Endogenous Liquidity and Contagion*

by

Rohit Rahi
Department of Finance,
Department of Economics,
and Financial Markets Group,
London School of Economics,
Houghton Street, London WC2A 2AE

and

Jean-Pierre Zigrand
Department of Finance
and Financial Markets Group,
London School of Economics,
Houghton Street, London WC2A 2AE

July 6, 2009.

*This paper was circulated earlier under the title "A Theory of Strategic Intermediation and Endogenous Liquidity." We thank seminar participants at the LSE/FMG Conference on Liquidity, the EFA meetings in Maastricht, Cass Business School, University of Venice, and New Economic School in Moscow.

Abstract

Market liquidity is typically characterized by a number of ad hoc metrics, such as depth, volume, bid-ask spreads etc. No general coherent definition seems to exist, and few attempts have been made to justify the existing metrics on welfare grounds. In this paper we propose a welfare-based definition of liquidity and characterize its relationship to the usual proxies. Our analysis rests on a general equilibrium model with multiple assets and restricted investor participation. Strategic intermediaries pursue profit opportunities by providing intermediation services (i.e. "liquidity") in exchange for an endogenous fee. Our model is well-suited to study the contagion-like effects of liquidity shocks.

Journal of Economic Literature classification numbers: G10, G20, D52, D53.
Keywords: Liquidity, intermediation, arbitrage, segmented markets, contagion.

1 Introduction

Market liquidity has long been a puzzle to financial economists. Given its myriad connotations, it is a bit odd that more attempts have not been made to analyze and reconcile these various aspects within one equilibrium model. Some well-known attributes of liquidity are depth (the market impact of a trade), breadth (the size of bid-ask spreads, also referred to as tightness), volume, intermediation and transaction costs (e.g. brokerage fees), as well as timeliness and ease of execution of trades. Rather than define liquidity by its attributes, we define liquidity by the underlying function that gives rise to those attributes. While any one model may be too specialized to capture all, or even many, of the salient features of liquidity, we believe that a general equilibrium model such as the one proposed here may help to guide our intuition as to which of these features are worthy of analysis. Indeed, since most papers on market liquidity are partial equilibrium models or partial equilibrium empirical studies, it is not obvious why the focus has been on one or the other asset or one or the other measure of liquidity.

We believe that the study of liquidity needs to ultimately be unambiguously grounded in a general equilibrium welfare analysis. Liquidity affects trades, which then may affect depth, tightness and timeliness, which in turn affect liquidity and welfare. Partial equilibrium liquidity concepts, such as depth or bid-ask spreads of individual securities, also ignore interdependencies across assets and markets. A particular asset may not be liquid, but substitutes may be liquid enough to compensate for it. For instance, the fact that the market for treasury futures contracts on Eurex US was not truly deep did not indicate that the treasury futures market was not liquid in general. Indeed, the reason why that market was not deep was precisely because most of the trade in treasury futures occurred on the CBOT. Similarly, the liquidity of the market for certain derivatives, such as calls and puts on a share, depends on the liquidity of the underlying securities. This calls for a *global* point of view that considers multiple assets traded in multiple markets.

In order to find a metric that is at the same time intuitive and welfare-based, one needs to resort to a realistic general equilibrium model with liquidity demanders and liquidity suppliers. In this paper, liquidity is provided by both investors and financial intermediaries.[1] Within this setup, we introduce and defend a particular liquidity metric and show that there is an unambiguous relationship of this metric to the attributes of liquidity mentioned above and to welfare. The metric we propose is not model-dependent, but its properties of course will be. Roughly, we define liquidity as the gains from trade achieved in equilibrium through the trading of securities. Markets are liquid if they allow investors to execute large amounts of welfare-enhancing security trades. The gains from trade are determined by the magnitude of the change in both prices and quantities, i.e. by the extent to which the

[1]We agree with Dewatripont and Tirole (1993) when they say: "The behavior of financial intermediaries ... largely determines the liquidity of financial markets. ... A more complete understanding of financial markets should thus explicitly integrate financial intermediation."

marginal valuations of investors change relative to autarky, and by the scale of the accompanying trades. Heuristically, our liquidity metric can be written as follows:

$$\begin{aligned} \text{liquidity} &= \text{gains from trade mediated through security markets} \\ &= (\text{scale of trades}) \times (\text{change in marginal valuations}) \end{aligned}$$

This measure of liquidity is intuitive. The first component, the scale of trades, is related to the market impact of trades, or depth. If markets are deep, an agent can trade a large amount without adversely affecting the terms of trade. By itself, however, this is not a sufficient measure of liquidity. First, depth is primarily a measure of the price impact of an exogenous market order from outside the model—what could be called *outside liquidity*. It does not capture the liquidity already provided to agents whose behavior is modeled as part of the equilibrium, or *inside liquidity*. Second, with multiple assets, there are as many ways to impact markets as there are portfolios that can be perturbed. Not all perturbations are economically useful. For instance imposing a small additional trade in a security that leads to a change in the intertemporal marginal rate of substitution that is uncorrelated with the payoff of the security being perturbed will have zero market impact and reflect a very deep market, although nobody desires or trades that economically irrelevant security.

The second component of our liquidity metric, i.e. the change in marginal valuations induced by trading, measures the usefulness of security markets in terms of the gain in efficiency that trading secures for investors. This efficiency gain is reflected in the degree to which marginal valuations are aligned relative to autarky as agents trade their way from the endowment point towards the contract curve. This will naturally depend upon the potential gains from trade, the degree of competition in intermediation, and the payoff characteristics of the securities available for trade. By itself, alignment of marginal valuations is not a sufficient characteristic of liquid markets either, for it could be that there is a large adjustment in marginal valuations, and yet the amount traded and its welfare impact are small.

That liquidity manifests itself in the interaction of the scale of trades and the alignment of marginal valuations is commonsensical to market practitioners. For instance, those derivatives contracts that survive by succeeding to attract liquidity do so precisely because they play to a natural hedging need between natural counterparties (the second component) with a need for sizable trades (the first component). It also turns out that this definition follows naturally from a general equilibrium model with trade occurring both directly and through intermediaries. For example, liquidity so-defined has a purely pecuniary characterization as the additional amount of consumption investors can enjoy due to more efficient pricing.

To make room for liquidity-providing intermediaries, we model asset markets as segmented. While there are assets that each given group of investors can trade among themselves, trades with other groups of investors require the intervention of intermediaries. As an obvious example we can cite the fact that buyers and sellers of the more exotic financial instruments rely on inter-dealer brokers to facilitate liquidity

by gathering pricing information and identifying counterparties with reciprocal interests. As in the real world, intermediaries in our model are larger, strategic entities maximizing trading profits, given the trades of other intermediaries. Asset prices and bid-ask spreads are determined endogenously at a Nash equilibrium. Since we allow entry into the intermediation sector (with a fixed cost of entry), the number of intermediaries is endogenous as well. In other words, liquidity depends on the number of intermediaries, and the number of intermediaries depends on liquidity. Liquidity in our model can be thought of as provided partly by the endogenous number of intermediaries ("across markets liquidity" in O'Hara (1995)) and partly through direct centralized trading in markets ("within markets liquidity" in O'Hara (1995)). This contrasts with many market microstructure studies where all trades must pass through market makers. When linking liquidity to welfare, the endogenous spreads are not treated as deadweight costs; they are rather viewed as forming the intermediaries' profits, which need to be accounted for in any welfare analysis.

We also evaluate the impact on liquidity of asset innovation by intermediaries (as is the case, for example, with many categories of OTC derivatives). We find that such innovation enhances overall market liquidity, though liquidity in some sectors of the economy may be adversely affected.

Our model lends itself directly to the study of the contagion effects of liquidity, i.e. how a liquidity shock in one sector of the economy is transmitted by intermediaries to other sectors, and which markets bear the brunt of the shock. We find a feedback effect through which a detrimental liquidity shock lowers the number of intermediaries, which in turn lowers liquidity and so on. A recent example of such a "liquidity spiral" can be seen in the demise of Lehman Brothers. This was caused by a liquidity shock originating in the US housing sector. The exit of Lehman in turn led to a further deterioration of liquidity, forcing other intermediaries to curtail their operations.

Our model also provides a framework within which one can understand the logic of securitization in the first place. The CDO boom was made possible not only by the low interest rate environment that led investors to seek out higher yields, but also by the arbitrage profits reaped by CDO structurers due to the difference between the price paid for debt, and the monies raised by selling tranches of that debt tailored to the needs of individual clienteles. Our framework provides a rationale for the CDO mechanism. Quite naturally, it also illustrates the dangers inherent in such a mechanism: should the demand for one of the tranches wane, this local liquidity shock ripples through all the tranches.

In order to keep the model tractable, we abstract from some attributes of liquidity. In particular, we choose to study an economy with only two dates, so that the temporal aspect of liquidity as the price of immediacy an investor needs to pay to an intermediary in order to transact now rather than later (as in Grossman and Miller (1988)) cannot be captured. Nor can we capture resiliency, the tendency that order flows do or do not have to induce return reversals. A second simplifying assumption that we impose in this paper is that information is symmetric.

Related Literature. There are is a vast literature studying market liquidity directly or indirectly. However, we are not aware of any papers that define liquidity via an explicit metric that itself has a clear welfare meaning, or that relate this definition to the different attributes of liquidity, such as depth, bid-ask spreads, transaction costs, immediacy services and the like. Grossman and Stiglitz (1980) directly assume exogenous liquidity *trades*, rather than liquidity shocks that may give rise to optimal liquidity trades. This is also the case in the models of Kyle (1985) and Glosten and Milgrom (1985) and in much of the ensuing literature on market microstructure. A number of papers, following Diamond and Verecchia (1981), have taken this further and study how different specifications of "liquidity shocks" translate into optimal trades and equilibrium outcomes. Another line of research can be traced to Diamond and Dybvig (1983) where investors realize in the interim period whether they are early or late consumers. This "liquidity shock" defines the role of intermediaries and gives rise to various financial contracts that intermediaries can engage in with investors.

Traditionally, liquidity has been studied mostly in single-asset models (see, for example, the papers cited in Chordia et al. (2000)), with little attention given to multi-asset liquidity, common factors, liquidity substitutes and so forth. Recently, however, a few empirical market microstructure papers have started to address this omission, among them Chordia et al. (2000), Hasbrouck and Seppi (2001) and Korajczyk and Sadka (2008). As far as theoretical modeling of multi-asset liquidity is concerned, less work has been done, be it in market microstructure or otherwise. Fernando (2003) models "liquidity shocks" as non-informative additive shocks that affect investors' marginal valuations of risky assets. His main interests are the price effects of idiosyncratic versus systematic liquidity shocks as well as how liquidity shocks to one asset affect prices of other assets. Brunnermeier and Pedersen (2009) model liquidity needs as arising from the asynchronous arrival of investors. Their main concern is the link between the capital or margin constraints faced by speculators and the liquidity they provide. In both these papers, while "liquidity shocks" are specified, no definition or metric of "liquidity" is proposed. Acharya and Pedersen (2005) specify illiquidity as an exogenous per-share cost of selling an individual security. They derive a liquidity-adjusted CAPM in which this illiquidity is priced.

There is a growing empirical literature in support of segmentation in asset markets. In particular it documents how assets in different market segments are priced by distinct groups of investors. The reader is referred to Rahi and Zigrand (2008) for a discussion of this literature. Financial contagion has been studied by Allen and Gale (2000), Freixas et al. (2000), Fernando (2003), and Gromb and Vayanos (2007), among others.

Our analysis builds on our earlier work in Rahi and Zigrand (2008, 2009). We use some results from these papers, in particular on the characterization of equilibrium (we summarize these results at the end of Section 3).

The paper is organized as follows. In the next section we introduce our definition of liquidity and outline some of its general properties. In Section 3 we describe and characterize our notion of equilibrium. In Section 4 we elaborate on the role played

by intermediaries in the provision of liquidity. In the next few sections we relate our liquidity measure to depth, bid-ask spreads, individual asset liquidity, volume, and welfare. In Section 9 we allow intermediaries to introduce new securities and analyze the impact on liquidity. In Sections 10 we show how our setup can be used to study contagion. An illustration of contagion in the CDO market follows in Section 11. Section 12 is devoted to extensions of our main results. Proofs are collected in the Appendix.

2 Liquidity as Realized Gains from Trade

In this section we formally define liquidity as "realized gains from trade." We argue that this metric captures the overall *economic meaning* of liquidity, as reflected also in welfare. Insofar as the aforementioned attributes of liquidity are not a good proxy for this liquidity measure, they may not be truly economically relevant.

At the most fundamental level, markets are more liquid the more diverse are the valuations of agents in the absence of trading and the larger the desired amounts of trade. In the extreme case where all agents have the same no-trade valuations, there are no gains from trade to be realized—trading volume is zero and markets can be deemed to be completely illiquid. For example, a situation in which all agents want to be on the same side of a trade, so that these trades cannot be consummated, is often referred to as a "drying up of liquidity."

We formalize this idea in a two-period economy in which assets are traded at date 0 and pay off at date 1. Our measure of liquidity involves a comparison of state-price deflators. Given a collection of J assets with random payoffs $d := (d_1, \ldots, d_J)$ and prices $q := (q_1, \ldots, q_J)$, a positive random variable p is called a state-price deflator[2] if $q_j = E[d_j p]$ for every asset j, or more compactly, $q = E[dp]$.

Consider first the benchmark case of a frictionless economy with complete markets. Let p^i be the no-trade valuation of agent i, i.e. the state-price deflator at which the agent chooses not to trade. Let p^W be a Walrasian state-price deflator. Then we measure the gains from trade of agent i in the equilibrium under consideration by[3] $\nu^i E[(p^i - p^W)^2]$, where ν^i is an agent-specific weight reflecting the scale of additional trades that the agent engages in when offered a marginally better price.[4] The corresponding liquidity measure is

$$\mathcal{L} := \sum_i \nu^i E[(p^i - p^W)^2]. \tag{1}$$

It captures the extent of diversity of individual valuations and the extent of mutually

[2] Other terms used in the literature for "state-price deflator" are "state-price density," "stochastic discount factor," and "pricing kernel."

[3] We restrict all random variables to lie in the linear space L^2 of square-integrable random variables.

[4] The precise meaning of ν^i will depend on the way preferences and endowments are modeled. In our formulation, it will turn out to be a preference parameter (see Section 4).

beneficial trades reaped by trading to the Walrasian first best equilibrium (in a sense that will be made precise below in Lemma 2.1).[5]

This definition is unambiguous if markets are complete. If markets are incomplete, however, there are multiple state-price deflators consistent with the same asset prices and payoffs. Consider the set of marketable payoffs $M := \{x : x = d \cdot \theta, \text{ for some portfolio } \theta \in \mathbb{R}^J\}$. Among all the state-price deflators p that price the payoffs in M identically, there is a unique one, p_M, that lies in M. This *traded* state-price deflator p_M is the least-squares projection on M of any of the deflators p (see Lemma 2.1 below). The liquidity metric (1) can therefore be extended to the incomplete markets case as follows:

$$\mathcal{L} := \sum_i \nu^i E[(p_M^i - p_M^W)^2]. \tag{2}$$

Liquidity thus defined is a measure of the gains from trade realized in equilibrium. However, there is no sense in which it can reflect bid-ask spreads, transaction costs, or intermediation costs, as these are absent in an economy with no frictions. Accordingly, we introduce a particular kind of market imperfection, which is empirically well-founded, namely market segmentation. This provides a role for intermediaries to exploit price differentials across market segments and in the process to provide liquidity.

We formalize market segmentation as follows. Assets are traded in several locations or "exchanges." There are K such exchanges, with I^k investors on exchange k. We also use K and I^k to denote the set of exchanges and the set of investors on exchange k,[6] i.e. $K := \{1, \ldots, K\}$ and $I^k := \{1, \ldots, I^k\}$. There are J^k assets available to agents on k, with the random payoff of a typical asset j denoted by d_j^k. Asset payoffs on exchange k can then be summarized by the random payoff vector $d^k := (d_1^k, \ldots, d_{J^k}^k)$.

While the "location" or "exchange" metaphor is a helpful one, it is more natural to think of the segmentation as being functional rather than geographical, e.g. in terms of investors restricted to certain asset classes (long-term bonds versus treasuries, equities versus derivatives etc.).

In a segmented economy agents can trade among themselves within each segment, and they can also trade across segments via intermediaries. Regardless of how intermediation is modeled, there is a natural generalization of the liquidity measure (2). Consider an equilibrium of the intermediated economy in which a state-price deflator for exchange k is given by $\hat{p}^k$, $k \in K$. Except in an ideal world of perfect intermediation, the $\hat{p}^k$'s will typically be different across exchanges. Let $p^{k,i}$ be a no-trade state-price deflator of the i'th agent on exchange k. Then we define the

[5]Lemma 2.1 shows that the mean-square distance between the state-price deflators p and p', given by $E[(p - p')^2]$, is equal to the largest difference in price for any security between the market priced by p and the market priced by p'. It is therefore a very intuitive measure of mispricing that does not depend on the precise securities chosen.

[6]Following standard convention, we use the same symbol to denote a set and its cardinality.

liquidity metric for exchange k as

$$\mathcal{L}^k := \sum_{i \in I^k} \nu^{k,i} E[(p^{k,i}_{M^k} - \hat{p}^k_{M^k})^2], \tag{3}$$

where M^k is the marketed subspace for exchange k. The corresponding aggregate liquidity measure is

$$\mathcal{L} := \sum_{k \in K} \mathcal{L}^k. \tag{4}$$

In the complete-markets frictionless case, all payoffs are marketable, and $\hat{p}^k = p^W$ for all k, so that (4) reduces to (1).

A complete characterization of this liquidity measure, and an analysis of its relationship to attributes such as bid-ask spreads and trading volume, must await a full description of the model. At this stage we motivate and describe some of its general properties that do not depend on the particular way in which equilibrium prices (the $\hat{p}^k$'s) are determined.

The term $E[(p^{k,i}_{M^k} - \hat{p}^k_{M^k})^2]$ in the definition of liquidity is the mean-square distance between agent (k, i)'s (traded) valuation $p^{k,i}_{M^k}$ and the equilibrium (traded) valuation of exchange k, $\hat{p}^k_{M^k}$. This has the interpretation of gains from trade reaped by agent (k, i) constrained by the assets available for trade on k (in particular, if there are no markets on k, these gains are zero). More generally, we can rely on the work of Chen and Knez (1995) on market integration to provide a characterization of mean-square distance between state-price deflators:

Lemma 2.1 *Given random variables p and p', and a marketed subspace M for some collection of assets, we have:*

1. *$p_M = p'_M$ if and only if $E[dp] = E[dp']$, for all payoffs $d \in M$.*

2. $$E[(p_M - p'_M)^2] = \max_{d:E[(d_M)^2]=1} [E(dp_M) - E(dp'_M)]^2$$

   i.e. $E[(p_M - p'_M)^2]$ is the maximal squared pricing error induced by p_M and p'_M among payoffs d with $E[(d_M)^2] = 1$.

3. $$E[(p_M - p'_M)^2] = \max_{d \in M:E[(d)^2]=1} [E(dp) - E(dp')]^2$$

   i.e. $E[(p_M - p'_M)^2]$ is the maximal squared pricing error induced by p and p' among marketed payoffs d with $E[(d)^2] = 1$.

The first statement says that two random variables are valid state-price deflators for a given collection of assets if and only if their marketed components are the same. Thus our liquidity measure does not depend on which state-price representation is chosen (i.e. $p^{k,i}$ could be any no-trade state-price deflator for agent (k, i) and $\hat{p}^k$ could

be any equilibrium state-price deflator for exchange k). The last two statements characterize the mean-square distance between the traded state-price deflators p_M and p'_M as a bound on the difference in asset valuations induced by them. More precisely, it is the maximal squared pricing error using p and p' to price (normalized) payoffs in M, or alternatively it is the maximal squared pricing error using the traded state-price deflators themselves to price all (normalized) payoffs, whether marketed or not.

Liquidity in our setting is provided by both investors and intermediaries. We can isolate the first component as follows. Let p^k be an autarky state-price deflator for exchange k. In the absence of intermediaries, $\hat{p}^k = p^k$, so that liquidity on k is

$$\mathcal{L}^k\big|_{N=0} = \sum_{i \in I^k} \nu^{k,i} E[(p^{k,i}_{M^k} - p^k_{M^k})^2], \tag{5}$$

This is the liquidity generated from the batch auction on exchange k, without any intervention of the intermediaries. It reflects the realized gains from trade of investors on k from trading among themselves.

For much of this paper we will be studying the case in which intra-exchange liquidity, given by (5), is zero, so that all liquidity is intermediated. This is a special case of our setup in which all investors within an exchange have the same no-trade valuations, but where endowments, preferences and asset spans differ across exchanges. We call this economy a *clientele economy.* In a clientele economy, $p^{k,i} = p^k$, for all k. Then liquidity on exchange k is

$$\mathcal{L}^k = \nu^k E[(p^k_{M^k} - \hat{p}^k_{M^k})^2],$$

where $\nu^k := \sum_{i \in I^k} \nu^{k,i}$. In the absence of intermediation $\hat{p}^k = p^k$ and $\mathcal{L}^k = 0$, for all k: markets are completely illiquid, as there are no liquidity providers. Liquidity on each exchange is also zero if the p^k's are all the same, so that there is no reason to trade across exchanges to begin with (in this case $\hat{p}^k$ must be equal to p^k for all k, as there are no profit opportunities for intermediaries).

It has been usual in the literature on liquidity, especially in applied work, to focus on depth and on spreads. While we will be more precise later on the relationship between our measure of liquidity and these proxies, a few general remarks are in order.

Analysis of depths and spreads in individual assets, as has been typical in the literature, suffers from the usual pitfalls of partial equilibrium analysis. Spreads have been analyzed by picking a few assets and then arguing that the spread in these assets is representative of the economy as a whole. For instance, refer to the excellent monograph by Marston (1995) where the integration of various national financial markets is measured by the degree of closeness with which these markets price various money market and fixed-income securities. Our framework, on the other hand, leads naturally to a measure of spreads that is a function of the mean-square distances between the state-price deflators $\{\hat{p}^k\}_{k \in K}$. The advantage of such a measure is that it considers willingness to pay directly, rather than indirectly through proxies

computed from a limited number of securities. In the latter procedure, a judgment must be made as to the most relevant assets or asset classes to compare. Furthermore, since identical assets, or more generally payoffs, may not exist on multiple exchanges, one would need to compare substitute assets. Both points raise a Pandora's box of judgmental issues which can be avoided entirely by using state prices instead. As shown in Lemma 2.1, the mean-square distance between the traded state-price deflators on two exchanges is equal to the bound on the squared pricing errors in using these state-price deflators to price any (normalized) payoff, whether marketed or not. In other words, it exactly represents what one is looking for when computing price differentials, and has the virtue of using and representing all available information.

It is easy to see that the level of mispricing, e.g. the size of bid-ask spreads, of individual securities need not have any relationship with the level of overall liquidity. Consider, for the sake of illustration, an asset with payoff d, $E[d] = 0$, that is traded on two exchanges, 1 and 2. The mispricing of this asset, given by $E[(\hat{p}^1 - \hat{p}^2)d]$, may be very low. For instance it is zero if the covariance between d and $\hat{p}^1 - \hat{p}^2$ is zero. Yet markets may be very illiquid, for instance if there are no intermediaries or if the potential gains from trade are insignificant. And the same applies to the converse: liquidity may be relatively high and yet bid-ask spreads for some asset may be large. In other words, the bid-ask spread for one particular asset may not necessarily provide a reliable indication as to the level of liquidity in the markets. All information impounded into the pricing relationships and gathered from the equilibrium actions of all agents needs to be taken into consideration, as is the case when using state prices.

In summary, market liquidity as we see it is a general snapshot spread, properly aggregated across all payoffs and all market segments. The apparent drawback of our definition is that it involves terms, such as autarky state-price deflators, which are hard to estimate. In the next few sections, we provide several characterizations of our liquidity metric in terms of variables that are in principle observable, such as the number of intermediaries and the cost of intermediation.

3 Equilibrium

The definition of liquidity proposed in this paper does not crucially depend on any particular choice of timing, agent characteristics or market structure, and is therefore of universal application. However, in order to derive closed-form solutions and to relate liquidity to welfare, a modeling choice must be made. The major difficulty that needs to be overcome is the fact that strategic intermediaries play a game whose payoffs are functions of the outcomes of a general equilibrium.

A tractable framework is obtained by making assumptions that yield a local CAPM on each exchange, as follows. Investor $i \in I^k$ on exchange $k \in K$ has date 0

endowment $\omega_0^{k,i}$, and date 1 endowment $\omega^{k,i}$. He has quadratic preferences:

$$U^{k,i}(x_0^{k,i}, x^{k,i}) = x_0^{k,i} + E\left[x^{k,i} - \frac{1}{2}\beta^{k,i}(x^{k,i})^2\right],$$

where $\beta^{k,i}$ is a positive parameter, $x_0^{k,i}$ is date 0 consumption, and $x^{k,i}$ is date 1 consumption. Investors behave competitively and can trade only on their own exchange.

In addition to the price-taking investors, there are N arbitrageurs (with the set of arbitrageurs also denoted by N) who possess the trading technology which allows them to also trade *across* exchanges, or in other words, which allows them to act as intermediaries if they so wish. Arbitrageurs only care about date 0 consumption. They are imperfectly competitive in our model, as they clearly are in actual financial markets. They have no endowments, so they can be interpreted as pure intermediaries.

We assume that all random variables (asset payoffs and endowments) have finite support. Then we can represent the uncertainty by a finite state space $S := \{1, \ldots, S\}$.

The interaction between price-taking investors and strategic arbitrageurs involves a Nash equilibrium concept with a Walrasian fringe, pioneered by Gabszewicz and Vial (1972). Let $y^{k,n}$ be the supply of assets on exchange k by arbitrageur n, and $y^k := \sum_{n \in N} y^{k,n}$ the aggregate arbitrageur supply on exchange k. For given y^k, $q^k(y^k)$ is the market-clearing asset price vector on exchange k, with the asset demand of investor i on exchange k denoted by $\theta^{k,i}(q^k)$.

Definition 1 *Given an asset structure $\{d^k\}_{k \in K}$, a Cournot-Walras equilibrium (CWE) of the economy is an array of asset price functions, asset demand functions, and arbitrageur supplies, $\{q^k : \mathbb{R}^{J^k} \to \mathbb{R}^{J^k}, \theta^{k,i} : \mathbb{R}^{J^k} \to \mathbb{R}^{J^k}, y^{k,n} \in \mathbb{R}^{J^k}\}_{k \in K, i \in I^k, n \in N}$, such that*

1. *Investor optimization: For given q^k, $\theta^{k,i}(q^k)$ solves*

$$\max_{\theta^{k,i} \in \mathbb{R}^{J^k}} x_0^{k,i} + E\Big[x^{k,i} - \frac{\beta^{k,i}}{2}(x^{k,i})^2\Big]$$

subject to the budget constraints:

$$\begin{aligned} x_0^{k,i} &= \omega_0^{k,i} - q^k \cdot \theta^{k,i} \\ x^{k,i} &= \omega^{k,i} + d^k \cdot \theta^{k,i}. \end{aligned}$$

2. *Arbitrageur optimization: For given $\{q^k(y^k), \{y^{k,n'}\}_{n' \neq n}\}_{k \in K}$, $y^{k,n}$ solves*

$$\max_{y^{k,n} \in \mathbb{R}^{J^k}} \sum_{k \in K} y^{k,n} \cdot q^k\Big(y^{k,n} + \sum_{n' \neq n} y^{k,n'}\Big)$$

$$s.t. \quad \sum_{k \in K} d^k \cdot y^{k,n} \leq 0.$$

3. *Market clearing:* $\left\{q^k(y^k)\right\}_{k\in K}$ *solves*

$$\sum_{i\in I^k} \theta^{k,i}(q^k(y^k)) = y^k, \qquad \forall k \in K.$$

A complete characterization of the CWE can be found in Rahi and Zigrand (2008, 2009). In the remainder of this section, we provide a brief synopsis of the relevant results. We refer the reader to the original papers for more details, including proofs.

Let $\beta^k := [\sum_i (\beta^{k,i})^{-1}]^{-1}$ and $\omega^k := \sum_i \omega^{k,i}$. Also define $p^{k,i} := 1 - \beta^{k,i}\omega^{k,i}$ and $p^k := 1 - \beta^k \omega^k$. This is consistent with our usage of $p^{k,i}$ and p^k in Section 2, as it can be shown that $p^{k,i}$ is a no-trade state-price deflator for agent (k,i) and p^k is an autarky state-price deflator for exchange k. Indeed, for given arbitrageur supply y^k,

$$q^k(y^k) = E\left[d^k[p^k - \beta^k(d^k \cdot y^k)]\right]. \tag{6}$$

Thus $p^k - \beta^k(d^k \cdot y^k)$ is a state-price deflator for exchange k. The autarky state-price deflator p^k is obtained by setting $y^k = 0$. We denote asset prices in autarky by $\mathring{q}^k := q^k(0) = E[d^k p^k]$.

Proposition 3.1 (Cournot-Walras equilibrium: Rahi and Zigrand (2008)**)** *There is a unique CWE.*[7]

1. *Equilibrium arbitrageur supplies are given by*

$$d^k \cdot y^{k,n} = \frac{1}{(1+N)\beta^k}\left(p^k_{M^k} - p^A_{M^k}\right), \qquad k \in K, \tag{7}$$

where $p^A \geq 0$ *is a state-price deflator for the arbitrageurs.*

2. *Equilibrium asset prices on exchange* k *are given by* $\hat{q}^k := E[d^k \hat{p}^k]$*, where*

$$\hat{p}^k := \frac{1}{1+N} p^k + \frac{N}{1+N} p^A. \tag{8}$$

Thus $\hat{p}^k$ *is an equilibrium state-price deflator for exchange* k*.*

3. *Aggregate arbitrageur profits originating from exchange* k *are given by*

$$\Phi^k := \hat{q}^k \cdot y^k = \frac{N}{(1+N)^2 \beta^k} E[(p^k_{M^k} - p^A_{M^k})^2]. \tag{9}$$

4. *The equilibrium demands of investors are given by*

$$d^k \cdot \theta^{k,i} = \frac{1}{\beta^{k,i}}(p^{k,i}_{M^k} - \hat{p}^k_{M^k}), \qquad i \in I^k, k \in K. \tag{10}$$

[7] Unlike Rahi and Zigrand (2008), here we denote equilibrium asset prices on exchange k by $\hat{q}^k$ instead of q^k.

5. *The equilibrium utilities of investors are given by*

$$U^{k,i} = \bar{U}^{k,i} + \frac{1}{2}\beta^{k,i} E[(d^k \cdot \theta^{k,i})^2], \qquad i \in I^k, k \in K, \tag{11}$$

where $\bar{U}^{k,i}$ is a constant that does not depend on the asset structure or investor portfolios.

The random variable p^A is a state-price deflator for the arbitrageurs in the sense that it is a state-price deflator, i.e. $\hat{q}^k = E[d^k p^A]$ for all k, and moreover $p^A(s)$ is the arbitrageurs' marginal shadow value of consumption in state s (formally, $p^A(s)$ is the Lagrange multiplier associated with the arbitrageurs' no-default constraint in state s). Note that p^A can be chosen so that it does not depend on N.

Given the centrality of the arbitrageur valuation p^A, it is important to provide an explicit characterization of it. To this end, we define a Walrasian equilibrium with restricted consumption as an equilibrium in which agents can trade any asset on a centralized exchange, facing a common state-price deflator p^{RC}, but agents on exchange k can consume claims in M^k only.[8] There are no arbitrageurs.

Proposition 3.2 (Arbitrageur valuations: Rahi and Zigrand (2009) **)** *Arbitrageur valuations in the CWE coincide with valuations in the Walrasian equilibrium with restricted consumption, i.e. $p^A_{M^k} = p^{RC}_{M^k}$, for all k. Consequently $\lim_{N\to\infty} \hat{q}^k = E[d^k p^{RC}]$.*

Thus asset prices in the arbitraged economy converge to asset prices in the restricted-consumption Walrasian equilibrium, as the number of arbitrageurs goes to infinity (note that this is an immediate consequence of (8), once it is established that $p^A_{M^k} = p^{RC}_{M^k}$).[9]

We obtain a sharper characterization of p^A under some restrictions on the asset structure $\{d^k\}_{k\in K}$. Let p^* denote the complete-markets Walrasian state-price deflator of the entire integrated economy with no participation constraints. It can be shown that

$$p^* = \sum_{k\in K} \lambda^k p^k,$$

where

$$\lambda^k := \frac{\frac{1}{\beta^k}}{\sum_{j=1}^K \frac{1}{\beta^j}}, \qquad k \in K.$$

The state-price deflator p^* reflects the autarky valuation of each exchange in proportion to its depth.

Now consider the following spanning condition:

[8] In other words, each investor can arbitrage all markets, but must then purchase a final consumption stream in the span of his local assets. See Rahi and Zigrand (2009) for a formal definition, and also for a discussion of the subtle difference between this notion of equilibrium and Walrasian equilibrium with restricted participation. In the latter, agents face a common state-price deflator, but agents on exchange k can trade claims in M^k only.

[9] The equilibrium allocation (for investors) in the arbitraged economy also converges to the restricted-consumption Walrasian equilibrium allocation.

(S) Either (a) $M^k = M$, $k \in K$, or (b) $p^k - p^* \in M^k$, $k \in K$.

Under **S**(a) we have an standard incomplete markets economy in which all investors trade the same payoffs, though on different exchanges. **S**(b) is the condition that characterizes an equilibrium security design (see Section 9). We have the following analogue of Proposition 3.2:

Proposition 3.3 (Arbitrageur valuations II: Rahi and Zigrand (2008) **)** *Suppose condition* **S** *holds. Then, arbitrageur valuations in the CWE coincide with valuations in the complete-markets Walrasian equilibrium, i.e.* $p^A_{M^k} = p^*_{M^k}$*, for all* k*. Consequently* $\lim_{N\to\infty} \hat{q}^k = E[d^k p^*]$.

4 Intermediation and Liquidity

Now that we are armed with a model with a closed-form solution of the unique equilibrium, we can explicitly characterize the properties of the liquidity measure defined in Section 2. In our model, the natural choice of the weight $\nu^{k,i}$ is (k,i)'s contribution to depth, $1/\beta^{k,i}$. Then the liquidity measure for exchange k is

$$\mathcal{L}^k = \sum_{i \in I^k} \frac{1}{\beta^{k,i}} E[(p^{k,i}_{M^k} - \hat{p}^k_{M^k})^2], \tag{12}$$

with economy-wide liquidity

$$\mathcal{L} = \sum_{k \in K} \mathcal{L}^k. \tag{13}$$

For a clientele economy ($p^{k,i} = p^k$, all k), liquidity on exchange k is

$$\mathcal{L}^k = \frac{1}{\beta^k} E[(p^k_{M^k} - \hat{p}^k_{M^k})^2]. \tag{14}$$

In words, liquidity on exchange k is equal to the gains from trade due to the improvement in pricing. We shall henceforth restrict ourselves to a clientele economy. The intuition for the general case where investors can trade among themselves on their own exchange and also across exchanges via intermediaries is very similar. The only difference is that, in general, liquidity has two components, one coming from the batch auctions on individual exchanges and the other one coming from intermediation, as explained in Section 2. Our primary focus in this paper is on intermediation. The relevant extensions of the main results to the general case are gathered in Section 12.

So how does intermediation create liquidity? Intermediation does not affect the spans $\{M^k\}_{k\in K}$, as there is no asset with a new dimension of spanning that becomes available due to pure intermediation.[10] What is achieved through intermediation is

[10]The case where intermediaries can issue assets to optimally intermediate is studied in Section 9.

that the existing assets can be used more fruitfully. Intermediaries provide liquidity in the very direct sense of being the counterparties to trades made possible due to their diverse customer base that reaches across various clienteles. Without intermediaries, those gains from trade cannot be reaped.

Thanks to intermediation, investors can trade on better terms. Suppose, for example, there are two exchanges, 1 and 2, with the same asset structure. Suppose there is an asset with payoff d for which the autarky price on exchange 1, $\mathring{q}^1 = E[dp^1]$, is lower than the autarky price on exchange 2, $\mathring{q}^2 = E[dp^2]$. Investors on 1 want to short the asset, while investors on 2 want to go long. By Proposition 3.3, we can choose $p^A = p^*$, which is a convex combination of p^1 and p^2. Hence the arbitrageurs' valuation of this asset, $q^A := E[dp^A]$, lies between p^1 and p^2. In the intermediated equilibrium, q^1 is pushed up and q^2 is pulled down (due to (8), $\hat{p}^k$ is closer to p^A than is p^k, for both exchanges). Intermediaries allow investors on exchange 1 to sell on better terms, while investors on exchange 2 can buy on better terms, with the spread narrowing. The welfare of investors increases even though intermediaries take home some profits.

Notice that liquidity for clientele k is scaled by $1/\beta^k$. From (6), it is clear that β^k is the price impact of a unit of arbitrageur trading on exchange k: the state s value of the state-price deflator $p^k - \beta^k(d^k \cdot y^k)$ falls by β^k for a unit increase in arbitrageur supply of s-contingent consumption. Thus $1/\beta^k$ is the depth of exchange k.

The equilibrium arbitrageur supply, given by (7), is very intuitive. Assuming for the moment that markets are complete on all exchanges, an arbitrageur supplies state s consumption to those exchanges which value it more than he does ($p^k_s - p^A_s > 0$). How much he supplies to exchange k depends on the size of the mispricing $|p^k_s - p^A_s|$, on the depth $1/\beta^k$, with more consumption supplied the deeper the exchange, and finally on the degree of competition N. If markets are incomplete, however, the difference between state prices may not be marketable. The arbitrageur would then supply state-dependent consumption as close to $p^k - p^A$ as permissible by the available assets d^k. The closest such choice is the projection $(p^k - p^A)_{M^k} = p^k_{M^k} - p^A_{M^k}$. The greater the number of arbitrageurs competing for the given opportunities, the smaller is each arbitrageur's residual demand, and so the less each one supplies. In the limiting equilibrium, as N goes to infinity, arbitrageurs virtually disappear in that individual arbitrageur trades vanish, as does their aggregate consumption, $\sum_k \Phi^k$, and they perform the reallocative job of the Walrasian auctioneer at no cost to society (as formalized in Proposition 3.2).

Another way to see this is to compare realized and potential gains from trade. Since arbitrageur valuations are Walrasian (Proposition 3.2), we can define the potentially achievable or total gains from trade as

$$\overline{\mathcal{L}} := \sum_{k \in K} \overline{\mathcal{L}}^k, \tag{15}$$

where

$$\overline{\mathcal{L}}^k := \frac{1}{\beta^k} E[(p^k_{M^k} - p^A_{M^k})^2]. \tag{16}$$

$\overline{\mathcal{L}}$ measures the gains from trade that can be reaped if the economy moves from the autarky equilibrium to a perfectly intermediated, Walrasian, equilibrium, with the asset spans remaining unchanged. $\overline{\mathcal{L}}^k$ measures the total gains from trade between k and the rest of the economy. These gains ultimately arise from differences in preferences (e.g. risk aversion) and endowments. In that sense, one can interpret date zero as the time when investors learn about their preferences and endowments, i.e. about their idiosyncratic "liquidity shocks."

Proposition 4.1 (Competition and liquidity) *In a clientele economy,*

$$\mathcal{L}^k = \left(\frac{N}{1+N}\right)^2 \overline{\mathcal{L}}^k, \qquad k \in K. \tag{17}$$

In particular, $\mathcal{L}^k$ is strictly increasing in N, $\mathcal{L}^k = 0$ at $N = 0$, and $\lim_{N\to\infty} \mathcal{L}^k = \overline{\mathcal{L}}^k$. Consequently, aggregate liquidity $\mathcal{L}$ is increasing in N, $\mathcal{L} = 0$ at $N = 0$, and $\lim_{N\to\infty} \mathcal{L} = \overline{\mathcal{L}}$.

This result follows from the fact that $p^k - \hat{p}^k = \frac{N}{1+N}(p^k - p^A)$, due to (8). The expression (17) shows how our liquidity measure captures the general costs of trading due to the noncompetitive nature of the intermediation business. More competition improves upon the extent of gains from trade realized in the markets. In the limit, as competition becomes perfect, all potential gains from trade are exploited.

One of the advantages of our setup is that it is straightforward to endogenize the number of intermediaries as a function of the cost of entry into the intermediation business. While there are a number of related concepts of entry, the following is simple and sensible. Suppose each arbitrageur must bear a fixed cost c in order to set up shop and intermediate across all markets. First we determine the number of arbitrageurs N', not necessarily a natural number, so that each one of the N' arbitrageurs makes a profit of 0 after having borne the fixed costs. Using (9), (15) and (16), N' solves

$$c = \frac{1}{N'} \sum_k \Phi^k(N') = \frac{\overline{\mathcal{L}}}{(1+N')^2}. \tag{18}$$

Second, this number is rounded down to the nearest natural number:

Proposition 4.2 *In a clientele economy, the equilibrium level of intermediation is given by*

$$N = \mathrm{rd}\left(\sqrt{c^{-1}\overline{\mathcal{L}}} - 1\right), \qquad c \leq \frac{\overline{\mathcal{L}}}{4}.$$

The operator "rd" rounds the real number in parenthesis down to the next natural number. In particular, arbitrageurs are allowed to make profits in equilibrium, but not enough to attract one further arbitrageur. We must have $c \leq \overline{\mathcal{L}}/4$ in order for intermediation to arise (this will be a standing assumption for the rest of the paper). N increases as c falls, with $\lim_{c\to 0} N = \infty$.

The assumption of unrestricted but costly entry provides us with a simple proxy for liquidity. Using (17) and (18), and ignoring integer constraints on N, we get:

Proposition 4.3 *In a clientele economy, liquidity is given by*

$$\mathcal{L} = cN^2.$$

With estimates of c and N, an estimate of liquidity is then simply the cost of entry times the square of the number of intermediaries, or equivalently the total cost borne by the intermediation sector times the number of intermediaries. Notice that even though depth is a crucial ingredient of liquidity, it appears only insofar as it affects the endogenous number of intermediaries N. An added bonus is that N is a variable which can in principle be observed directly rather than having to be estimated.

Finally, it follows from Propositions 4.2 and 4.3 (again ignoring integer constraints) that

$$\mathcal{L} = \left(\sqrt{\overline{\mathcal{L}}} - \sqrt{c}\right)^2.$$

Liquidity is increasing in the maximal amount of gains from trade allowed by preferences and securities in a first-best world, $\overline{\mathcal{L}}$, and decreasing in the entry costs c. Lower entry costs mean more competition amongst arbitrageurs, which leads to improved terms of trade and improved quantities offered to investors, and consequently higher liquidity.

5 Depth and Spreads

Depth, $1/\beta^k$, enters directly into the liquidity measure $\mathcal{L}^k$, as one would expect. It is constant, and in particular independent of arbitrage trades. This is a very convenient feature of our model, for it allows us to show the endogenous nature of liquidity, even though depth is constant.

While depth is constant, the supply of an asset on exchange k has a differential impact on the prices of other assets on k depending on the payoff structure d^k. From (6),

$$\frac{\partial q_j^k(y^k)}{\partial y_{j'}^k} = -\beta^k E[d_j^k d_{j'}^k]. \tag{19}$$

The price impact of one unit of trade in asset j' on exchange k is more pronounced for those assets on k that are close substitutes in the sense of having a higher noncentral comovement with j'. For normalized payoffs d, with $E[d^2] = 1$, β^k measures the own-price effect.

Since arbitrageur supply is scaled by depth, there is a natural connection between depth and volume of trade. We will return to this in Section 7, where we discuss the relationship between volume and liquidity.

Turning now to spreads, we define $S^k := E[(\hat{p}_{M^k}^k - p_{M^k}^A)^2]$ as the generalized true[11] "bid-ask spread" on exchange k. It is the spread between the valuation on k

[11] "True" in the sense that it measures the mispricing between the transaction price and the true value, here the shadow price p^A.

and the average valuation in the whole economy as measured by p^A (which is also the Walrasian valuation p^{RC}). As the number of arbitrageurs grows without bound, $\hat{p}^k$ converges to p^A, so that the spread S^k converges to zero.

Under the spanning condition **S** (for example if the same assets are traded on all exchanges), we can relate the true spread S^k to pairwise spreads between k and other exchanges:

Proposition 5.1 (Spreads) *If the spanning condition* **S** *holds, then*

$$
\begin{aligned}
S^k &= E\left[\left(\sum_{\ell \in K} \lambda^\ell (\hat{p}^k_{M^k} - \hat{p}^\ell_{M^k})\right)^2\right] \\
&= \frac{1}{(1+N)^2} \cdot E\left[\left(\sum_{\ell \in K} \lambda^\ell (p^k_{M^k} - p^\ell_{M^k})\right)^2\right].
\end{aligned}
$$

Under **S**, the spread S^k is the squared pricing error between k and the rest of the economy, with the pricing error relative to another exchange being weighted by its relative depth. This equilibrium spread is in fact the same as the autarky spread scaled by the number of intermediaries. As N increases, the spread S^k falls, in tandem with the increase in liquidity $\mathcal{L}^k$.

6 Individual Asset Liquidity

We have defined liquidity as the overall ease with which gains from trade can be exploited. In this section we deduce asset-by-asset liquidity measures from the aggregate measure. The main reason for doing so is to be able to contrast our theoretical results with the existing empirical literature.

Intuitively, the empirical findings of Chordia et al. (2000) that liquidity can be correlated between certain assets is not surprising from a theoretical point of view. The assets supplied in large amounts by arbitrageurs all share the characteristic of being valuable to investors, and those assets will all see higher volumes and liquidity than the remaining assets. Assets that do not contribute towards the realization of gains from trade will not see active trading. In other words, from an economic point of view, the commonality in liquidity across various assets is their contribution to the portfolio mimicking the gains from trade (for exchange k, this is the portfolio whose payoff is $p^k - \hat{p}^k$).

Recall that $\mathring{q}^k = E[d^k p^k]$ is the autarky asset price vector on exchange k, and $\hat{q}^k = E[d^k \hat{p}^k]$ is the equilibrium asset price vector on k. We can formally disaggregate liquidity $\mathcal{L}^k$ into the diverse contributions of the J^k assets on exchange k as follows:

Proposition 6.1 *In a clientele economy,*

$$
\mathcal{L}^k = \frac{1}{\beta^k} b^k \cdot (\mathring{q}^k - \hat{q}^k),
$$

where $b^k := \{b^k_j\}_{j \in J^k}$ is the regression coefficient of the multiple regression of $p^k - \hat{p}^k$ on d^k.

The coefficient b^k_j is the portion of the variation of the trading gains $p^k - \hat{p}^k$ that is explained by asset j on exchange k. Accordingly, we define the local liquidity of this asset on k as

$$\mathcal{L}^k_j := \frac{1}{\beta^k} b^k_j (\mathring{q}^k_j - \hat{q}^k_j),$$

so that indeed

$$\mathcal{L}^k = \sum_{j=1}^{J^k} \mathcal{L}^k_j.$$

The liquidity of asset j on exchange k equals its depth times the usefulness of asset j in generating overall gains from trade on k, b^k_j, times the gains from trade directly reaped from trading j, $\mathring{q}^k_j - \hat{q}^k_j$, regardless of the indirect gains from trade reflected in the other assets. The term $\frac{1}{\beta^k} b^k_j$ is in fact equal to θ^k_j, the equilibrium holding of asset j by clientele k (see Rahi and Zigrand (2008)). The local liquidity of asset j can therefore be characterized as follows:

Proposition 6.2 (Local asset liquidity) *In a clientele economy,*

$$\mathcal{L}^k_j = \theta^k_j (\mathring{q}^k_j - \hat{q}^k_j),$$

i.e. the liquidity of asset j on exchange k equals the amount of resources saved due to the more favorable equilibrium asset prices induced by intermediation.

Note that $\mathcal{L}^k_j$ is positive. The equilibrium holding θ^k_j is equal to the arbitrageur supply y^k_j. From (19) we can see that the own-price effect of arbitrageur supply is negative. For example, if $y^k_j > 0$, then $\hat{q}^k_j < \mathring{q}^k_j$.

Thus liquidity has a purely pecuniary interpretation as the additional amount of time zero consumption investors can enjoy due to more efficient pricing. This benefit is larger the greater the degree of competition among intermediaries.

Finally, consider the case in which the same assets (or, more generally, payoffs) trade in all locations. Let $\mathcal{L}_j := \sum_{k \in K} \mathcal{L}^k_j$ be the global, or economy-wide, liquidity of asset j.

Proposition 6.3 (Global asset liquidity) *Consider a clientele economy with $d^k = d$, for all $k \in K$. Then*

$$\mathcal{L}_j = N\Phi_j,$$

where $\Phi_j := \sum_k y^k_j \hat{q}^k_j$ is the aggregate arbitrageur profit in asset j.

Thus the global liquidity of asset j is the number of arbitrageurs times the total profits reaped by them in intermediating this asset.

Propositions 6.2 and 6.3 suggest a tight relationship between volume and liquidity, which is the subject of the next section.

7 Liquidity and Volume

We define the *inter-exchange volume* originating from exchange k as

$$\tilde{\mathcal{V}}^k := E[(d^k \cdot y^k)^2].$$

This is the overall equilibrium volume of trade in state-contingent consumption implied by intermediated asset trades on exchange k. From (7),

$$\tilde{\mathcal{V}}^k = \left[\frac{N}{(1+N)\beta^k}\right]^2 E[(p^k_{M^k} - p^A_{M^k})^2].$$

Using (16) and (17), we obtain the following result:

Proposition 7.1 (Liquidity and volume) *In a clientele economy, liquidity equals volume per unit of depth:* $\mathcal{L}^k = \beta^k \tilde{\mathcal{V}}^k$.

As one would expect, a welfare-based notion of liquidity is associated not with the volume of asset transactions, but with the volume of the induced net trade in the underlying state-contingent consumption. It is the latter that empirical researchers should try to measure when looking for a volume-based proxy for liquidity. Implicit in these trades are the motivations that gave rise to them as well as the microstructure considerations of asset spans and degree of competition in the intermediation sector.

The relationship between volume and liquidity highlighted in Proposition 7.1 is quite intuitive. For a given volume, more gains from trade are realized the closer state prices move towards Walrasian ones. State prices do not move very much in deep markets. Therefore volume needs to be large relative to depth to exploit and exhaust gains from trade, which are measured by liquidity. Of course, volume is itself increasing in depth, and the net effect of depth on liquidity is positive, indicating that the volume effect of depth dominates the direct depth effect.

8 Welfare

Equilibrium welfare of investors is given by (11). In a clientele economy we measure the welfare of clientele k by $U^k := \sum_{i \in I^k} U^{k,i}$ and economy-wide welfare by $U := \sum_{k \in K} U^k$. Using (10) and (11),

$$U^k = \bar{U}^k + \frac{1}{2}\mathcal{L}^k$$

and

$$U = \bar{U} + \frac{1}{2}\mathcal{L},$$

where $\bar{U}^k := \sum_{i \in I^k} \bar{U}^{k,i}$ and $\bar{U} := \sum_{k \in K} \bar{U}^k$. Similarly, from (9), (16) and (17), total arbitrageur profits originating from exchange k are

$$\begin{aligned}\Phi^k &= \frac{N}{(1+N)^2}\overline{\mathcal{L}}^k \\ &= \frac{1}{N}\mathcal{L}^k,\end{aligned} \tag{20}$$

so that aggregate economy-wide profits are

$$\sum_{k \in K} \Phi^k = \frac{1}{N}\mathcal{L}. \tag{21}$$

This leads us to the following result on the relationship between the various concepts:

Proposition 8.1 (Liquidity, welfare and volume) *In a clientele economy, the following concepts, local as well as global, are isomorphic: liquidity, investor welfare, profits, social welfare and volume.*

As we have argued in the introduction, we feel that any measure or metric of market liquidity would have to be tightly related to welfare and profits in order to be economically meaningful. The above proposition confirms that this is indeed the case in our model.

9 Security Design

In this section we allow intermediaries to innovate and add assets to the ones already available for trade on the exchanges. We shall see that the optimally innovated assets not only augment intermediary profits, but also allow a better exploitation of gains from trade, leading to higher liquidity, volume and welfare.

One might guess that any innovation would lead to more liquid markets. The reasoning might be as follows: since intermediaries can always choose not to trade the new assets, volumes, and therefore liquidity, cannot be lower than in the absence of innovation. The reality is more complicated though, since liquidity as defined here captures the extent to which markets allow the economy to move closer to the ideal Walrasian equilibrium for the given asset structure. Since an asset innovation perturbs the Walrasian equilibrium also (in particular the deflator p^A), it is not necessarily true that pricing at the new equilibrium is closer to the new Walrasian equilibrium than the old pricing was to the old Walrasian equilibrium. It turns out, however, that the aforementioned logic is correct if the innovations are optimal for arbitrageurs.

We have already seen, in Section 3, that there is a unique CWE for any given asset structure $\{d^k\}_{k \in K}$. We now allow each arbitrageur to add assets to each exchange before any trading takes place. This determines a new asset structure $\{d^k_{innov}\}_{k \in K}$. The payoffs of the arbitrageurs in this security design game are the profits they

earn in the ensuing CWE.[12] Which asset(s) would arbitrageurs introduce at a Nash equilibrium of this game? Rahi and Zigrand (2008) show that there is a unique asset added to each exchange (if not already present):

Proposition 9.1 (Optimal innovation: Rahi and Zigrand (2008)**)**
For a given $\{d^k\}_{k \in K}$, the asset structure

$$\begin{array}{ccc} [d^k \quad (p^k - p^*)] & \text{if} & p^k - p^* \notin M^k; \\ d^k & \text{if} & p^k - p^* \in M^k; \end{array}$$

is

1. *a minimal optimal asset structure for arbitrageurs; and*

2. *a minimal Nash equilibrium of the security design game.*

The reader is referred to Rahi and Zigrand (2008) for a proof and a detailed discussion of this result. The term "minimal" refers to the fact that there are other optimal (or equilibrium) configurations, but involving more assets—all of these configurations span an asset structure that is minimal. If there is an innovation cost, howsoever small, the chosen structure would unambiguously be a minimal one.

Since arbitrageur profits are higher in the post-innovation economy (condition 1 of Proposition 9.1), so is liquidity due to the isomorphism between profits and liquidity (Proposition 8.1)):

Proposition 9.2 (Innovation and liquidity) *In a clientele economy, liquidity $\mathcal{L}$ increases when intermediaries can innovate assets.*

A clear distinction needs to be made between local and global liquidity. While liquidity overall improves with optimal innovation, even though the intermediaries act strategically, it is also shown in Rahi and Zigrand (2008) that profits on any particular exchange may fall. Invoking the isomorphism between *local* profits and liquidity (Proposition 8.1), this means that innovation may hurt liquidity on some exchanges. The intuition goes as follows. If due to the innovation one of the exchanges sees decreased volume due to decreased usefulness of trade, then liquidity falls on that exchange. This occurs for instance if the exchange in question had an initial asset structure that permitted intermediaries to execute some crucial trades, say to borrow some state-contingent resources. When intermediaries can innovate optimally, they build such trades into the assets they innovate, thereby reducing the need to execute the trades on the exchange in question.

[12]Note that all arbitrageurs are able to trade the assets introduced by any one arbitrageur. Also, due to the symmetry of the CWE (Proposition 3.1), all arbitrageurs have the same equilibrium payoff.

10 Transmission of Liquidity Shocks

We now turn to the study of how liquidity shocks are transmitted across the economy. Starting from an initial equilibrium, we perturb fundamentals on one of the exchanges and analyze the economy-wide repercussions of this local shock. For simplicity, this is not a temporal shock that could have been anticipated. In this regard we follow most of the literature on contagion.

In order to simplify the analysis, we shall assume that the spanning condition **S** holds, i.e. either the security design is optimal, or the same set of payoffs are tradable on all exchanges. Then we can choose $p^A = p^* = \sum_k \lambda^k p^k$ by Proposition 3.3. We shall also continue to restrict ourselves to a clientele economy.

We consider a local shock on exchange ℓ. There are a number of ways to model this shock. The following turns out to be analytically tractable. Consider a shock to the number of investors active in the market, I^ℓ, while preserving the relative distribution of preferences and endowments on ℓ, $\{\beta^{\ell,i}, \omega^{\ell,i}\}_{i\in I^\ell}$. A negative participation shock on exchange ℓ lowers its depth $1/\beta^\ell$ while keeping its autarky state-price deflator, $p^\ell = 1 - \beta^\ell \omega^\ell$, constant. Consequently p^ℓ plays a less prominent role in p^A, but without making the economy more risk averse as would have happened had we simply lowered the depth of exchange ℓ.

Let

$$\vartheta^{k\ell} := \frac{E[(p^k_{M^k} - p^A_{M^k})(p^\ell_{M^k} - p^A_{M^k})]}{E[(p^k_{M^k} - p^A_{M^k})^2]}.$$

Thus $\vartheta^{k\ell}$ is the regression coefficient of the (projected) mispricing on exchange ℓ, $p^\ell_{M^k} - p^A_{M^k}$, on the mispricing on exchange k, $p^k_{M^k} - p^A_{M^k}$; this measure of covariation is a noncentral "beta" in the language of the CAPM. Ignoring integer constraints on N, we have the following result:[13]

Proposition 10.1 (Contagion) *Consider a clientele economy satisfying the spanning condition* **S**. *Then the effect on exchange k of a population (or participation) shock on exchange ℓ is given by*

$$\frac{d\log\mathcal{L}^k}{d\log I^\ell} = \underbrace{\mathbf{1}_{k=\ell} - 2\lambda^\ell\vartheta^{k\ell}}_{\left.\frac{d\log\mathcal{L}^k}{d\log I^\ell}\right|_N} + \frac{\mathcal{L}^\ell}{N\mathcal{L}}.$$

The indicator function $\mathbf{1}_{k=\ell}$ takes the value 1 if $k = \ell$, and is zero otherwise. Effects can be split into two categories: direct effects for a given N, captured by the term $(\mathbf{1}_{k=\ell} - 2\lambda^\ell\vartheta^{k\ell})$, and indirect effects via entry or exit which are represented by the term $\mathcal{L}^\ell/(N\mathcal{L})$.

---

[13] This result requires the assumption that **S** holds in a neighborhood of I^ℓ, so that we can set p^A equal to p^* both before and after the shock. This is clearly not an issue if the same payoffs are traded on all exchanges (condition **S**(a)). However, if we invoke **S**(b), the result should be interpreted as the long-run effect of a population shock, allowing for optimal adjustment of the security design. While it is difficult to obtain an analytical result if we fix the (initially optimal) security design, numerical examples can be worked out, as we do in the next section.

Consider first an exchange $k \neq \ell$, and suppose N is fixed. The effect on liquidity on k is $-2\lambda^{\ell}\vartheta^{k\ell}$. If the parameter $\vartheta^{k\ell}$ is negative, exchanges k and ℓ are *complements* in the sense that arbitrageurs tend to buy on one when they are selling to the other, i.e. there is intermediated trade between the two exchanges. If exchange ℓ experiences a reduction in its investor base, and a consequent deterioration of its depth, these intermediated trades become less valuable and less plentiful in equilibrium, thus reducing liquidity on k.

With endogenous N, this effect is exacerbated: fewer investors and lower depth on ℓ lead to less trade and to lower liquidity, which in turn leads to lower profits and thereby to fewer intermediaries, which in turn affects liquidities adversely and so forth. It is this cascade of deteriorating liquidities that has received significant attention in the contagion literature. The net effect of this feedback loop is represented by the term $\mathcal{L}^{\ell}/(N\mathcal{L})$. The effect is more pronounced the larger is the role of exchange ℓ in generating trades, as measured by its relative size $\mathcal{L}^{\ell}/\mathcal{L}$, and the smaller the initial N. A smaller initial N means that the feedback loop of liquidity on N and again of N on liquidity etc. is stronger as each arbitrageur is more powerful and holds a larger portfolio the less competition there is.

So far we have assumed k and ℓ to be complements. On the other hand, if $\vartheta^{k\ell} > 0$, valuations on exchanges k and ℓ are similar in the sense of being on average on the same side as the economy-wide valuation p^*. The two exchanges therefore compete for trades, and can be said to be *substitutes*. In this case, a shallower ℓ induces intermediaries to migrate to k, thereby increasing liquidity on k, for given N. The contagion effect operating through a lower N is however the same as in the case of complementary exchanges.

Finally, consider the effect of a population shock on exchange ℓ on its own liquidity. For fixed N, this effect is given by $(1 - 2\lambda^{\ell})$. If λ^{ℓ} is small, this has the straightforward interpretation of the direct loss of liquidity due to the flight of investors. This is compounded by the consequent flight of intermediaries in the same way as for the rest of the economy. If λ^{ℓ} is non-negligible, however, there is a countervailing effect. Indeed, if the parameter $\lambda^{\ell} > 1/2$, $\mathcal{L}^{\ell}$ actually increases when the population on ℓ falls, for given N. This might at first appear odd, but the effect stems from the endogenous nature of Walrasian prices. Fewer investors on exchange ℓ lower the depth of exchange ℓ, and everything else constant, liquidity is lower. But the smaller size of this clientele also means that it will now play a less prominent role in the determination of the economy-wide valuation p^*. The valuation p^* will become more dissimilar from p^{ℓ}, thereby increasing the potential gains from trade between ℓ and the rest of the economy, stimulating intermediated trades and increasing liquidity on ℓ. If $\lambda^{\ell} > 1/2$, this effect is strong enough to compensate for the loss of depth, before accounting for the knock-on effect on the number of intermediaries.

Evidently, in an economy with many exchanges, loss of liquidity is more likely to go hand in hand with a decline of active investors. But there might be situations where a dominant exchange optimally limits or rations participants. There may be situations in which a lower number of investors can sustain a higher level of liquidity,

or conversely where the arrival of more (identical) investors can hurt local liquidity. The converse implication is that liquidity can suffer on an exchange that experiences a rise in its investor population while substitute exchanges at the same time benefit from higher liquidity. These various examples show that there is a clear liquidity externality in our economy that can go in either direction.

What is the effect on asset prices of a liquidity shock? It is instructive to consider the case where the same assets trade on all exchanges so that price comparisons are straightforward. Accordingly, we assume that $d^k = d$, all k. Then $q^* := E[dp^*]$ is the asset price vector implied by the hypothetical complete-markets state-price deflator for the entire integrated economy.

Proposition 10.2 *Suppose $d^k = d$, for all $k \in K$. Then*

$$\frac{\partial \hat{q}^k}{\partial I^\ell} = \frac{N}{1+N}\frac{\lambda^\ell}{I^\ell}(\mathring{q}^\ell - q^*), \qquad k \in K.$$

Thus, if exchange ℓ in isolation values assets more than does the economy as a whole ($\mathring{q}^\ell > q^*$), an adverse participation shock on ℓ depresses asset prices worldwide. This is because the tendency of exchange ℓ to pull up asset prices, via intermediated trades, is reduced when its weight in the world economy is lower. Quite naturally, the effect is more pronounced the greater the degree of intermediation.

As an illustration of contagion, consider the episode documented by Peek and Rosengren (1997). They study the liquidity shock emanating from Japan at the end of the 1980s and beginning of the 1990s. We can interpret this shock as a drop in the Japanese local investor base. While Japan was a major financial power, it is safe to assume that it did not constitute more than half the world's financial depth. Given that the flow of capital was from Japan to the US, Japan and the US were complements and on average assets were cheaper abroad than in Japan. The adverse shock to Japanese liquidity depressed stock prices in Japan. The authors documented that the result of this liquidity shock was a sharp decline in Japanese investment in the US, which in turn adversely affected liquidity in the US, an instance of contagion along the lines suggested by our model.

The following section is devoted to a more elaborate example, in which contagion occurs not across national markets, but across different segments of the fixed income market.

11 An Example of Contagious Illiquidity: CDO Boom and Bust

Consider the CDO mechanism. The profit to intermediaries from structuring and marketing CDOs ultimately stems from the fact that the tranched cash flows can be sold for more than the procurement cost of the cash flows from credit, such as loans and mortgages.

For simplicity, the following example consists of four clienteles. Exchange $k = 4$ represents the clientele from which the credit originates, modeled as a single security with payoff d^4. Suppose there are three states of the world, and the promised cash flows from credit are 3. Due to default, however, the effective cash flows are $d^4 = (3, 2, 1)$, where we write the random variable d^4 as a vector of state-contingent payoffs. In other words, in state $s = 1$ all loans are repaid, in state $s = 2$ two-thirds are repaid, and in state $s = 3$ only one-third are repaid. Intermediaries slice these cash flows into three tranches. The supersenior tranche is sold off to the highest bidders, here represented by investors of type $k = 1$. We assume that the supersenior tranche always pays off,[14] with $d^1 = (1, 1, 1)$. The mezzanine tranche, paying off $d^2 = (1, 1, 0)$, is sold to the highest bidding clientele, $k = 2$. Notice that the mezzanine tranche suffers a loss in state 3. Finally, the highest bidders for the junior tranche are investors on exchange $k = 3$. The junior tranche only pays off in state $s = 1$ as it is the first to absorb any losses: $d^3 = (1, 0, 0)$. To summarize, the asset structure is:

$$d^1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad d^2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad d^3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad d^4 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}. \tag{22}$$

We construct an economy in which the equilibrium strategies of the arbitrageurs consist of buying the debt on exchange 4, tranching it, and selling each tranche off to the clientele that values it most. We are interested in the transmission of liquidity shocks across this economy. In particular, based on current accounts of the subprime crisis, the relevant question is what the repercussions on overall liquidity are of a diminished clientele for the supersenior tranche.

To simplify our calculations, we assume that the three states are equally probable, and all investors have the same preference parameter $\beta^{k,i} = 1/4$. Furthermore, we assume that exchanges 2, 3 and 4 have the same population, which we normalize to one (i.e. $I^2 = I^3 = I^4 = 1$). We denote the population on exchange 1 by I (i.e. $I^1 = I$). We shall reduce I to reflect investor flight from the supersenior CDO tranche. Date 1 endowments are as follows:

$$\omega^{1,i} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \omega^{2,i} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \omega^{3,i} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \omega^{4,i} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}.$$

The corresponding autarky state-price deflators, given by $p^k = p^{k,i} = 1 - \beta^{k,i}\omega^{k,i}$, are:

$$p^1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad p^2 = \begin{bmatrix} 1 \\ 1 \\ \frac{3}{4} \end{bmatrix}, \quad p^3 = \begin{bmatrix} 1 \\ \frac{3}{4} \\ \frac{3}{4} \end{bmatrix}, \quad p^4 = \begin{bmatrix} 0 \\ \frac{1}{4} \\ \frac{1}{2} \end{bmatrix}.$$

Thus clientele 1 has the highest willingness to purchase the supersenior payoff d^1. Likewise, clienteles 2 and 3 are the highest bidders for the mezzanine and junior tranches, d^2 and d^3, respectively.

[14] This is irrelevant for our results. With more states, superseniors can default as well.

To understand the rationale for the CDO structure, consider first the benchmark case in which $I = 1$. Then the complete-markets Walrasian state-price deflator for the integrated economy, p^*, is $3/4$ in all three states. It is easy to check that the asset structure (22) is the optimal security design, i.e. tranching is optimal for arbitrageurs. For every unit of d^4 that arbitrageurs buy, they sell one unit each of the tranches d^1, d^2 and d^3. The arbitrageurs' valuation p^A is equal to p^*.

Compare this, for instance, to the case in which a pass-through security is sold to all investors. Then the asset structure is $(3, 2, 1)$ on all exchanges. The arbitrageurs' valuation is the same as above and equal to p^*. For every unit that arbitrageurs buy on exchange 4, they sell 6/14, 5/14 and 3/14 units on exchanges 1, 2 and 3, respectively. Maximal liquidity, $\overline{\mathcal{L}}^k$, is unchanged for exchange 4 but lower for the other exchanges. The equilibrium level of intermediation is lower, leading to lower liquidity (and welfare) on all four exchanges.

While the CDO structure is optimal for $I = 1$, it is not so for other values of I. In particular, we are interested in what happens if appetite for the supersenior tranche diminishes, given this CDO structure. For $I \neq 1$, the spanning property **S** fails, which means that we cannot use the convenient condition $p^A = p^*$. The following can be verified to be a Lagrange multiplier for the arbitrageurs' first order conditions, and therefore a valid state-price deflator:

$$p^A = \frac{3}{17I + 3}\begin{bmatrix} 4I + 1 \\ 4I + 1 \\ 9I - 4 \end{bmatrix},$$

provided $I \geq 4/9$, which we will henceforth assume.[15] Equilibrium arbitrageur supplies are:

$$y^{1,n} = y^{2,n} = y^{3,n} = -y^{4,n} = \frac{1}{1+N} \cdot \frac{20I}{17I+3}.$$

Thus the pattern of trade is the same as in the benchmark case of $I = 1$. These trades are simply scaled down as I falls. Notice that arbitrageur trades are exactly offsetting, so that $\sum_k y^{k,n} d^k = 0$. Equilibrium asset prices are given by:

$$\hat{q}^1 = 1 - \tfrac{N}{1+N}\tfrac{5}{17I+3}, \qquad \hat{q}^2 = \tfrac{2}{3} - \tfrac{N}{1+N}\tfrac{10I}{3(17I+3)},$$
$$\hat{q}^3 = \tfrac{1}{3} - \tfrac{N}{1+N}\tfrac{5I}{3(17I+3)}, \qquad \hat{q}^4 = \tfrac{1}{3} + \tfrac{N}{1+N}\tfrac{70I}{3(17I+3)}.$$

Maximal economy-wide liquidity is

$$\overline{\mathcal{L}} = \frac{100I}{3(17I+3)}.$$

As I falls, so does $\overline{\mathcal{L}}$. This means that, even for fixed N, overall liquidity $\mathcal{L}$, which is given by $(\frac{N}{1+N})^2\overline{\mathcal{L}}$, falls. In fact, it can be verified that the same is true for the

[15]The results are less clear-cut when I falls below 4/9. This is because there are not enough investors to absorb consumption in state 3, so it ends up in the hands of the arbitrageurs. Then our assumption that arbitrageurs only care about consumption at date zero, which is fairly innocuous as long as the asset structure does not deviate too far from one that satisfies **S**, starts to matter.

liquidity of tranches 2 and 3, and the liquidity of the underlying debt. Moreover, as I falls, intermediaries start going out of business, with N given by $\mathrm{rd}(\sqrt{c^{-1}\mathcal{L}} - 1)$. This exacerbates the drying up of liquidity.

Figures 1 and 2 illustrate the effects on liquidity and intermediation of a change in I, both above and below 1, for $c = .001$.

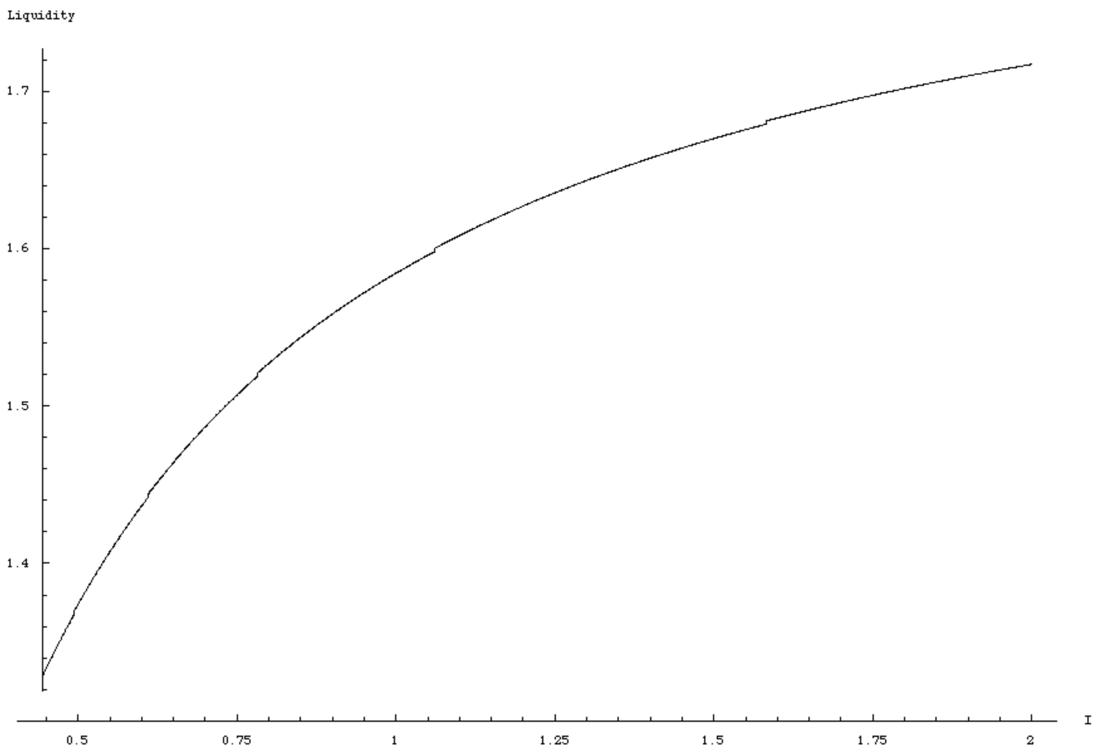

Figure 1: Overall Liquidity, $\mathcal{L}$, as a function of I

Contagion is seen to be very strong: as the natural clientele for the supersenior tranche is eroded, the entire CDO market seizes up. A 50% decline in the size of this clientele (starting from $I = 1$) causes overall liquidity to decline by more than 13%. This effect aggregates the impact of a change in I on relative depths, on shadow prices p^A, as well as on N. The plots for the liquidity of tranches 2 and 3, and for the liquidity of the securitized debt, are similar to that for overall liquidity.

During the boom phase, before doubts about the creditworthiness of CDOs and related products became prevalent, demand for tranches was in part fueled by the quest for yield in a low interest rate environment. In our model, the CDO mechanism leads to lower prices of the various tranches than would have obtained in its absence (i.e. $\hat{q}^k < \mathring{q}^k, k = 1, 2, 3$). In other words, the CDOs allow the credit and money markets to deliver higher yields. Likewise, the CDO mechanism allows debtors to borrow at a more attractive rate ($\hat{q}^4 > \mathring{q}^4$).

Everything else constant, higher demand for the supersenior tranche leads to

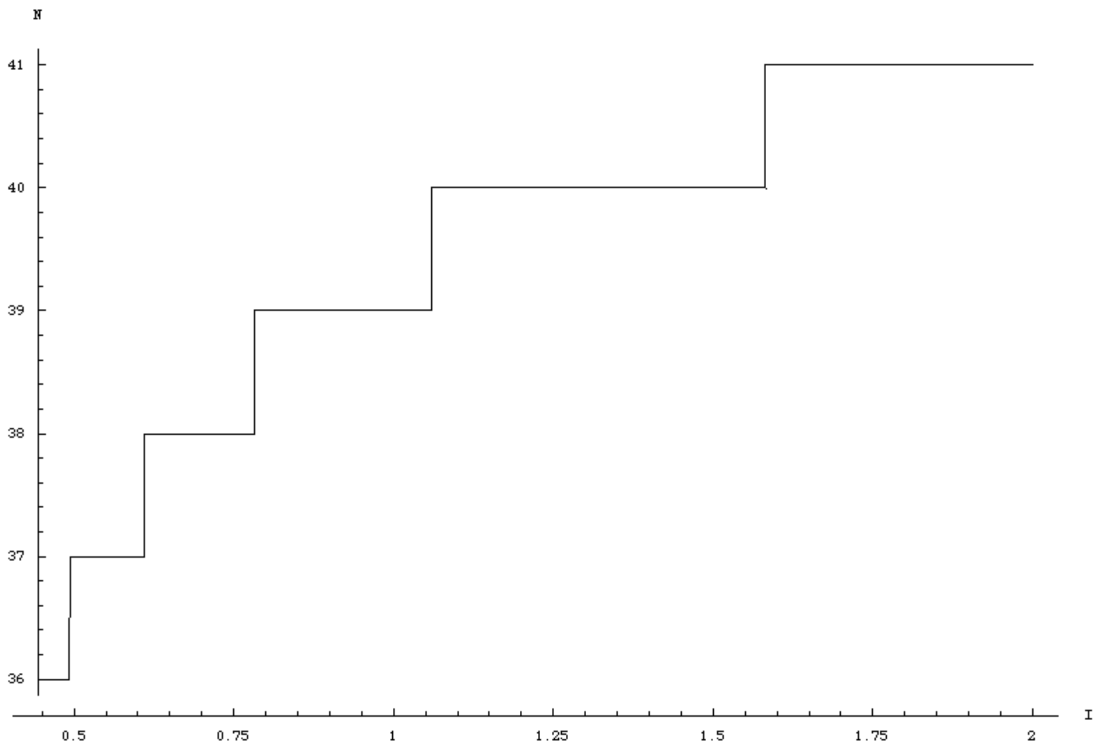

Figure 2: Equilibrium Number of Arbitrageurs, N, as a function of I

higher supersenior prices,[16] as well as higher prices for the underlying securitized debt. Concurrently, prices for the other tranches fall – and yields rise – since these investors find more counterparties for their trades. This is contagion. And if on the contrary demand for the supersenior tranche wanes, these effects are reversed: prices for tranches 2 and 3 rise and the corresponding yields fall as arbitrageurs are forced to reduce their shorts and buy back those tranches.

The crisis events unfolding in the credit markets from Summer 2007 onwards cannot be fully captured by this simple version of our model. Contrary to our assumptions here, banks in the real world did have their own capital and used it to keep the supersenior tranches when they found no buyers for them. They went on structuring CDOs and selling the remaining lower graded tranches off, pocketing the "arbitrage" profits (they were arbitrage trades for the structuring desks, who sold the supersenior tranches to the treasury department of the same organization, but not for the intermediary as a whole). This overextension into CDOs then became plain when an "unexpected" state was realized wherein the supersenior tranches were no longer perceived to pay back their face value. More elaborate versions of our model can be constructed to allow for arbitrageur capital and for default, but this is beyond the scope of this paper.

[16] One can check that this is true in spite of the countervailing effect of higher N.

12 Heterogeneous Investors

In this section we allow a diversity of investors within each exchange. Liquidity is then defined by (12) and (13). Most of our results continue to hold in this more general setup, but we only mention those extensions which add to our understanding of liquidity, in particular the relationship between liquidity on the one hand and volume and welfare on the other.

The volume of trade originating from agent (k, i) is

$$\mathcal{V}^{k,i} := E[(d^k \cdot \theta^{k,i})^2].$$

The *total volume* of trade on exchange k is $\mathcal{V}^k := \sum_{i \in I^k} \mathcal{V}^{k,i}$. This contrasts with the *inter-exchange volume* originating from exchange k defined above as $\tilde{\mathcal{V}}^k := E[(d^k \cdot y^k)^2]$. Evidently, $\tilde{\mathcal{V}}^k \leq \mathcal{V}^k$ since part of the trading volume on k arises from direct trades among the local investors.

Plugging the expression for $d^k \cdot \theta^{k,i}$, given by (10), into the definition of volume, we get

$$\mathcal{L}^k = \sum_{i \in I^k} \beta^{k,i} \mathcal{V}^{k,i}.$$

The intuition is the same as for Proposition 7.1. Since liquidity also considers the gains from trade realized from intra-exchange trade, liquidity is related to total volume, not just to the inter-exchange portion of volume as was the case in a clientele economy.

We now turn to the welfare properties of liquidity. If we give each investor (k, i) the same welfare weight we see that the results derived above in the clientele case again hold exactly:

$$U^k = \bar{U}^k + \frac{1}{2} \mathcal{L}^k$$

and

$$U = \bar{U} + \frac{1}{2} \mathcal{L}.$$

We now summarize these results:

Proposition 12.1 (Liquidity, welfare and volume: general case) *The following measures are isomorphic:*

- *liquidity, investor welfare and total volume, local as well as global,*
- *local profits and local inter-exchange volume.*

Since arbitrageur profits on exchange k only depend on inter-exchange trade, the expression for Φ^k given by (20) is still valid here, i.e. $\Phi^k = \mathcal{L}^k / N$, where $\mathcal{L}^k$ is the liquidity measure for a clientele economy. Thus there is a wedge between the objectives of arbitrageurs and investors. Investors consider trades among themselves as valuable, whereas arbitrageurs do not. This will have obvious consequences if intermediaries design securities. As shown in Rahi and Zigrand (2008), while the designed securities are socially optimal in a clientele economy, this is no longer true in a more general economy.

13 Conclusion

In this paper we have attempted to provide an equilibrium notion of liquidity that encompasses the disparate attributes commonly attached to the word "liquidity": depth, ease of trade, volume, low transaction costs, and so forth. We define liquidity as the extent to which gains from trade are realized in equilibrium. Liquidity originates from local batch auctions as well as from intermediation.

This simple definition has a number of pleasant characteristics. We show that the definition is welfare-grounded: liquidity equals social welfare. We also provide a number of operational forms which can in principle be tested and estimated, and which coincide with some of the intuitive attributes of liquidity. For instance, overall liquidity is equal to the total cost of intermediation times the number of intermediaries. Liquidity can also be disaggregated into individual asset liquidity, which equals the amount of resources saved due to the more favorable asset prices induced in equilibrium by the intermediaries. Some of the standard liquidity measures, on the other hand, such as individual asset bid-ask spreads, have only a tenuous connection to welfare—they fail to take the bigger picture into account wherein investors can use substitute assets to satisfy their portfolio needs.

Having defined and characterized the relevant concepts, the paper studies the transmission of liquidity shocks originating in one sector of the economy. The signs of the "contagion" effects are shown to depend on the degree of substitutability of the market segments, while the intensity is shown to depend on the relative depth and liquidity of the segment from where the shock emanates, as well as on the number of intermediaries.

Our analysis does not require us to assume that assets are exogenously given. In fact, we show that if intermediaries innovate assets with the sole aim of augmenting their own private profits, the equilibrium innovation improves liquidity, and therefore welfare, at least in an economy in which all liquidity is intermediated.

Appendix

In the Appendix we adopt matrix notation in order to simplify the proofs. We represent asset payoffs d^k by the $S \times J^k$ matrix R^k whose j'th column lists the state-by-state payoffs of the j'th asset. The set of traded payoffs M^k is then the column space of R^k.

Let Π be the diagonal matrix whose diagonal elements are the probabilities of the states, $\pi_1, \ldots, \pi_S$. A state-price deflator for (q, R) is a vector $p \in \mathbb{R}^S$ such that $q = R^\top \Pi p$.[17] In other words, state-price deflators can be viewed as vectors instead of random variables. Similarly, the expectation $E[xy]$ can be written as $x^\top \Pi y$, where the random variables x and y are viewed as vectors in $\mathbb{R}^S$. In our finite-dimensional setting, the inner product space L^2 is the space $\mathbb{R}^S$ endowed with the inner product $\langle x, y \rangle_2 := x^\top \Pi y$. Then $x_{M^k} = P^k x$, where P^k is the orthogonal projection operator in L^2 onto M^k, given by the idempotent matrix $P^k := R^k ({R^k}^\top \Pi R^k)^{-1} {R^k}^\top \Pi$. An explicit derivation of P^k can be found in Rahi and Zigrand (2008). P^k depends on R^k only through the span M^k. The L^2-norm of $x \in \mathbb{R}^S$ is $\|x\|_2 := (x^\top \Pi x)^{\frac{1}{2}}$.

In this notation, liquidity for exchange k, in the general case, is

$$\mathcal{L}^k = \sum_{k \in K} \frac{1}{\beta^{k,i}} \| P^k (p^{k,i} - \hat{p}^k) \|_2^2.$$

For a clientele economy,

$$\mathcal{L}^k = \frac{1}{\beta^k} \| P^k (p^k - \hat{p}^k) \|_2^2.$$

Proof of Lemma 2.1 The proof is adapted from arguments in Chen and Knez (1995). Let the asset payoff matrix be R with marketed subspace M and corresponding projection matrix P.

The first statement says that $P(p - p') = 0$ if and only if $R^\top \Pi (p - p') = 0$. If $P(p - p') = 0$, then $R^\top \Pi P (p - p') = 0$. But $R^\top \Pi P = R^\top \Pi$, so that $R^\top \Pi (p - p') = 0$. Conversely, $R^\top \Pi (p - p') = 0$ implies that $(p - p') \in M^\perp$. Hence $P(p - p') = 0$.

As to the second statement, consider a payoff d, not necessarily in M. The mispricing of d using Pp versus Pp' is $m(d) := d^\top \Pi P (p - p')$. Since $\Pi P = P^\top \Pi P$, by the Cauchy-Schwartz inequality we have $m(d) \leq \|Pd\|_2 \|P(p - p')\|_2$; equality occurs if and only if Pd and $P(p - p')$ are collinear. It follows that

$$\|P(p - p')\|_2 = \max_{d : \|Pd\|_2 = 1} d^\top \Pi P (p - p'). \tag{23}$$

For the last statement, consider a payoff $d \in M$. Then $d = R\theta$ for some $\theta \in \mathbb{R}^N$. Using again the fact that $R^\top \Pi P = R^\top \Pi$, we see that $m(d) = d^\top \Pi (p - p')$. Hence,

[17]The symbol $\top$ denotes "transpose." We adopt the convention of taking all vectors to be column vectors by default, unless transposed.

(23) can be written as

$$\|P(p-p')\|_2 = \max_{d\in M:\|d\|_2=1} d^\top \Pi(p-p').$$

■

Proof of Proposition 5.1 If **S** holds, we can choose $p^A = p^* = \sum \lambda^k p^k$ by Proposition 3.3. Using (8), it is easy to check that $p^A = \sum_k \lambda^k \hat{p}^k$. This gives us the first expression for S^k. Another consequence of (8) is $\hat{p}^k - \hat{p}^\ell = \frac{1}{1+N}(p^k - p^\ell)$. This gives us the second expression. ■

Proof of Proposition 6.1

$$\begin{aligned}\mathcal{L}^k &= \frac{1}{\beta^k}\|P^k(p^k-\hat{p}^k)\|_2^2 \\ &= \frac{1}{\beta^k}(p^k-\hat{p}^k)^\top {P^k}^\top \Pi P^k (p^k-\hat{p}^k) \\ &= \frac{1}{\beta^k}\underbrace{(p^k-\hat{p}^k)^\top \Pi R^k({R^k}^\top \Pi R^k)^{-1}}_{{b^k}^\top}\underbrace{{R^k}^\top \Pi(p^k-\hat{p}^k)}_{\mathring{q}^k-\hat{q}^k}.\end{aligned}$$

■

Proof of Proposition 6.3 Using (8), we have

$$\hat{q}^k = \frac{1}{1+N}\mathring{q}^k + \frac{N}{1+N}q^*,$$

where $q^* = E[dp^*]$. It follows that $\mathring{q}^k - \hat{q}^k = N(\hat{q}^k - q^*)$. From Proposition 6.2, $\mathcal{L}_j^k = N\theta_j^k(\hat{q}_j^k - q_j^*) = Ny_j^k(\hat{q}_j^k - q_j^*)$. Since the same assets trade on all exchanges, arbitrageur holdings of any asset, aggregated across all exchanges, must be zero, i.e. $\sum_k y^k = 0$. Hence, $\mathcal{L}_j = N\sum_k y_j^k \hat{q}_j^k$. ■

Proof of Proposition 10.1 In order to calculate the effect of a proportional change in I^ℓ, $d\log I^\ell$, it is convenient to write the population of exchange ℓ as αI^ℓ, with corresponding depth α/β^ℓ, and compute derivatives with respect to α evaluated at $\alpha = 1$. Using (17), we can write $\mathcal{L}^k$ as a function of α, p^A and N:

$$\begin{aligned}\mathcal{L}^k(\alpha, p^A(\alpha), N(\alpha)) &= \frac{\alpha^k}{\beta^k}\left(\frac{N}{1+N}\right)^2 \|P^k(p^k-p^A)\|_2^2 \\ &= \frac{\alpha^k}{\beta^k}\left(\frac{N}{1+N}\right)^2 (p^k-p^A)^\top \Pi P^k(p^k-p^A),\end{aligned}$$

where $\alpha^k = \alpha$ for $k = \ell$, and $\alpha^k = 1$ for $k \neq \ell$. The total derivative of $\mathcal{L}^k$ with respect to α is

$$\frac{d\mathcal{L}^k}{d\alpha} = \underbrace{\frac{\partial \mathcal{L}^k}{\partial \alpha} + \frac{\partial \mathcal{L}^k}{\partial p^A}\cdot {p^A}'(\alpha)}_{\left.\frac{d\mathcal{L}^k}{d\alpha}\right|_N} + \frac{\partial \mathcal{L}^k}{\partial N}N'(\alpha). \tag{24}$$

Noting that

$$p^A(\alpha) = \frac{\frac{\alpha}{\beta^\ell} p^\ell + \sum_{k \neq \ell} \frac{1}{\beta^k} p^k}{\frac{\alpha}{\beta^\ell} + \sum_{k \neq \ell} \frac{1}{\beta^k}},$$

we have

$$\begin{aligned}
\frac{\partial \mathcal{L}^k}{\partial \alpha} &= \mathcal{L}^\ell \mathbf{1}_{k=\ell} \\
\frac{\partial \mathcal{L}^k}{\partial p^A} &= -\frac{2}{\beta^k} \left(\frac{N}{1+N} \right)^2 \Pi P^k (p^k - p^A) \\
p^{A\prime}(\alpha) = \lambda^\ell (p^\ell - p^A) &= \frac{\beta^k \lambda^k}{\beta^\ell} (p^\ell - p^A),
\end{aligned} \tag{25}$$

where all the derivatives are evaluated at $\alpha = 1$. Hence the effect on $\mathcal{L}^k$ for given N is

$$\left. \frac{d\mathcal{L}^k}{d\alpha} \right|_N = \mathcal{L}^\ell \mathbf{1}_{k=\ell} - \frac{2\lambda^k}{\beta^\ell} \left(\frac{N}{1+N} \right)^2 \varphi^{k\ell}, \tag{26}$$

where

$$\varphi^{k\ell} := (p^k - p^A)^\top \Pi P^k (p^\ell - p^A).$$

We now solve for $N'(\alpha)$. With free entry, $\mathcal{L} = cN^2$ (Proposition 4.3). Therefore, the function $N(\alpha)$ is defined by the identity

$$\sum_k \mathcal{L}^k(\alpha, p^A(\alpha), N(\alpha)) - c[N(\alpha)]^2 \equiv 0.$$

Implicit differentiation gives us

$$N'(\alpha) = \frac{\sum_k \left. \frac{d\mathcal{L}^k}{d\alpha} \right|_N}{2cN - \sum_k \frac{\partial \mathcal{L}^k}{\partial N}}.$$

Now

$$\frac{\partial \mathcal{L}^k}{\partial N} = \frac{2}{N(1+N)} \mathcal{L}^k, \tag{27}$$

so that

$$\sum_k \frac{\partial \mathcal{L}^k}{\partial N} = \frac{2}{N(1+N)} \mathcal{L} = \frac{2cN}{1+N},$$

where we have once again used the result that $\mathcal{L} = cN^2$.

Under the spanning condition **S**, either $P^k = P$ or $P^k(p^k - p^A) = p^k - p^A$. In both cases $\sum_k \lambda^k \varphi^{k\ell} = 0$, since $p^A = p^*$. Therefore, from (26),

$$\sum_k \left. \frac{d\mathcal{L}^k}{dI^\ell} \right|_N = \mathcal{L}^\ell.$$

Altogether, this yields

$$N'(\alpha) = \frac{(1+N)\mathcal{L}^\ell}{2cN^2} = \frac{(1+N)\mathcal{L}^\ell}{2\mathcal{L}}. \tag{28}$$

Substituting (26), (27) and (28) into (24) gives us

$$\frac{d\mathcal{L}^k}{d\alpha} = \mathcal{L}^\ell \mathbf{1}_{k=\ell} - \frac{2\lambda^k}{\beta^\ell}\left(\frac{N}{1+N}\right)^2 \varphi^{k\ell} + \frac{\mathcal{L}^k\mathcal{L}^\ell}{N\mathcal{L}}.$$

Dividing through by $\mathcal{L}^k$ we get the desired result (note that $\vartheta^{k\ell} = \frac{\varphi^{k\ell}}{\varphi^{k,k}}$). ■

Proof of Proposition 10.2 Let $R^k = R$, all k. Then, using (8),

$$\begin{aligned} \hat{q}^k &= R^\top \Pi \hat{p}^k \\ &= R^\top \Pi \left(\frac{1}{1+N} p^k + \frac{N}{1+N} p^A \right). \end{aligned}$$

Therefore, from (25),

$$\frac{\partial \hat{q}^k}{\partial d \log I^\ell} = \frac{N}{1+N} \lambda^\ell R^\top \Pi (p^\ell - p^A).$$

Moreover, since condition $\mathbf{S}$(a) is satisfied, we have $p^A = p^*$. This yields the desired result. ■

References

Acharya, V. and Pedersen, L. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, 77(2):375–410.

Allen, F. and Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Brunnermeier, M. and Pedersen, L. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2199.

Chen, Z. and Knez, P. J. (1995). Measurement of market integration and arbitrage. *Review of Financial Studies*, 8(2):287–325.

Chordia, T., Roll, R., and Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics*, 56:3–28.

Dewatripont, M. and Tirole, J. (1993). *The Prudential Regulation of Banks*. MIT Press.

Diamond, D. and Dybvig, P. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91:401–419.

Diamond, D. and Verecchia, R. (1981). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics*, 9:221–235.

Fernando, C. S. (2003). Commonality in liquidity: Transmission of liquidity shocks across investors and securities. *Journal of Financial Intermediation*, 12:233–254.

Freixas, X., Parigi, B., and Rochet, J.-C. (2000). Systemic risk, interbank relations and liquidity provision by the central bank. *Journal of Money Credit and Banking*, 32(3):611–638.

Gabszewicz, J. J. and Vial, J.-P. (1972). Oligopoly 'a la Cournot' in a general equilibrium analysis. *Journal of Economic Theory*, 4(3):381–400.

Glosten, L. and Milgrom, P. (1985). Bid ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 13:71–100.

Gromb, D. and Vayanos, D. (2007). Financially constrained arbitrage and cross-market contagion. Mimeo, London School of Economics.

Grossman, S. and Miller, M. (1988). Liquidity and market structure. *Journal of Finance*, 43:617–633.

Grossman, S. and Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70:393–408.

Hasbrouck, J. and Seppi, D. (2001). Common factors in prices, order flows and liquidity. *Journal of Financial Economics*, 59:383–411.

Korajczyk, R. and Sadka, R. (2008). Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics*, 87(1):45–72.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica*, 53:1315–1336.

Marston, R. (1995). *International Financial Integration: A Study of Interest Differentials Between the Major Industrial Countries*. Cambridge University Press.

O'Hara, M. (1995). *Market Microstructure Theory*. Blackwell.

Peek, J. and Rosengren, E. S. (1997). The international transmission of financial shocks: The case of Japan. *The American Economic Review*, 87(4):495–505.

Rahi, R. and Zigrand, J.-P. (2008). Strategic financial innovation in segmented markets. *Review of Financial Studies*. Forthcoming.

Rahi, R. and Zigrand, J.-P. (2009). Walrasian foundations for equilibria in segmented markets. Mimeo, London School of Economics.