediaries, or banks, make available to agents at date 0 a state-contingent contract (ξ_1, ξ_2). Under this contract, which can be thought of as a deposit contract, the agent can obtain from the intermediary, at the agent's discretion, either ξ_1 at date 1 or ξ_2 at date 2 (but not both). Let us normalize the price of this contract to e, i.e., an agent who accepts a contract deposits his whole initial wealth with a bank. Also, as before, agents can borrow from and lend to each other privately in the hidden retrade market at date 1.

Under this market structure, an agent has the following choices to make. At date 0, he decides whether to deposit his wealth e with a bank or to invest directly in assets s and x. If he deposits, after he learns his type θ, he chooses whether to withdraw at date 1 or 2, and how much, if at all, to borrow or lend in the hidden retrade market at the market rate R. If the agent chooses not to deposit at date 0 but rather to invest directly, he selects a portfolio (s, x). At date 1, after he learns his type and his cash investment s matures, the agent decides how much to borrow or lend in the retrade market at the market rate R.

Competition among banks (existing or potential entrants) drives banks' profits to zero and forces each active bank to offer the same contract (namely, the contract that maximizes the ex ante expected utility of the representative agent, for otherwise agents would deposit with a different bank). Since intermediation is an activity with constant returns to scale in this model, it is without loss of generality to assume

[13] For a formal statement a version of this economy see Section 3.1 of Farhi, Golosov, and Tsyvinski (2009) or Allen and Gale (2004).

that a single large bank operates in equilibrium (the market, however, is perfectly contestable).

The bank's contract design problem is similar to the social planning problem in that in both cases the objective is to maximize the agent's expected utility. There is, however, a key difference. The planner can control date-0 investment, which enables her to have an (indirect) impact on the retrade market interest rate R. The bank cannot force the agents to deposit, which means it must act competitively, i.e., take prices as given. In particular, the bank takes as given the retrade market interest rate R.

Given this difference, it is not hard to see that the optimal allocation (c_1^*, c_2^*) cannot be an equilibrium allocation. If (c_1^*, c_2^*) were to be an equilibrium allocation, the interest rate R in the hidden retrade market would have to be equal to the shadow rate R^* given in (17), for otherwise agents would use that market to trade away from this allocation. But R^* cannot be an equilibrium retrade interest rate because the fact that R^* is strictly smaller than $\hat{R}$ creates an arbitrage opportunity. This arbitrage opportunity is similar to the one that in the incomplete-markets model discussed in the previous section pinned down the secondary-market asset price p at 1.

The arbitrage strategy, described in Jacklin (1987), calls for investment $x = e$ at date 0. If the agent executing this arbitrage is patient, i.e., his $\theta = 1$, he consumes nothing at date 1 and $\hat{R}e > c_2^*$ at date 2. If he turns out impatient, i.e., his $\theta = 0$, he can access the retrade market and borrow at rate R^*, which gives him date-1 consumption $\frac{\hat{R}e}{R^*} > c_1^*$. In either case, thus, he consumes more than (c_1^*, c_2^*), which shows that (c_1^*, c_2^*), with its shadow interest rate R^*, cannot be an equilibrium allocation of consumption.

What allocation can be a market equilibrium allocation in this model? The Jacklin arbitrage strategy pins down the interest rate in the retrade market at $R = \hat{R}$. With this interest rate, it is easy to check (or consult Allen and Gale [2004] or Farhi, Golosov, and Tsyvinski [2009]) that the equilibrium allocation (19) from the incomplete-markets model discussed in the previous section is a unique equilibrium allocation, also here in the richer model with fully state-contingent contracts.[14]

Why is the planner able to do better than the market in this model? The planner makes the Jacklin arbitrage strategy infeasible for the

[14] This conclusion applies to all conceivable market structures in which the Jacklin arbitrage strategy remains feasible. In particular, when the hidden retrade market is included, it applies to the general competitive private information model of Prescott and Townsend (1984) in which agents trade lotteries over allocations subject to incentive compatibility constraints.

agent by controlling initial investment (s, x). In the planning problem, although the agent has unfettered access to the retrade market, the agent does not have private control over his initial investment. The initial investment choice is publicly observable and therefore can be controlled by the planner/government. The Jacklin arbitrage strategy calls for the all-long investment $(s, x) = (0, e)$ at date 0. By forcing/choosing investment $(s, x) = (s^*, e - s^*)$, the planner eliminates this arbitrage. Moreover, this choice of date-0 investment pins down the amount of resources available at dates 1 and 2 and, thus, also the interest rate in the hidden retrade market, which with liquid investment s^* is $R = R^*$. In a competitive market economy, by contrast, firms have to respect the agents' freedom to not contract with them but instead to invest directly (or set up another firm that will do the investing for them, as in Farhi, Golosov, and Tsyvinski [2009]). Intermediaries thus cannot make the Jacklin arbitrage strategy infeasible for the agents. Having to respect this arbitrage condition, the best allocation they can provide is $(\hat{c}_1, \hat{c}_2) = (e, \hat{R}e)$ with the associated retrade market interest rate $R = \hat{R}$.

To recap, the planner internalizes the fact that her control of the initial investment changes the price in the equilibrium of the retrade market. Firms, in contrast, take all prices as given, including those in the retrade market. The discrepancy constitutes a pecuniary externality in this model and the equilibrium allocation is inefficient.

Efficiency Without Retrade

The Jacklin arbitrage strategy is clearly impossible to execute if arbitrageurs do not have access to the hidden retrade market. Absent retrade, competitive equilibrium with state-contingent contracts would be efficient. Indeed, if the retrade market is shut down, the value function $\tilde{V}(c, R; \theta)$ defined in (6) reduces to $\tilde{V}(c, R; \theta) = \max_{\tilde{\theta}} u(c_1(\tilde{\theta}) + \theta c_2(\tilde{\theta}))$, which no longer depends on R. The incentive constraint (7), therefore, no longer depends on a price.[15] This means that there is no pecuniary externality. The welfare theorems of Prescott and Townsend (1984) apply, and competitive equilibrium is efficient. In particular, it can be implemented as a banking equilibrium of Diamond and Dybvig (1983) with the equilibrium deposit contract $(\xi_1, \xi_2) = (c_1^*, c_2^*)$.

The theoretical results we reviewed in this section suggest that retrade generates a pecuniary externality and leads to equilibrium

[15] In particular, given Lemma 1, the impatient types will never misrepresent their type and the patient types' incentive constraint reduces to $c_2 \geq c_1$.

underprovision of liquidity. In practice, banks and other financial intermediaries have ample access to various retrade markets. Therefore, one might be tempted to take as an implication of this theory the prediction that markets will fail to provide sufficient liquidity. In the next section, we present a simple version of the analysis of Kilenthong and Townsend (2011) showing that this conclusion would be premature: If harnessed inside appropriate venues, retrade can be consistent with efficient functioning of markets in the provision of liquidity.

5. COMPETITIVE EQUILIBRIUM WITH SEGREGATED EXCHANGES

In this section, we consider the model of Kilenthong and Townsend (2011), in which a market-maker eliminates the Jacklin arbitrage by segmenting the retrade market and pricing entry into market segments as a function of the investment portfolio held by agents entering a given segment. With the Jacklin arbitrage eliminated, the pecuniary externality causing market failure is eliminated as well. The resulting equilibrium is efficient. We supplement the analysis of Kilenthong and Townsend (2011) by characterizing explicitly how the equilibrium exchange entry fees depend on the fundamentals of the exchange and on the portfolio of the agent (equation [27] and Figure 2). We conclude with a discussion of an important difference between the environment with pecuniary externality studied in the previous sections and the environment without it that we study here. The segregated-exchanges equilibrium is efficient, but, effectively, it requires that agents commit ex ante to not using the hidden retrade market ex post. Whether or not retrade leads to a pecuniary externality and inefficiency of market outcomes, therefore, depends on the practical feasibility of such a commitment.

Trade Inside Segregated Exchanges at Date 1

Before we define the general equilibrium concept with segregated exchanges proposed by KT, we describe in this subsection segregated exchanges, their fundamentals, and internal prices.

A segregated exchange is a competitive market for the long-term asset that opens at date 1 after types θ are realized. A defining characteristic of such an exchange is a set of fundamentals determining the market price p at which the long-term assets will be traded. The fundamentals and the price must be consistent: Given the fundamentals in an exchange, the price p must indeed be a competitive equilibrium price in that exchange. In the DD economy at hand, the level of the cash asset investment s held by each member of an exchange is a sufficient description of the fundamentals in the exchange. Thus, we will index exchanges by $S \in [0, e]$, where S represents the level of liquid investment held by each agent entering the exchange. Note that this definition assumes identical asset holdings by all exchange members. We will see later that this assumption is without loss of generality in the present environment.

Equilibrium price in exchange S

Let us derive an equilibrium consistency condition between fundamentals S and price p in the exchange $S \in [0, e]$. It is a simple equilibrium pricing condition in a competitive market with all agents holding the same portfolio of assets $(s, x) = (S, e - S)$ and experiencing shocks θ drawn from the same distribution. We will denote the equilibrium price in exchange S by $p(S)$.

The equilibrium condition for consistency between S and p is

$$p(S) = \min \left\{ \frac{(1-\pi)\, S}{\pi\, (e - S)}, \hat{R} \right\}. \tag{20}$$

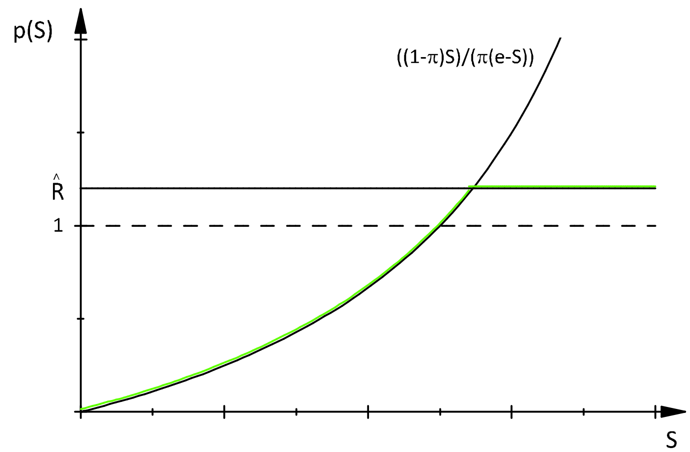
This condition is derived as follows. The equilibrium price of the illiquid asset is determined by supply and demand in exchange S in the same way as it was determined in the incomplete-markets model of Section 3. At date 1, the impatient agents want to sell their long-term asset in the market at any price. They supply $\pi(e - S)$ units of the long-term asset to the market. The behavior of the patient agents depends on the price p. If $p > \hat{R}$, a short position in the asset gives them a positive return, so patient agents want to sell their holdings of the asset, just like the impatient ones. This cannot be an equilibrium, as demand for the asset is zero and supply is positive. Thus, in any equilibrium, $p \leq \hat{R}$. With $p \leq \hat{R}$, a long position in the asset gives patient agents a non-negative return (strictly positive if $p < \hat{R}$). With any such price, the patient agents are willing to buy the long-term asset. They demand $(1 - \pi)\frac{S}{p}$ units. Thus, the equilibrium price $p(S)$ solves $\pi(e - S) = (1 - \pi)\frac{S}{p(S)}$, which gives us

$$p(S) = \frac{(1-\pi)\, S}{\pi\, (e - S)}, \tag{21}$$

provided that $p(S) \leq \hat{R}$. Solving $\hat{R} = \frac{(1-\pi)S}{\pi(e-S)}$ for S, we get a threshold

$$\bar{S} = \frac{\pi \hat{R} e}{\pi \hat{R} + 1 - \pi}. \tag{22}$$

Figure 1 Equilibrium Asset Price p in Exchange S



For all $S \geq \bar{S}$, the equilibrium price is flat at $\hat{R}$.[16] Combining this restriction with (21) gives us the consistency condition (20).

Figure 1 illustrates the derivation of the consistency condition (20) graphically. When S is small and $e - S$ is large, there is a large quantity of the illiquid asset in the market, supplied by the impatient agents, and very few units of the consumption good (cash), supplied by the patient agents, and so the price of the asset is low.[17] In exchanges with higher S, the proportion of cash to units of the asset in the market is higher, so the price $p(S)$ is higher. This is true up to the threshold $\bar{S}$. In exchanges with S larger than $\bar{S}$, the price $p(S)$ remains flat at $\hat{R}$ and the patient types are indifferent between buying and selling the asset. The price of the asset cannot exceed $\hat{R}$, as at a price higher than $\hat{R}$ the patient agents would switch from buying to selling the asset. As we see, the range of prices that can be consistent with some fundamentals $S \in (0, e]$ is

$$0 < p \leq \hat{R}. \tag{23}$$

[16] Note that $\bar{S}$ is the same threshold that in Proposition 1 results with the full-insurance allocation (an upper bound on s^*).

[17] We will exclude the exchange $S = 0$ from our analysis. In this exchange, the supply of resources at date 1 would be zero and thus welfare of the impatient agents would be extremely low. No agent would want to enter this exchange at date 0.

Markets at Date 0 and Equilibrium Definition

In this subsection, we use the segregated exchanges to define the KT notion of competitive equilibrium with segregated retrade.

At date 0, agents choose their investments s and x and join segregated exchanges. Each agent can physically join one exchange. Exchanges are defined by their fundamental level of liquid investment S. Associated with each exchange is an entry fee pricing any deviations of the investment portfolio of an agent wishing to join a given exchange from that exchange's fundamentals. If an agent joins an exchange S with liquid investment s, the amount of shortage of his liquid asset relative to the exchange fundamentals is $S - s$. Upon entry, the agent is charged a fee proportional to the amount of shortage of liquid investment in his portfolio. The price per unit of shortage in exchange S is $\delta(S)$. Thus, an agent entering exchange S with liquid investment s is charged an entry fee of $\delta(S)(S - s)$. This charge is assessed by the exchange as of the time of entry, i.e., at date 0. The unit price $\delta(S)$ can be positive or negative. Note that if $\delta(S) > 0$ and an agent joins exchange S with liquid investment $s > S$, the entry fee is negative, so the exchange makes a payment to the agent.

In sum, at date 0 agents choose investment portfolios (s, x) and exchange membership S subject to the budget constraint

$$s + x + \delta(S)(S - s) \leq e. \tag{24}$$

If, for example, an agent decides to join exchange S and go all-long, i.e., invest $s = 0$ and $x = e$, then the price for this shortage would be $\delta(S)S$. Clearly, public observability of the agent's portfolio is important for the assessment of fees. In particular, agents cannot avoid fees by "window dressing" or changing the composition of their portfolio after the fees are assessed but before the shock θ is realized and exchanges open for business.

What if an agent chooses not to join an exchange? The decision not to join is equivalent to joining an exchange in which the price of any "deviation" or "shortage" relative to the "fundamentals" is zero. Thus, not joining a segregated exchange is equivalent to maintaining access to the free exchange in which $\delta = 0$. As we will see shortly, the exchange $S = \pi e$ will have $\delta = 0$. This exchange corresponds to the incomplete-markets model of Section 3, where, as we saw earlier, all agents choose investment $s = \pi e$ at date 0. It is natural to default all agents who do not join a different exchange into this one. The model with segregated exchanges, therefore, nests the simple incomplete-markets model as a special case in which there is only one secondary market for the long-term asset, and access to this market is free.

Let us now discuss the agents' objective function as of date 0. Agents maximize

$$E[V_1(s, x, S; \theta)], \tag{25}$$

where $V_1(s, x, S; \theta)$ is the indirect utility function as of date 1, i.e., the value the agent can get in exchange S with an asset portfolio (s, x) and a liquidity shock realization θ. The indirect utility function

$$\begin{aligned} V(s, x, S; \theta) &= \max\ u(c_1 + \theta c_2), \\ & \quad s.t. \\ & \quad c_1 + p(S)n \leq s, \\ & \quad n \geq -x, \\ & \quad c_2 \leq (x + n)\hat{R}, \end{aligned} \tag{26}$$

where n is the agent's net demand at date 1 in the market for the illiquid asset inside exchange S.

Next, we define competitive equilibrium with segregated exchanges.

Definition 1 *(Kilenthong and Townsend) A price system* $(p(\cdot), \delta(\cdot))$*, ex ante investment and exchange membership choices* s*,* x*,* S*, value functions* $V_1(\cdot, \cdot, \cdot; \theta)$ *for* $\theta \in \{0, 1\}$*, and a consumption allocation* (c_1, c_2) *are an equilibrium with segregated exchanges if*

1. *expectations are correct: For each* S*, price* $p(S)$ *satisfies the consistency condition (20) and value functions* $V_1(\cdot, \cdot, \cdot; \theta)$ *solve (26);*

2. *agents optimize ex ante: Taking prices* $(\delta(\cdot), p(\cdot))$ *and value functions* $V_1(\cdot, \cdot, \cdot; \theta)$ *as given, agents' choices* s*,* x*,* S *maximize their ex ante utility (25) subject to the budget constraint (24);*

3. *market clearing: Consumption allocation* (c_1, c_2) *is an equilibrium allocation of consumption in the exchange* S*.*

Note that this definition does not allow for mixed strategies. In general, mixed strategies may be useful, as agents face a discrete choice of exchange membership. As the theorem presented next makes clear, in the environment at hand it is without loss of generality to restrict attention to equilibria in pure strategies, where all agents, being ex ante identical, join the same exchange.[18]

[18] In excluding random exchange assignments, this definition follows Definition 4 in Kilenthong and Townsend (2014a).

Efficient Equilibrium with Segregated Exchanges

Theorem 1 *Prices $p(S)$ as in (20) and*

$$\delta(S) = \min\left\{1 - \frac{\pi}{1-\pi}\left(\frac{e}{S} - 1\right), 1 - \hat{R}^{-1}\right\}, \tag{27}$$

ex ante investment and membership choices $s = s^*$, $x = e - s^*$, $S = s^*$, and consumption allocation $(c_1, c_2) = (c_1^*, c_2^*)$ are a competitive equilibrium with segregated exchanges.

The rest of this subsection is devoted to proving this theorem. We need to check the three equilibrium conditions in Definition 1.

We start by characterizing value functions (26). For $\theta = 0$, the optimized value of (26) is

$$V_1(s, x, S; 0) = u\left(s + p(S)x\right). \tag{28}$$

Clearly, the impatient agents want to sell their holdings x of the long-term asset at any price $p(S)$ and consume all their wealth at date 1, as they have no use for consumption at date 2. At price $p(S)$, an impatient agent can afford consumption $c_1 = s + p(S)x$, which gives us (28).

The patient type's value as of date 1 is

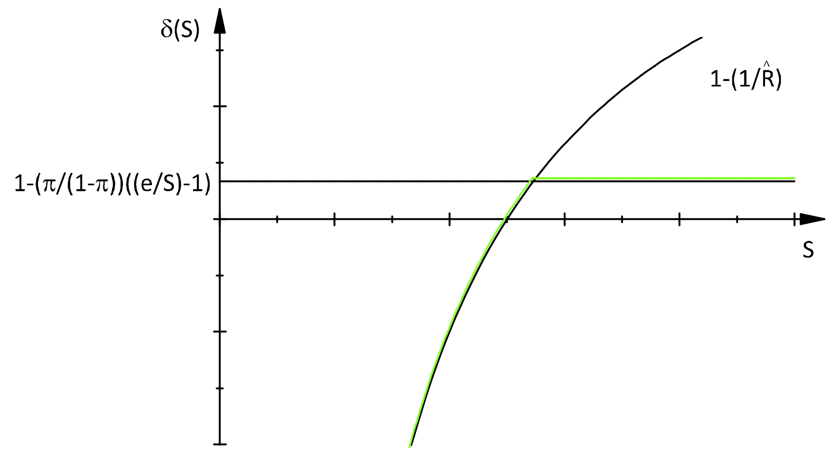
$$V_1(s, x, S; 1) = u\left(\left(x + \frac{s}{p(S)}\right)\hat{R}\right). \tag{29}$$

To see that this is the case, note that in each exchange S patient agents are happy to buy the long-term asset at date 1 because, by (23), $p(S) \leq \hat{R}$ in all exchanges S. This means that the rate of return on this investment, $\frac{\hat{R}}{p(S)}$, exceeds the patient type's rate of time preference, which is 1. A patient agent's demand for the long-term asset is $n = \frac{s}{p(S)}$, his consumption at date 1 is $c_1 = 0$, and consumption at date 2 is $c_2 = (x + \frac{s}{p(S)})\hat{R}$. These quantities substituted to (26) with $\theta = 1$ give us (29).

We can now confirm that with value functions (28) and (29) the first equilibrium condition (correct expectations) is satisfied, as these value functions and prices $p(S)$ defined in (20) are consistent with agents' optimization at date 1. Note that the general pattern of behavior at date 1 is the same in all exchanges. The impatient types sell and the patient types buy the long-term asset. The exchanges are different only in the composition of demand and supply, which gives rise to different equilibrium prices at which the asset is traded in each exchange.

In order to check the second equilibrium condition (agents' optimization ex ante), we now study the agents' behavior at date 0. Substituting the indirect utility functions (28) and (29) into the objective (25), we express the ex ante expected utility function of the

Figure 2 Unit Liquidity Shortage Price δ in Exchange S



representative agent as

$$\pi u\left(s+p(S)x\right)+(1-\pi)u\left(\left(x+\frac{s}{p(S)}\right)\hat{R}\right).$$

This expression gives the agent's expected value of being in exchange S with assets s and x. The representative agent chooses investment portfolio (s,x) and exchange membership S to maximize this value subject to the date-0 budget constraint (24).

The structure of the portfolio fees $\delta(S)$ charged upon exchange entry is a key part of the budget constraint. Figure 2 graphs against S the unit liquid asset shortage price $\delta(S)$ given in (27). As we argue, these prices support the efficient equilibrium.

It is easy to check directly in (27) that $1-\frac{\pi}{1-\pi}\left(\frac{e}{S}-1\right)<1-\hat{R}^{-1}$ for all $S<\bar{S}$, where $\bar{S}$ is, as before, given in (22). Thus,

$$\delta(S)=\begin{cases}1-\frac{\pi}{1-\pi}\left(\frac{e}{S}-1\right) & \text{for} \quad S\leq\bar{S}, \\ 1-\frac{1}{\hat{R}} & \text{for} \quad S\geq\bar{S}.\end{cases}$$

Note that $\delta(S)$ is increasing. This means that the portfolio charge per unit of liquidity shortage is higher in exchanges with higher fundamental liquidity S. Substituting in (27) $S=\pi e<\bar{S}$, we check that $\delta(\pi e)=0$. Thus, the exchange with $S=\pi e$ is a (unique) free-entry exchange, where portfolio charges are zero for all portfolios (s,x). In exchanges with $S>\pi e$, $\delta(S)>0$, i.e., agents are subject to a positive charge for shortage of liquidity in their portfolio. For all $S<\pi e$,

$\delta(S) < 0$, i.e., portfolio charges are positive if the long-term investment x is less than $e - S$.

We now can study the agents' date-0 problem of choice of investment (s, x) and exchange membership S. For each exchange S, we need to determine the investment portfolio (s, x) the agent will choose conditional on joining S and the consumption pair (c_1, c_2) he will be able to afford inside S. This will give us the ex ante expected value of joining S, which we will then use to determine the agent's most preferred exchange membership decision and thus the solution to his utility maximization problem.

We start by examining the exchanges with $S \geq \bar{S}$. What ex ante value can the representative agent obtain if he plans on joining one of these exchanges? All exchanges $S \geq \bar{S}$ have the same entry fees and long-term asset prices:

$$\begin{aligned} \delta(S) &= 1 - \hat{R}^{-1}, \\ p(S) &= \hat{R}. \end{aligned} \tag{30}$$

Given a portfolio (s, x), an agent in exchange $S \geq \bar{S}$ can afford consumption

$$c_1 = s + \hat{R}x$$

if impatient, or

$$c_2 = \left(x + \frac{s}{\hat{R}}\right)\hat{R} = s + \hat{R}x$$

if patient. As we see, the agent is fully insured against the liquidity shock θ in any exchange $S \geq \bar{S}$, as his optimal consumption in any such exchange is independent of the realization of the liquidity shock θ. His ex ante expected utility is therefore simply

$$u\left(s + \hat{R}x\right). \tag{31}$$

With the entry fee of $\delta(S) = 1 - \hat{R}^{-1}$ per unit of liquidity shortage, the agent's ex ante budget constraint (24) can be written as

$$s + \hat{R}x \leq \hat{R}e - (\hat{R} - 1)S. \tag{32}$$

Comparing the agent's objective (31) and his budget constraint (32), we see that the agent is indifferent between all portfolios (s, x) on the budget line $s + \hat{R}x = \hat{R}e - (\hat{R} - 1)S$. This is because any such portfolio gives the agent the same ex ante utility of $u\left(\hat{R}e - (\hat{R} - 1)S\right)$. Since $\hat{R} > 1$, this value is decreasing in S. Thus, among all exchanges $S \geq \bar{S}$, exchange $\bar{S}$ is the best one for the agent.

Next, let us consider the choices of an agent who plans on joining one of the exchanges with $S \leq \bar{S}$. The prices this agent faces are

$$\begin{aligned} \delta(S) &= 1 - \frac{\pi}{1-\pi}\left(\frac{e}{S} - 1\right), \\ p(S) &= \frac{(1-\pi)\, S}{\pi\,(e-S)}. \end{aligned} \tag{33}$$

Thus, given a portfolio (s, x), in exchange S the agent can afford consumption

$$c_1 = s + \frac{(1-\pi)\, S}{\pi\,(e-S)} x$$

if impatient, or

$$c_2 = \left(x + \frac{s}{\frac{(1-\pi)S}{\pi(e-S)}} \right) \hat{R} = \left(s + \frac{(1-\pi)\, S}{\pi\,(e-S)} x \right) \frac{\pi\,(e-S)}{(1-\pi)\, S} \hat{R}$$

if patient. Unlike in the previous case, these consumptions are not identical. They are, however, directly proportional to $s + \frac{(1-\pi)S}{\pi(e-S)}x$. Substituting these consumption values into the ex ante expected utility function, we have

$$\pi u\left(s + \frac{(1-\pi)\, S}{\pi\,(e-S)} x \right) + (1-\pi) u\left(\left(s + \frac{(1-\pi)\, S}{\pi\,(e-S)} x \right) \frac{\pi\,(e-S)}{(1-\pi)\, S} \hat{R} \right). \tag{34}$$

With the entry fee $\delta(S)$ given in (33), the agent's ex ante budget constraint (24) can be rewritten, after some algebra, as

$$s + \frac{(1-\pi)\, S}{\pi\,(e-S)} x \leq \frac{S}{\pi}.$$

Comparing this budget constraint and the agent's objective (34) we see that here, as in the previous case, the agent is indifferent between all portfolios (s, x) on the budget line $s + \frac{(1-\pi)S}{\pi(e-S)}x = \frac{S}{\pi}$ as any such portfolio gives him the same expected utility value of

$$\pi u\left(\frac{S}{\pi}\right) + (1-\pi) u\left(\frac{e-S}{1-\pi}\hat{R}\right).$$

Finally, we observe that this objective function, representing the agent's utility from joining exchange S, is mathematically the same as the objective function (10) in the social welfare maximization problem studied in Proposition 1. As we saw there, this objective is maximized by a unique $s^* < \bar{S}$. Thus, exchange $S = s^*$ is a unique maximizer in the agent's utility maximization problem we study here.[19] To simplify the notation, we will use S^* to denote the exchange $S = s^*$.

The last equilibrium condition that we need to check is to confirm that (c_1^*, c_2^*) is an equilibrium allocation of consumption in exchange S^* with the asset price $p(S^*)$. For the pair (c_1^*, c_2^*) to be resource-feasible in exchange S^*, agents must enter this exchange carrying the investment portfolio $(s^*, e - s^*)$. Portfolio $(s^*, e - s^*)$ is (weakly) optimal for an agent joining exchange S^* because, as we saw earlier, conditional on joining an exchange, agents are indifferent among all portfolios (s, x) on the budget line. Finally, since the asset price $p(S^*)$ satisfies the consistency condition (20), the market for the long-term asset inside the exchange S^* does clear.

We conclude that the prices and quantities specified in Theorem 1 are indeed a competitive equilibrium with segregated exchanges. This equilibrium is efficient, as the equilibrium consumption bundle is exactly the optimal consumption bundle (c_1^*, c_2^*).

Discussion

In the two equilibrium concepts without segregated exchanges that we discussed in Sections 3 and 4, arbitrage pinned at $p = 1$ the equilibrium price in the secondary market for the long-term asset or, equivalently, the retrade market interest rate at $R = \hat{R}$. In the model with segregated exchanges, agents trade the long-term asset in the secondary market inside the exchange S^* at the equilibrium price

$$p(S^*) = \frac{(1-\pi)\, s^*}{\pi\, (e - s^*)} = \frac{(1-\pi)\, \pi c_1^*}{\pi\, (1-\pi)\, \frac{c_2^*}{\hat{R}}} = \hat{R}\frac{c_1^*}{c_2^*} = \frac{\hat{R}}{R^*} > 1.$$

Why does arbitrage not force $p(S^*)$ down to 1 in the segregated exchanges model?

The Jacklin arbitrage strategy is infeasible in the segregated exchange model because of the entry fees ex ante and the separation of agents in different exchanges ex post. The Jacklin arbitrage strategy calls for the all-long initial investment $(s, x) = (0, e)$ and a subsequent sale of the long-term asset, or borrowing against it, in case the agent attempting arbitrage turns out needing funds at date 1. But which exchange should the arbitrageur join at date 0? If he defaults to the entry-fee-free market $S = \pi e$, he does not receive the favorable asset price $p(S^*) > 1$ but only the arbitrage-free price $p(\pi e) = 1$, so no arbitrage profit can be made in this exchange. If the arbitrageur joins

[19] Note in particular that the right inequality in (11) implies that exchange $S = s^*$dominates the exchange $S = \bar{S}$ and thus also all exchanges $S \geq \bar{S}$.

exchange S^*, he must pay the entry fee of $\delta(S^*)S^*$. This fee offsets exactly the profit he makes selling the long-term asset at the high price $p(S^*)$, thus eliminating the overall profitability of this attempt at arbitrage. The entry fee offsets exactly the asset sale profit because, conditional on joining an exchange, agents are indifferent between all feasible portfolio choices. In particular, the arbitrageur joining exchange S^* with the all-long portfolio $(0, e)$ does no better than an agent entering this exchange with the equilibrium portfolio $(s^*, e-s^*)$. Similarly, if the arbitrageur with portfolio $(0, e)$ joins any other exchange S, he is exactly as well off as an agent joining S with the fundamentals-consistent portfolio $(S, e-S)$. Thus, the arbitrageur joining S obtains the ex ante expected utility value of $W(S)$. As we saw in Proposition 1, this value is maximized at $S = S^*$. No arbitrage attempt therefore can be successful.

The agents' ability to commit to not trading across exchanges ex post is key in eliminating the Jacklin arbitrage. The segregated exchanges mechanism lets each agent join only one exchange. In addition, it requires that agents sign off their right to trade freely with the counterparty of their choice. Instead, it requires that agents commit to trading only with other members of the exchange they belong to. If agents do not have the ability to contractually give away their freedom to trade without counterparty restrictions, an impatient arbitrageur residing in the entry-fee-free exchange $S = \pi e$ can easily convince a patient agent in exchange S^* to buy the long-term asset from him rather than in exchange S^* because he can sell for less than $p(S^*)$ and still make a profit. As agents anticipate this at date 0, price expectations embedded in $p(S)$ are not credible and the equilibrium breaks down. Thus, the restriction of participation to one exchange only and the assumption of the agents' ability to commit to not step out of their exchanges ex post are crucial.

In the KT equilibrium, segregated exchanges can therefore be thought of as a commitment device allowing the agents to promise credibly to not access the hidden IOU market. Clearly, if in the KT model agents could access the hidden IOU retrade market *after* they trade in segregated exchanges, the equilibrium with segregated exchanges supporting the optimal asset price $p(S^*)$ would collapse. The argument for it is the same as in Section 4. The optimal allocation (c_1^*, c_2^*) is consistent with free access to the retrade market only if the interest rate in this market equals $R^* = c_2^*/c_1^*$. But with this interest rate, the Jacklin arbitrage can again be executed by investing all long, joining the entry-fee-free exchange $S = \pi e$, and not trading in this exchange but rather borrowing in the IOU market if liquidity is needed at date 1.[20]

In the banking model discussed in Section 4, the intermediary designing the state-contingent deposit contract cannot put any restrictions on retrade between depositors and non-depositors. The market-making firm in the segregated exchanges model, in contrast, can. In particular, an agent who did not join exchange S and subject his portfolio to the entry fee $\delta(S)$ cannot retrade with agents who did join exchange S. This additional power given to the market-maker in the segregated exchanges model makes her equally as effective as the social planner in Section 4 in controlling agents' investment at date 0. Unlike the planner, the market-maker does not control this investment directly but rather sets up prices (i.e., exchange entry fees) to induce efficient investment.

As we see, the model with segregated exchanges, where retrade does not lead to a pecuniary externality, requires a different economic environment than the models in Sections 3 and 4, where access to hidden retrade causes an externality. The segregated exchanges model requires that agents have the ability to commit themselves to refrain from trading in the *hidden* retrade market, which effectively makes this model equivalent to the model with observable trades that we discussed in Section 4.[21] If such commitment can be made credible, e.g., by physically separating agents ex post, then all agents would choose to extend it ex ante. If, however, it is a feature of the environment that such a commitment cannot be made credible, as in Farhi, Golosov, and Tsyvinski (2009), access to hidden retrade makes the Jacklin arbitrage strategy feasible, the pecuniary externality exists, and markets fail to provide sufficient liquidity in equilibrium.

Clearly, the cap-and-trade mechanism will not be successful at limiting greenhouse gas emissions if firms can emit completely privately/anonymously, i.e., without anyone observing it. If they can, the price of the right to emit one tonne of CO_2 will be zero. In the KT model, retrade is analogous to observable emissions that can be priced. In the pecuniary externality model, hidden retrade is analogous to anonymous emissions that cannot be priced or internalized with a cap-and-trade scheme.

Are then segregated exchanges a solution to the pecuniary externality problem caused by retrade? Segregated exchanges do not solve

[20] Better yet, the arbitrageur could join one of the exchanges $S < \pi e$, where $\delta(S) < 0$, which means with $s = 0$ he would get a payment from the exchange upon entry.

[21] That the segregated exchanges model requires a different environment than the unfettered hidden retrade model is clear from Table 1 on page 1,046 in Kilenthong and Townsend (2011).

the pecuniary externality problem, but they show that retrade does not have to lead to one. The literature on pecuniary externalities with complete markets and retrade assumes that agents have unfettered access to an anonymous, hidden retrade market and cannot do anything to make credible an ex ante promise to refrain from accessing this market ex post. The segregated exchanges model assumes that such a commitment is possible. The segregated exchanges model, therefore, does not solve the pecuniary externality problem associated with anonymous, hidden retrade. Instead, it points out that retrade by itself does not imply the existence of a pecuniary externality. The model shows that retrade can be accounted for within the competitive market framework without violating efficiency, provided that a sufficiently rich market structure, including markets for exchange membership, is allowed for.

In addition, the KT model shows that exclusivity and ex post trade restrictions can be socially valuable. Their role can be to serve as a commitment device that agents may be able to use to help them refrain from the "harmful," hidden retrade activity and still be able to engage in efficient, priced retrade.

6. CONCLUSION

The literature we review makes it clear that in the Diamond-Dybvig economy, the agents' access to retrade is key in understanding whether markets are efficient or require government intervention. The theory makes a distinction between two kinds of retrade: the "priced" kind and the "hidden" kind. Hidden, anonymous retrade leads to a pecuniary externality and market failure. Priced retrade, harnessed into access-controlled segregated exchanges with exchange- and portfolio-dependent entry fees does not cause market failure.

The observation of retrade itself in present-day financial markets does not therefore imply that markets are inefficient or efficient in providing liquidity. To answer the question of efficiency, one must assess which of the two kinds of retrade discussed in the model is a better reflection of reality. Kilenthong and Townsend (2014a) suggest that the assumption of restricted retrade is a good one in financial markets. In other applications, for example in the problem studied in Kehoe and Levine (1993) where pecuniary externalities result from workers' unrestricted access to spot labor markets, this assumption may be more problematic, as firms may lack the commitment to deny employment to workers who have defaulted on some financial obligations in the past. Given these theoretical predictions and their implications for the efficacy of government intervention, empirical research identifying the nature of retrade and the existence or nonexistence of pecuniary externalities is needed.

APPENDIX

Proof of Lemma 1

Suppose allocation c is optimal with $c_2(0) > 0$ and define an allocation $\hat{c} = \{\hat{c}_1(0), \hat{c}_2(0), \hat{c}_1(1), \hat{c}_2(1)\}$ as follows:

$$\begin{array}{ll} \hat{c}_1(0) = c_1(0), & \hat{c}_1(1) = c_1(1), \\ \hat{c}_2(0) = 0, & \hat{c}_2(1) = c_2(1) + \frac{\pi}{1-\pi} c_2(0). \end{array}$$

Allocation $\hat{c}$ is feasible because at each date $t = 1, 2$ it uses the same amount of resources as allocation c. Indeed:

$$\begin{aligned} & \pi\left(\hat{c}_1(0) + \frac{\hat{c}_2(0)}{\hat{R}}\right) + (1-\pi)\left(\hat{c}_1(1) + \frac{\hat{c}_2(1)}{\hat{R}}\right) \\ = \; & \pi c_1(0) + (1-\pi)\frac{c_2(1) + \frac{\pi}{1-\pi}c_2(0)}{\hat{R}} \\ = \; & \pi\left(c_1(0) + \frac{c_2(0)}{\hat{R}}\right) + (1-\pi)\left(c_1(1) + \frac{c_2(1)}{\hat{R}}\right) \\ \leq \; & e. \end{aligned}$$

Allocation $\hat{c}$, however, attains a higher value of the objective (5) because it provides the same utility $u(c_1(0))$ to the impatient type and a higher utility $u(c_1(1) + c_2(1) + \frac{\pi}{1-\pi}c_2(0)) > u(c_1(1) + c_2(1))$ to the patient type. This contradicts the supposed optimality of c.

To prove that $c_1(1) = 0$, suppose that c is optimal with $c_1(1) > 0$ and define an allocation $\hat{c} = \{\hat{c}_1(0), \hat{c}_2(0), \hat{c}_1(1), \hat{c}_2(1)\}$ as follows:

$$\begin{array}{ll} \hat{c}_1(0) = c_1(0), & \hat{c}_1(1) = 0, \\ \hat{c}_2(0) = c_2(0), & \hat{c}_2(1) = c_2(1) + \hat{R}c_1(1). \end{array}$$

Allocation $\hat{c}$ is feasible because it costs the same in present value terms as the feasible allocation c. Indeed:

$$\begin{aligned} & \pi\left(\hat{c}_1(0) + \frac{\hat{c}_2(0)}{\hat{R}}\right) + (1-\pi)\left(\hat{c}_1(1) + \frac{\hat{c}_2(1)}{\hat{R}}\right) \\ = \; & \pi\left(c_1(0) + \frac{c_2(0)}{\hat{R}}\right) + (1-\pi)\frac{c_2(1) + \hat{R}c_1(1)}{\hat{R}} \\ = \; & \pi\left(c_1(0) + \frac{c_2(0)}{\hat{R}}\right) + (1-\pi)\left(c_1(1) + \frac{c_2(1)}{\hat{R}}\right) \\ \leq \; & e. \end{aligned}$$

Table 1 Allocation Mechanisms, Frictions, and Outcomes

Allocation Mechanism	Private Shocks θ	Hidden IOU Retrade Market	Market Structure	Allocation	Page
First-best planning problem	Absent	Absent	No markets–planner chooses allocation	(c_1^*, c_2^*)	313
Planning problem with private θ	Present	Absent	No markets–planner chooses allocation	(c_1^*, c_2^*)	321
Planning problem with private θ and hidden retrade	Present	Present	No markets–planner chooses allocation	(c_1^*, c_2^*)	315
Incomplete markets model	Present	Present	Ex ante: direct investment (s, x) Ex post: market for the long-term asset and the hidden IOU market	$(\hat{c}_1, \hat{c}_2)$	316
Market model with state-contingent contracts, no hidden IOU retrade markets	Present	Absent	Ex ante: deposit contract (ξ_1, ξ_2) with price e Ex post: no markets	(c_1^*, c_2^*)	321
Market model with state-contingent contracts and hidden IOU retrade markets	Present	Present	Ex ante: deposit contract (ξ_1, ξ_2) with price e Ex post: hidden IOU retrade market	$(\hat{c}_1, \hat{c}_2)$	319
Market model with segregated exchanges	Present	Absent	Ex ante: continuum of segregated exchanges Ex post: retrade inside exchanges, but not across	(c_1^*, c_2^*)	322

Allocation $\hat{c}$, however, attains a higher value of the objective (5) than c, because it provides the same utility $u(c_1(0))$ to the impatient type and a higher utility $u(c_2(1) + \hat{R}c_1(1)) > u(c_1(1) + c_2(1))$ to the patient type. This contradicts the supposed optimality of c. QED

Proof of Proposition 1

Since $W''(s) = \frac{1}{\pi}u''\left(\frac{s}{\pi}\right) + \frac{1}{1-\pi}\hat{R}^2 u''\left(\frac{e-s}{1-\pi}\hat{R}\right) < 0$, we have that $W'(s) = u'\left(\frac{s}{\pi}\right) - \hat{R}u'\left(\frac{e-s}{1-\pi}\hat{R}\right)$ is continuous and strictly decreasing. The existence of a unique solution to $W'(s) = 0$ in $(0, e)$ thus follows from the fact that $\lim_{s\to 0} W'(s) \to \infty$ and $\lim_{s\to e} W'(s) \to \infty$.

For the two bounds on s^*, it is sufficient to show that $W'(\pi e) > 0$ and $W'\left(\pi e \frac{\hat{R}}{\pi\hat{R}+1-\pi}\right) < 0$. We first note that relative risk aversion everywhere strictly greater than one implies that the function $f(\alpha) = \alpha u'(\alpha e)$ is strictly decreasing. Indeed, with $f'(\alpha) = u'(\alpha e) + \alpha e u''(\alpha e)$ we have that $f'(\alpha) < 0$ follows from $1 < \frac{-\alpha e u''(\alpha e)}{u'(\alpha e)}$. Now, evaluating W' at $s = \pi e$, we have $W'(\pi e) = u'(e) - \hat{R}u'(e\hat{R}) = f(1) - f(\hat{R}) > 0$, where the strict inequality follows from f strictly decreasing and $\hat{R} > 1$. To show $W'\left(\pi e \frac{\hat{R}}{\pi\hat{R}+1-\pi}\right) < 0$, note that with $s = \pi \frac{\hat{R}e}{\pi\hat{R}+1-\pi}$ we have $\frac{s}{\pi} = \frac{e-s}{1-\pi}\hat{R} = \frac{\hat{R}e}{\pi\hat{R}+1-\pi}$. Therefore,

$$\begin{aligned} W'\left(\pi\frac{\hat{R}e}{\pi\hat{R}+1-\pi}\right) &= u'\left(\frac{\hat{R}e}{\pi\hat{R}+1-\pi}\right) - \hat{R}u'\left(\frac{\hat{R}e}{\pi\hat{R}+1-\pi}\right) \\ &= \left(1-\hat{R}\right)u'\left(\frac{\hat{R}e}{\pi\hat{R}+1-\pi}\right) \\ &< 0, \end{aligned}$$

where the inequality follows from $u' > 0$ and $\hat{R} > 1$. QED

Formal Definition of Incomplete-Markets Equilibrium in Section 3

At date 0, each agent chooses an investment portfolio (s, x). Agents solve

$$\begin{aligned} &\max_{(s,x)\geq(0,0)} \mathbb{E}[V_1(s,x;\theta)] \\ &s.t. \\ &s + x \leq e, \end{aligned} \tag{35}$$

where $V_1(s, x; \theta)$ is the indirect utility function representing the value the agent can obtain at date 1 if he holds investments (s, x) and receives realization θ of the liquidity shock. This indirect utility function is defined as follows:

$$\begin{aligned} V_1(s, x; \theta) &= \max_{(c_1, c_2) \geq (0,0), n, b} u(c_1 + \theta c_2), \\ & \quad s.t. \\ & \quad c_1 + pn + b \leq s, \\ & \quad n \geq -x, \\ & \quad c_2 \leq (x + n)\hat{R} + bR, \end{aligned} \tag{36}$$

where n represents net purchases of the long-term asset in the asset market at date 1 and b represents expenditures on the IOUs in the hidden retrade market. Let $n(\theta; s, x, p)$ denote net demand for the long-term asset of a type-θ agent.

Competitive equilibrium consists of initial investments s and x, value functions $V(s, x; \theta)$, a date-1 price p for the long-term asset, and a gross interest rate R in the hidden retrade market such that (i) given p and R, value functions solve (36); (ii) given V, investment choices s and x solve (35); and (iii) the date-1 market for the long-term asset clears, $\mathbb{E}[n(\theta; s, x, p)] = 0$, and R is an equilibrium interest rate on the hidden retrade market.

REFERENCES

Allen, Franklin, and Douglas Gale. 2004. "Financial Intermediaries and Markets." *Econometrica* 72 (July): 1,023–61.

Bianchi, Javier. 2011. "Overborrowing and Systemic Externalities in the Business Cycle." *American Economic Review* 101 (December): 3,400–26.

Bisin, Alberto, and Piero Gottardi. 2006. "Efficient Competitive Equilbria with Adverse Selection." *Journal of Political Economy* 91 (June): 485–516.

Dávila, Julio, Jay H. Hong, Per Krusell, and José-Víctor Ríos-Rull. 2012. "Constrained Efficiency in the Neoclassical Growth Model with Uninsurable Idiosyncratic Shocks." *Econometrica* 80 (November): 2,431–67.

Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 114 (June): 401–19.

Di Tella, Sebastian. 2014. "Optimal Financial Regulation and the Concentration of Aggregate Risk." Stanford Graduate School of Business Working Paper.

Ellerman, A. Denny, and Barbara K. Buchner. 2007. "The European Union Emissions Trading Scheme: Origins, Allocation, and Early Results." *Review of Environmental Economics and Policy* 1 (1): 66–87.

Ennis, Huberto M., and Todd Keister. 2010. "On the Fundamental Reasons for Bank Fragility." Federal Reserve Bank of Richmond *Economic Quarterly* 96 (1): 33–58.

Farhi, Emmanuel, Mikhail Golosov, and Aleh Tsyvinski. 2009. "A Theory of Liquidity and Regulation of Financial Intermediation." *The Review of Economic Studies* 76 (3): 973–92.

Geanakoplos, John D., and Heraklis M. Polemarchakis. 1986. "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations when the Asset Market is Incomplete." In *Essays in Honor of Kenneth Arrow*, Volume 3, edited by Walter P. Heller, Ross M. Starr, and David A. Starrett. New York: Cambridge University Press, 65–95.

Golosov, Mikhail, and Aleh Tsyvinski. 2007. "Optimal Taxation with Endogenous Insurance Markets." *The Quarterly Journal of Economics* 122 (2): 487–534.

Greenwald, Bruce C., and Joseph E. Stiglitz. 1986. "Externalities in Economies with Imperfect Information and Incomplete Markets." *The Quarterly Journal of Economics* 101 (May): 229–64.

Jacklin, Charles J. 1987. "Demand Deposits, Trading Restrictions, and Risk Sharing." In *Contractual Arrangements for Intertemporal Trade*, edited by Edward C. Prescott and Neil Wallace. Minneapolis: University of Minnesota Press, 26–47.

Kehoe, Timothy J., and David K. Levine. 1993. "Debt-Constrained Asset Markets." *The Review of Economic Studies* 60 (October): 865–88.

Kilenthong, Weerachart T., and Robert M. Townsend. 2011. "Information-Constrained Optima with Retrading: An Externality and Its Market-Based Solution." *Journal of Economic Theory* 146 (May): 1,042–77.

Kilenthong, Weerachart T., and Robert M. Townsend. 2014a. "Segregated Security Exchanges with Ex Ante Rights to Trade: A Market-Based Solution to Collateral-Constrained Externalities." Working Paper 20086. Cambridge, Mass.: National Bureau of Economic Research (May).

Kilenthong, Weerachart T., and Robert M. Townsend. 2014b. "A Market Based Solution to Price Externalities: A Generalized Framework." Working Paper 20275. Cambridge, Mass.: National Bureau of Economic Research (July).

Lorenzoni, Guido. 2008. "Inefficient Credit Booms." *The Review of Economic Studies* 75 (3): 809–33.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. New York: Oxford University Press.

Prescott, Edward C., and Robert M. Townsend. 1984. "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard." *Econometrica* 52 (January): 21–45.

Stiglitz, Joseph E. 1982. "The Inefficiency of the Stock Market Equilibrium." *The Review of Economic Studies* 49 (April): 241–61.

Viner, Jacob. 1932. "Cost Curves and Supply Curves." *Zeitschrift für Nationalökonomie* 3: 23–46.