

Leaning against the wind

Pierre-Olivier Weill*

January 18, 2007

Abstract

During financial disruptions, marketmakers provide liquidity by absorbing external selling pressure. They buy when the pressure is large, accumulate inventories, and sell when the pressure alleviates. This paper studies optimal dynamic liquidity provision in a theoretical market setting with large and temporary selling pressure, and order-execution delays. I show that competitive marketmakers offer the socially optimal amount of liquidity, provided they have access to sufficient capital. If raising capital is costly, this suggests a policy role for lenient central-bank lending during financial disruptions.

Keywords: marketmaking capital, marketmaker inventory management, financial crisis.

*Department of Economics, University of California, Los Angeles, 8283 Bunche, Box 951477, Los Angeles, CA, 90095, e-mail: poweill@econ.ucla.edu. First version: July 2003. This is the first chapter of my Stanford PhD dissertation. I am deeply indebted to Darrell Duffie and Tom Sargent, for their supervision, their encouragements, many detailed comments and suggestions. I also thank Narayana Kocherlakota for fruitful discussions and suggestions. I benefited from comments by Manuel Amador, Andy Atkeson, Marco Bassetto, Bruno Biais, Vinicius Carrasco, John Y. Campbell, William Fuchs, Xavier Gabaix, Ed Green, Bob Hall, Ali Hortaçsu, Steve Kohlhagen, Arvind Krishnamurthy, Hanno Lustig, Erzo Luttmer, Eva Nagypal, Lasse Heje Pedersen, Esteban Rossi-Hansberg, Tano Santos, Carmit Segal, Stijn Van Nieuwerburgh, François Velde, Tuomo Vuolteenaho, Ivan Werning, Randy Wright, Mark Wright, Bill Zame, Ruilin Zhou, participants of Tom Sargent's reading group at the University of Chicago, Stanford University 2003 SITE conference, and of seminar at Stanford University, NYU Economics and Stern, UCLA Anderson, Columbia GSB, Harvard University, University of Pennsylvania, University of Michigan Finance, MIT Economics, the University of Minnesota, the University of Chicago, Northwestern University Economics and Kellogg, the University of Texas at Austin, the Federal Reserve Bank of Chicago, the Federal Reserve Bank of Cleveland, UCLA Economics, the Federal Reserve Bank of Atlanta, THEMA, and MIT Sloan. The financial support of the Kohlhagen Fellowship Fund at Stanford University is gratefully acknowledged. I am grateful to Andrea Prat (the editor) and two anonymous referees for comments that improved the paper. All errors are mine.

1 Introduction

When disruptions subject financial markets to unusually strong selling pressures, NYSE specialists and NASDAQ marketmakers typically *lean against the wind* by absorbing the market’s selling pressure and creating liquidity: they buy large quantity of assets and build up inventories when selling pressure in the market is large, then dispose of those inventories after that selling pressure has subsided.¹ In this paper, I develop a model of optimal dynamic liquidity provision. To explain how much and when liquidity should be provided, I solve for socially optimal liquidity provision. I argue that some features of the socially optimal allocation would be regarded by a policymaker as symptoms of poor liquidity provision. In fact, these symptoms can be consistent with efficiency. I also show that when they can maintain sufficient capital, competitive marketmakers supply the socially optimal amount of liquidity. If capital-market imperfections prevent marketmakers from raising sufficient capital, this suggests a policy role for lenient central-bank lending during financial disruptions.

The model studies the following scenario. In the beginning at time zero, outside investors receive an aggregate shock which lowers their marginal utility for holding assets relative to cash. This creates a sudden need for cash and induces a large selling pressure. Then, randomly over time, each investor recovers from the shock, implying that the initial selling pressure slowly alleviates. This is how I create a stylized representation of a “flight-to-liquidity” (Longstaff [2004]) or a stock-market crash such as that of October 1987. All trades are intermediated by marketmakers who do not derive any utility for holding assets and who are located in a central marketplace which can be viewed, say, as the floor of the New-York Stock Exchange. I assume that the asset market can be illiquid in the sense that investors make contact with marketmakers only after random delays. This means that, at each time, only a fraction of investors can trade, which effectively imposes an upper limit on the fraction of outside orders marketmakers can execute per unit of time. The random delays are designed to represent, for example, front-end order capture, clearing, and settlement. While one expects such delays to be short in normal times, the Brady [1988] report suggests that they were unusually long and variable during the crash of October 1987. Similarly, during the crash of October 1997, customers complained about “poor or untimely execution from broker dealers” (SEC Staff Legal Bulletin No.8 of September 9, 1998). Lastly, McAndrews and Potter [2002] and Fleming and Garbade [2002] document payment and transaction delays, due to disruption of the communication network after the terrorist attacks of September 11, 2001.

¹This behavior reflects one aspect of the U.S. Securities and Exchange Commission (SEC) Rule 11-b on maintaining fair and orderly markets.

In this economic environment, marketmakers offer buyers and sellers quicker exchange, what [Demsetz \[1968\]](#) called “immediacy”. Marketmakers anticipate that after the selling pressure subsides, they will achieve contact with more buyers than sellers, which will allow them then to transfer assets to buyers in two ways. They can either contact additional sellers, which is time-consuming because of execution delays; or they can sell from their own inventories, which can be done much more quickly. Therefore, by accumulating inventories early, when the selling pressure is large, marketmakers mitigate the adverse impact on investors of execution delays.

The socially optimal asset allocation maximizes the sum of investors’ and marketmakers’ intertemporal utility, subject to the order-execution technology. Because agents have quasi-linear utilities, any other asset allocation could be Pareto improved by reallocating assets and making time-zero consumption transfers. The upper panel of [Figure 1](#) shows the socially optimal time path of marketmakers’ inventory. (The associated parameters and modelling assumptions are described in [Section 2](#).) The graph shows that marketmakers accumulate inventories only temporarily, when the selling pressure is large. Moreover, in this example, it is not socially optimal that marketmakers start accumulating inventories at time zero when the pressure is strongest. This suggests that a regulation forcing marketmakers to promptly act as “buyers of last resort” could in fact result in a welfare loss. For example, if the initial preference shock is sufficiently persistent, marketmakers acting as buyers of last resort will end up holding assets for a very long time, which cannot be efficient given that they are not the final holder of the asset. Lastly, when the economy is close to its steady state (interpreted as a “normal time”) marketmakers should effectively act as “matchmakers” who never hold assets but merely buy and re-sell instantly.

If marketmakers maintain sufficient capital, I show that the socially optimal allocation is implemented in a competitive equilibrium, as follows. Investors can buy and sell assets only when they contact marketmakers. Marketmakers compete for the order flow and can trade among each other at each time. The lower panel of [Figure 1](#) shows the equilibrium price path. It jumps down at time zero, then increases, and eventually reaches its steady-state level. A marketmaker finds it optimal to accumulate inventories only temporarily, when the asset price grows at a sufficiently high rate. This growth rate compensates for the time value of the money spent on inventory accumulation, giving a marketmaker just enough incentive to provide liquidity. A marketmaker thus buys early at a low price and sells later at a high price, but competition implies that the present value of her profit is zero.

Ample anecdotal evidence suggests that marketmakers do not maintain sufficient capital ([Brady \[1988\]](#), [Greenwald and Stein \[1988\]](#), [Marès \[2001\]](#), and [Greenberg \[2003\]](#).) I find that

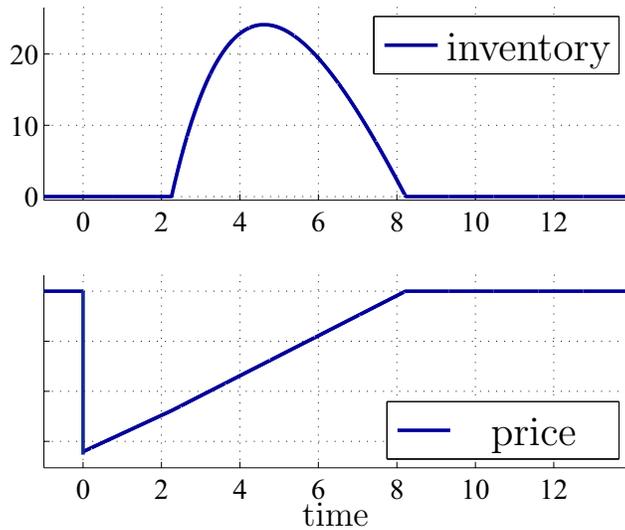


Figure 1: Features of the Competitive Equilibrium.

if marketmakers do not maintain sufficient capital, then they are not able to purchase as many assets as prescribed by the socially optimal allocation. If capital-market imperfections prevent marketmakers from raising sufficient capital before the crash, lenient central-bank lending during the crash can improve welfare. Recall that during the crash of October 1987, the Federal Reserve lowered the funds rate while encouraging commercial banks to lend to security dealers (Parry [1997], Wigmore [1998].)

It is often argued that marketmakers should provide liquidity in order to maintain price continuity and to smooth asset price movements.² The present paper steps back from such price-smoothing objective and instead studies liquidity provision in terms of the Pareto criterion. The results indicate that Pareto-optimal liquidity provision is consistent with a discrete price decline at the time of the crash. This suggests that requiring marketmakers to maintain price continuity at the time of the crash might result in a welfare loss.

Related Literature

Liquidity provision in normal times has been analyzed in traditional inventory-based models of marketmaking (see Chapter 2 of O’Hara [1995] for a review). Because they study inventory management in normal times, these models assume exogenous, time-invariant supply and demand curves. The present paper, by contrast, derives time-varying supply and demand

²For instance, the glossary of www.nyse.com states that NYSE specialists “use their capital to bridge temporary gaps in supply and demand and help reduce price volatility.” See also the NYSE information memo 97-55.

curves from the solutions of investors' inter-temporal utility maximization problems. This allows to address the welfare impact of liquidity provision under unusual market conditions. Another difference with this literature is that I study the impact of scarce marketmaking capital on marketmakers' profit and price dynamics.

In [Grossman and Miller \[1988\]](#) and [Greenwald and Stein \[1991\]](#), the social benefit of marketmakers' liquidity provision is to share risk with sellers before the arrival of buyers. In the present model, by contrast, the social benefit of liquidity provision is to facilitate trade, in that it speeds up the allocation of assets from the initial sellers to the later buyers. Moreover Grossman and Miller study a two-period model, which means that the timing of liquidity provision is effectively exogenous. With its richer intertemporal structure, my model sheds light on the optimal timing of liquidity provision.

[Bernardo and Welch \[2004\]](#) explain a financial-market crisis in a two-period model, along the line of [Diamond and Dybvig \[1983\]](#), and they study the liquidity provision of myopic marketmakers. The main objective of the present paper is not to explain the cause of a crisis, but rather to develop an inter-temporal model of marketmakers optimal liquidity provision, after an aggregate liquidity shock.

Search-and-matching models of financial markets study the impact of trading delays in security markets (see, for instance, [Duffie et al. \[2005\]](#), [Weill \[2004\]](#), [Vayanos and Wang \[2006\]](#), [Vayanos and Weill \[2006\]](#), [Lagos \[2006\]](#), [Spulber \[1996\]](#) and [Hall and Rust \[2003\]](#).) The present model builds specifically on the work of Duffie, Gârleanu and Perdersen. In their model, marketmakers are matchmakers who, by assumption, cannot hold inventory. By studying investment in marketmaking capacity, they focus on liquidity provision in the long run. By contrast, I study liquidity provision in the short run and view marketmaking capacity as a fixed parameter. In the short run, marketmakers provide liquidity by adjusting their inventory positions.

Another related literature studies the equilibrium size of the middlemen sector in search-and-matching economies, and provides steady-states in which the aggregate amount of middlemen's inventories remains constant over time (see, among others, [Rubinstein and Wolinsky \[1987\]](#), [Li \[1998\]](#), [Shevchenko \[2004\]](#), and [Masters \[2004\]](#)). The present paper studies intermediation during a financial crisis, when it is arguably reasonable to take the size of the marketmaking sector as given. In the short run, the marketmaking sector can only gain capacity by increasing its capital and aggregate inventories fluctuate over time.

The remainder of this paper is organized as follows. Section 2 describes the economic environment, Section 3 solves for socially optimal dynamic liquidity provision, Section 4 studies the implementation of this optimum in a competitive equilibrium, and introduces

borrowing-constrained marketmakers. Section 5 discusses policy implications, and Section 6 concludes. The appendix contains the proofs.

2 The Economic Environment

This section describes the economy and introduces the two main assumptions of this paper. First, there is a large and temporary selling pressure. Second, there are order-execution delays.

2.1 Marketmakers and Investors

Time is treated continuously, and runs forever. A probability space (Ω, \mathcal{F}, P) is fixed, as well as an information filtration $\{\mathcal{F}_t, t \geq 0\}$ satisfying the usual conditions (Protter [1990]). The economy is populated by a non-atomic continuum of infinitely lived and risk-neutral agents who discount the future at the constant rate $r > 0$. An agent enjoys the consumption of a non-storable numéraire good called “cash,” with a marginal utility normalized to 1.³

There is one asset in positive supply. An agent holding q units of the asset receives a stochastic utility flow $\theta(t)q$ per unit of time. Stochastic variations in the marginal utility $\theta(t)$ capture a broad range of trading motives such as changes in hedging needs, binding borrowing constraints, changes in beliefs, or risk-management rules such as risk limits. There are two types of agents, marketmakers and investors, with a measure one (without loss of generality) of each. Marketmakers and investors differ in their marginal-utility processes $\{\theta(t), t \geq 0\}$, as follows. A marketmaker has a constant marginal utility $\theta(t) = 0$ while an investor’s marginal utility is a two-state Markov chain: the high-marginal-utility state is normalized to $\theta(t) = 1$, and the low-marginal-utility state is $\theta(t) = 1 - \delta$, for some $\delta \in (0, 1)$. Investors transit randomly, and pair-wise independently, from low to high marginal utility with intensity⁴ γ_u , and from high to low marginal utility with intensity γ_d .

These independent variations over time in investors’ marginal utilities create gains from trade. A low-marginal-utility investor is willing to sell his asset to a high-marginal-utility investor in exchange for cash. A marketmaker’s zero marginal utility could capture a large exposure to the risk of the market she intermediates. In addition, it implies that in the equilibrium to be described, a marketmaker will not be the final holder of the asset. In particular, a marketmaker would choose to hold assets only because she expects to make

³Equivalently, one could assume that agents can borrow and save cash in some “bank account,” at the interest rate $\bar{r} = r$. Section 4 adopts this alternative formulation.

⁴For instance, if $\theta(t) = 1 - \delta$, the time $\inf\{u \geq 0 : \theta(t+u) \neq \theta(t)\}$ until the next switch is exponentially distributed with parameter γ_u . The successive switching times are independent.

some profit by buying and reselling.⁵

Asset Holdings

The asset has $s \in (0, 1)$ shares outstanding per investor’s capita. Marketmakers can hold any positive quantity of the asset. The time t asset inventory $I(t)$ of a representative marketmaker satisfies the short-selling constraint⁶

$$I(t) \geq 0. \tag{1}$$

An investor also cannot short-sell and, moreover, he cannot hold more than one unit of the asset. This paper restricts attention to allocations in which an investor holds either zero or one unit of the asset. In equilibrium, because an investor has linear utility, he will find it optimal to hold either the maximum quantity of one or the minimum quantity of zero.

An investor’s type is made up of his marginal utility (high “ h ,” or low “ ℓ ”) and his ownership status (owner of one unit, “ o ,” or non-owner, “ n ”). The set of investors’ types is $\mathcal{T} \equiv \{\ell o, h n, h o, \ell n\}$. In anticipation of their equilibrium behavior, low-marginal-utility owners (ℓo) are named “sellers,” and high-marginal-utility non-owners ($h n$) are “buyers.” For each $\sigma \in \mathcal{T}$, $\mu_\sigma(t)$ denotes the fraction of type- σ investors in the total population of investors. These fractions must satisfy two accounting identities. First, of course,

$$\mu_{\ell o}(t) + \mu_{h n}(t) + \mu_{\ell n}(t) + \mu_{h o}(t) = 1. \tag{2}$$

Second, the assets are held either by investors or marketmakers, so

$$\mu_{h o}(t) + \mu_{\ell o}(t) + I(t) = s. \tag{3}$$

2.2 Crash and Recovery

I select initial conditions representing the strong selling pressure of a financial disruption. Namely, it is assumed that, at time zero, all investors are in the low-marginal-utility state (see Table 1). Then, as earlier specified, investors transit to the high-marginal-utility state. Under suitable measurability requirements (see Sun [2000], Theorem C), the law of large

⁵The results of this paper hold under the weaker assumption that marketmakers’ marginal utility is $\theta(t) = 1 - \delta_M$, for some holding cost $\delta_M > \delta$. Proofs are available from the author upon request.

⁶The short-selling constraint means that both marketmakers and investors face an infinite cost of holding a negative asset position. This may be viewed as a strong assumption because it is typically easier for marketmakers to go short than for investors. However, one can show that, in the present setup, marketmakers find it optimal to choose $I(t) \geq 0$, as long as they incur a finite but sufficiently large cost $c > 1 - \delta\gamma_d/(r + \rho + \gamma_u + \gamma_d)$ per unit of negative inventory. This cost could capture, for example, the fact that short positions are more risky than long positions.

numbers applies, and the fraction $\mu_h(t) \equiv \mu_{ho}(t) + \mu_{hn}(t)$ of high-marginal-utility investors solves the ordinary differential equation (ODE)

$$\begin{aligned}\dot{\mu}_h(t) &= \gamma_u(\mu_{\ell o}(t) + \mu_{\ell n}(t)) - \gamma_d(\mu_{ho}(t) + \mu_{hn}(t)) = \gamma_u(1 - \mu_h(t)) - \gamma_d\mu_h(t) \\ &= \gamma_u - \gamma\mu_h(t),\end{aligned}\tag{4}$$

where $\dot{\mu}_h(t) = d\mu_h(t)/dt$ and $\gamma \equiv \gamma_u + \gamma_d$. The first term in (4) is the rate of flow of low-marginal-utility investors transiting to the high-marginal-utility state, while the second term is the rate of flow of high-marginal-utility investors transiting to the low-marginal-utility state. With the initial condition $\mu_h(0) = 0$, the solution of (4) is

$$\mu_h(t) = y(1 - e^{-\gamma t}),\tag{5}$$

where $y \equiv \gamma_u/\gamma$ is the steady-state fraction of high-marginal-utility investors. Importantly for the remainder of the paper, it is assumed that

$$s < y.\tag{6}$$

In other words, in steady state, the fraction y of high-marginal-utility investors exceeds the asset supply s . This will ensure that, asymptotically in equilibrium, the selling pressure has fully alleviated. Figure 2 plots the time dynamic of $\mu_h(t)$, for some parameter values that satisfy (6). On the Figure, the unit of time is one hour. Years are converted into hours assuming 250 trading days per year, and 10 hours of trading per days. The parameter values used for all of the illustrative computations of this paper, are in Table 2.

Table 1: Initial conditions.

$\mu_{\ell o}(0)$	$\mu_{hn}(0)$	$\mu_{\ell n}(0)$	$\mu_{ho}(0)$	$I(0)$
s	0	$1 - s$	0	0

2.3 Order-execution delays

Marketmakers intermediate all trades from a central marketplace which can be viewed, say, as the floor of the New York Stock exchange. This market, however, is illiquid in the sense that investors cannot contact that marketplace instantly. Instead, an investor establishes contact with marketmakers at Poisson arrival times with intensity $\rho > 0$. Contact times are pairwise independent across investors and independent of marginal utility processes.⁷ Therefore, an

⁷These random contact time provide a simple way to formalize [Biais et al. \[2005\]](#)'s view that "only small subset of all economic agents become full-time traders and stand ready to accommodate the trading needs of the rest of the population."

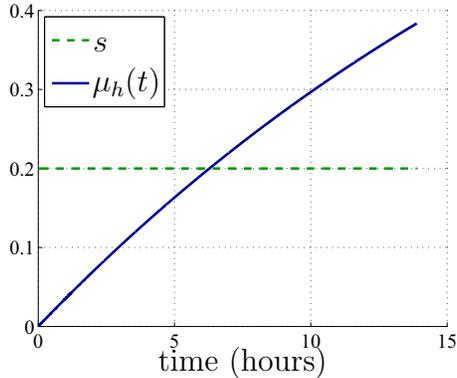


Figure 2: Dynamic of $\mu_h(t)$.

Table 2: Parameter Values.

Parameters		Value
Measure of Shares	s	0.2
Discount Rate	r	5%
Contact Intensity	ρ	1000
Intensity of Switch to High	γ_u	90
Intensity of Switch to Low	γ_d	10
Low marginal utility	$1 - \delta$	0.01

Time is measured in years. Assuming that the stock market opens 250 days a year, $\rho = 1000$ means that it takes 2.5 hours to execute an order, on average. The parameter $\gamma = \gamma_u + \gamma_d$ measures the speed of the recovery. Specifically, with $\gamma = 100$, $\mu_h(t)$ reaches half of its steady-state level in about 1.73 days.

application of the law of large numbers (under the technical conditions mentioned earlier) implies that contacts between type- σ investors and marketmakers occur at a total (almost sure) rate of $\rho\mu_\sigma(t)$. Hence, during a small time interval of length ε , marketmakers can only execute a fixed fraction $\rho\varepsilon$ of randomly chosen orders originating from type- σ investors.⁸

The random contact times represent a broad range of execution delays, including the time to contact a marketmaker, to negotiate and process an order, to deliver an asset, or to transfer a payment. One might argue that such execution delays are usually quite short and perhaps therefore of little consequence to the quality of an allocation. The Brady [1988] report shows, however, that during the October 1987 crash, overloaded execution systems created

⁸Instead of imposing a limit on the *fraction* of orders that marketmakers can execute per unit of time, one could impose a limit on the total *number* of orders they can execute. One can show that, under such an alternative specification, competitive marketmakers also provide the socially optimal amount of liquidity.

delays that were much longer and variable than in normal times. It suggests that these delays might have amplified liquidity problems in a far-from-negligible manner. Although the trading technology improved after the crash of October 1987, substantial execution delays also occurred during the crash of October 1997. The SEC reported that “broker-dealers web servers had reached their maximum capacity to handle simultaneous users” and “telephone lines were overwhelmed with callers who were frustrated by the inability to access information online.” As a result of these capacity problems, customers could not be “routed to their designated market center for execution on a timely basis” and “a number of broker dealers were forced to manually execute some customers orders.”⁹

3 Optimal Dynamic Liquidity Provision

The first objective of this section is to explain the benefit of liquidity provision, addressing how much and when liquidity should be provided. Its second objective is to establish a benchmark against which to judge the market equilibria studied in Sections 4 and 4.2. To these ends, I temporarily abstract from marketmakers’ incentives to provide liquidity and solve for socially optimal allocations, maximizing the sum of investors and marketmakers’ intertemporal utility, subject to order-execution delays. The optimal allocation is found to resemble “leaning against the wind.” Namely, it is socially optimal that a marketmaker accumulates inventories when the selling pressure is strong.

3.1 Asset Allocations

At each time, a representative marketmaker can transfer assets only to her own account or among those of investors who are currently contacting her. For instance, the flow rate $u_\ell(t)$ of assets that a marketmaker takes from low-marginal-utility investors is subject to the order-flow constraint

$$-\rho\mu_{\ell n}(t) \leq u_\ell(t) \leq \rho\mu_{\ell o}(t). \tag{7}$$

The upper (lower) bound shown in (7) is the flow of ℓo (ℓn) investors who establish contact with marketmakers at time t . Similarly, the flow $u_h(t)$ of assets that a marketmaker transfers to high-marginal-utility investors is subject to the order-flow constraint

$$-\rho\mu_{ho}(t) \leq u_h(t) \leq \rho\mu_{hn}(t). \tag{8}$$

⁹SEC Staff Legal Bulletin No.8, <http://www.sec.gov/interp/legals/slbmr8.htm>

When the two flows $u_\ell(t)$ and $u_h(t)$ are equal, a marketmaker is a matchmaker, in the sense that she takes assets from some ℓo investors (sellers) and transfers them instantly to some hn investors (buyers). If the two flows are not equal, a marketmaker is not only matching buyers and sellers, but she is also changing her inventory position. For example, if both $u_\ell(t)$ and $u_h(t)$ are positive, a marketmaker is matching sellers and buyers at the rate $\min\{u_\ell(t), u_h(t)\}$. The net flow $u_\ell(t) - u_h(t)$ represents the rate of change of a marketmaker's inventory, in that

$$\dot{I}(t) = u_\ell(t) - u_h(t). \quad (9)$$

Similarly, the rate of change of the fraction $\mu_{\ell o}(t)$ of low-marginal-utility owners is

$$\dot{\mu}_{\ell o}(t) = -u_\ell(t) - \gamma_u \mu_{\ell o}(t) + \gamma_d \mu_{ho}(t), \quad (10)$$

where the terms $\gamma_u \mu_{\ell o}(t)$ and $\gamma_d \mu_{ho}(t)$ reflect transitions of investors from low to high marginal utility, and from high to low marginal utility, respectively. Likewise, the rate of change of the fractions of hn , ℓn , and ho investors are, respectively,

$$\dot{\mu}_{hn}(t) = -u_h(t) - \gamma_d \mu_{hn}(t) + \gamma_u \mu_{\ell n}(t) \quad (11)$$

$$\dot{\mu}_{\ell n}(t) = u_\ell(t) - \gamma_u \mu_{\ell n}(t) + \gamma_d \mu_{hn}(t) \quad (12)$$

$$\dot{\mu}_{ho}(t) = u_h(t) - \gamma_d \mu_{ho}(t) + \gamma_u \mu_{\ell o}(t). \quad (13)$$

Definition 1 (Feasible Allocation). *A feasible allocation is some distribution $\mu(t) \equiv (\mu_\sigma(t))_{\sigma \in \mathcal{T}}$ of types, some inventory holding $I(t)$, and some piecewise continuous asset flows $u(t) \equiv (u_h(t), u_\ell(t))$ such that*

(i) *At each time, the short-selling constraint (1) and the order-flow constraints (7)-(8) are satisfied.*

(ii) *The ODEs (9)-(13) hold.*

(iii) *The initial conditions of Table 1 hold.*

Since $u(t)$ is piecewise continuous, $\mu(t)$ and $I(t)$ are piecewise continuously differentiable. A feasible allocation is said to be constrained Pareto optimal if it cannot be Pareto improved by choosing another feasible allocation and making time-zero cash transfers. As it is standard with quasi-linear preferences, it can be shown that a constrained Pareto optimal allocation must maximize

$$\int_0^{+\infty} e^{-rt} \left(\mu_{ho}(t) + (1 - \delta) \mu_{\ell o}(t) \right) dt, \quad (14)$$

the equally weighted sum of investors' intertemporal utilities for holding assets.¹⁰ This criterion is deterministic, reflecting pairwise independence of investors' marginal-utility and contact-time processes. Conversely, an asset allocation maximizing (14) is constrained Pareto optimal. This discussion motivates the following definition of an optimal allocation.

Definition 2 (Socially Optimal Allocation). *A socially optimal allocation is some feasible allocation maximizing (14).*

3.2 The Benefit of Liquidity Provision

This subsection illustrates the social benefits of accumulating inventories. Namely, it considers the no-inventory allocation ($I(t) = 0$, at each time), and shows that it can be improved if marketmakers accumulate a small amount of inventory, when the selling pressure is strong. I start by describing some features of the no-inventory allocation. Substituting $I(t) = 0$ into equation (3) gives

$$\mu_{\ell o}(t) = s - \mu_h(t) + \mu_{hn}(t). \quad (15)$$

The “crossing time” is the time t_s at which $\mu_h(t_s) = s$. This is, as Figure 2 illustrates, the time at which the fraction $\mu_h(t)$ of high-marginal-utility investors crosses the supply s of assets. Because $\mu_h(t)$ is increasing, equation (15) implies that

$$\rho\mu_{hn}(t) < \rho\mu_{\ell o}(t) \quad (16)$$

if and only if $t < t_s$. Therefore, in the no-inventory allocation, before the crossing time, the selling pressure is “positive,” meaning that marketmakers are in contact with more sellers (ℓo) than buyers (hn). After the crossing time, they are in contact with more buyers than sellers.

Intuitively, the no-inventory allocation can be improved as follows. A marketmaker can take an additional asset from a seller before the crossing time, say at $t_1 = t_s - \varepsilon$, and transfer it to some buyer after the crossing time, at $t_2 = t_s + \varepsilon$. Because the transfer occurs around the crossing time, the transfer time 2ε can be made arbitrarily small.

The benefit is that, for a sufficiently small ε , this asset is allocated almost instantly to some high-marginal-utility investor. Without the transfer, by contrast, this asset would continue to be held by a low-marginal-utility investor until either i) the seller transits to a high marginal utility with intensity γ_u , or ii) the seller establishes another contact with

¹⁰Marketmakers intertemporal utility for holding assets is equal to zero and hence does not appear in (14). If a marketmaker marginal utility for holding asset is $(1 - \delta_M) > 0$, then one has to add a term $\int_0^\infty e^{-rt}(1 - \delta_M)I(t) dt$ to the above criterion.

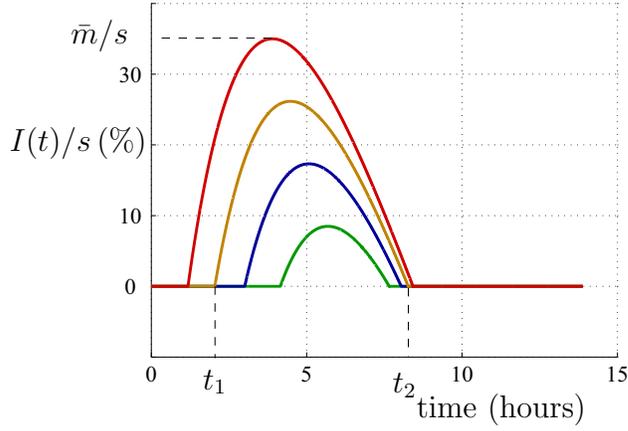


Figure 3: Illustrative Buffer Allocations.

a marketmaker with intensity ρ . This means that, without the transfer, this asset would continue to be held by a seller and not by a buyer, with an instantaneous utility cost of δ , incurred for a non-negligible average time of $1/(\gamma_u + \rho)$.

The cost of the transfer is that the asset is temporarily held by a marketmaker and not by a seller, implying an instantaneous utility cost of $1 - \delta$. If ε is sufficiently small, this cost is incurred for a negligible time and is smaller than the benefit. This intuitive argument can be formalized by studying the following family of feasible allocations.

Definition 3 (Buffer Allocation). *A buffer allocation is a feasible allocation defined by two times $(t_1, t_2) \in [0, t_s] \times [t_s, +\infty)$, called “breaking times,” such that*

$$\begin{aligned}
 u_\ell(t) &= \rho\mu_{hn}(t) & \text{and} & & u_h(t) &= \rho\mu_{hn}(t) & t \in [0, t_1) \\
 u_\ell(t) &= \rho\mu_{\ell o}(t) & \text{and} & & u_h(t) &= \rho\mu_{hn}(t) & t \in [t_1, t_2] \\
 u_\ell(t) &= \rho\mu_{\ell o}(t) & \text{and} & & u_h(t) &= \rho\mu_{\ell o}(t), & t \in (t_2, \infty],
 \end{aligned}$$

and $I(t_2) = 0$.

The no-inventory allocation is the buffer allocation for which $t_1 = t_2 = t_s$. A buffer allocation has the “bang-bang” property: at each time, either $u_\ell(t) = \rho\mu_{\ell o}(t)$ or $u_h(t) = \rho\mu_{hn}(t)$. Because of the linear objective (14), it is natural to guess that a socially optimal allocation will also have this bang-bang property. In the next subsection, Theorem 1 will confirm this conjecture, showing that the socially optimal allocation belongs to the family of buffer allocations.

In a buffer allocation, a marketmaker acts as a “buffer,” in that she accumulates assets when the selling pressure is strong and unwinds these trades when the pressure alleviates.

Specifically, as illustrated in Figure 3, a buffer allocation (t_1, t_2) has three phases. In the first phase, when $t \in [0, t_1]$, a marketmaker does not accumulate inventory ($u_\ell(t) = u_h(t)$ and $I(t) = 0$). In the second phase, when $t \in (t_1, t_2)$, a marketmaker first builds up ($u_\ell(t) > u_h(t)$ and $I(t) > 0$) and then unwinds ($u_\ell(t) < u_h(t)$ and $I(t) > 0$) her inventory position. At time t_2 , her inventory position reaches zero. In the third phase $t \in [t_2, +\infty)$, a marketmaker does not accumulate inventory ($u_\ell(t) = u_h(t)$ and $I(t) = 0$). The following proposition characterizes buffer allocations by the maximum inventory position held by marketmakers.

Proposition 1. *There exist some $\bar{m} \in \mathbb{R}_+$, some strictly decreasing continuous function $\psi : [0, \bar{m}] \rightarrow \mathbb{R}_+$, and some strictly increasing continuous functions $\phi_i : [0, \bar{m}] \rightarrow \mathbb{R}_+$, $i \in \{1, 2\}$, such that, for all $m \in [0, \bar{m}]$ and all buffer allocations (t_1, t_2) ,*

$$m = \max_{t \in \mathbb{R}_+} I(t) \quad \text{and} \quad \psi(m) = \arg \max_{t \in \mathbb{R}_+} I(t) \quad (17)$$

$$t_1 = \psi(m) - \phi_1(m) \quad \text{and} \quad t_2 = \psi(m) + \phi_2(m), \quad (18)$$

where \bar{m} is the unique solution of $\psi(z) - \phi_1(z) = 0$. Furthermore, $\psi(0) = t_s$ and $\phi_1(0) = \phi_2(0) = 0$.

In words, the breaking times (t_1, t_2) of a buffer allocation can be written as functions of the maximum inventory position m . The maximum inventory position is achieved at time $\psi(m)$. In addition, the larger is a marketmaker's maximum inventory position, the earlier she starts to accumulate and the longer she accumulates. Lastly, if she starts to accumulate at time zero, then her maximum inventory position is \bar{m} .

The social welfare (14) associated with a buffer allocation can be written as $W(m)$, for some function $W(\cdot)$ of the maximum inventory position m . As anticipated by the intuitive argument, one can prove the following result.

Proposition 2.

$$\lim_{m \rightarrow 0^+} \frac{W(m) - W(0)}{m} > 0. \quad (19)$$

This demonstrates that the no-inventory allocation ($m = 0$) is improved by accumulating a small amount of inventory near the crossing time t_s .

3.3 The Socially Optimal Allocation

Having shown that accumulating some inventory improves welfare, this section explains how much inventory marketmakers should accumulate. Namely, it provides first-order sufficient conditions for, and solves for, a socially optimal allocation. The reader may wish to skip the

following paragraph on first-order conditions, and go directly to Theorem 1, which describes the socially optimal allocation.

First-Order Sufficient Conditions

The first-order sufficient conditions are based on Seierstad and Sydsæter [1977]. The accounting identities $\mu_{ho}(t) = \mu_h(t) - \mu_{hn}(t)$ and $\mu_{ln}(t) = 1 - \mu_h(t) - \mu_{ln}(t)$ are substituted into the objective and the constraints, reducing the state variables to $(\mu_{\ell o}(t), \mu_{hn}(t), I(t))$. The “current-value” Lagrangian (see Kamien and Schwartz [1991], Part II, Section 8) is

$$\begin{aligned} \mathcal{L}(t) = & \mu_h(t) - \mu_{hn}(t) + (1 - \delta)\mu_{\ell o}(t) & (20) \\ & + \lambda_\ell(t) (-u_\ell(t) - \gamma_u\mu_{\ell o}(t) - \gamma_d\mu_{hn}(t) + \gamma_d\mu_h(t)) \\ & - \lambda_h(t) (-u_h(t) - \gamma_u\mu_{\ell o}(t) - \gamma_d\mu_{hn}(t) + \gamma_u(1 - \mu_h(t))) \\ & + \lambda_I(t) (u_\ell(t) - u_h(t)) \\ & + w_\ell(t) (\rho\mu_{\ell o}(t) - u_\ell(t)) + w_h(t) (\rho\mu_{hn}(t) - u_h(t)) + \eta_I(t) I(t). \end{aligned}$$

The multiplier $\lambda_\ell(t)$ of the ODE (10) represents the social value of increasing the flow of investors from the ln type to the ℓo type or, equivalently, the value of transferring an asset to an ln investor. One gives a similar interpretation to the multipliers $\lambda_h(t)$ and $\lambda_I(t)$ of the ODEs (11) and (9), respectively.¹¹ The multipliers $w_\ell(t)$ and $w_h(t)$ of the flow constraints (7) and (8) represent the social value of increasing the rate of contact with ℓo and hn investors, respectively.¹² The multiplier on the short-selling constraint (1) is $\eta_I(t)$. The first-order condition with respect to the controls $u_\ell(t)$ and $u_h(t)$ are

$$w_\ell(t) = \lambda_I(t) - \lambda_\ell(t) \tag{21}$$

$$w_h(t) = \lambda_h(t) - \lambda_I(t), \tag{22}$$

respectively. For instance, (21) decomposes $w_\ell(t)$ into the opportunity cost $-\lambda_\ell(t)$ of taking assets from ℓo investors, and the benefit $\lambda_I(t)$ of increasing a marketmaker’s inventory. The positivity and complementary-slackness conditions for $w_\ell(t)$ and $w_h(t)$, respectively, are

$$w_\ell(t) \geq 0 \quad \text{and} \quad w_\ell(t)(\rho\mu_{\ell o}(t) - u_\ell(t)) = 0, \tag{23}$$

$$w_h(t) \geq 0 \quad \text{and} \quad w_h(t)(\rho\mu_{hn}(t) - u_h(t)) = 0. \tag{24}$$

The multipliers $w_\ell(t)$ and $w_h(t)$ are non-negative because a marketmaker can ignore additional contacts. The complementary-slackness condition (23) means that, when the marginal

¹¹In equation (20), the minus sign in front of $\lambda_h(t)$ is contrary to conventional notations but turns out to simplify the exposition.

¹²It is anticipated that the left-hand constraints in (7) and (8) never bind. In other words, a marketmaker never transfers asset from a high-marginal-utility to a low-marginal-utility investor.

value $w_\ell(t)$ of additional contact is strictly positive, a marketmaker should take the assets of all ℓo investors with whom she is currently in contact. One also has the positivity and complementary-slackness conditions

$$\eta_I(t) \geq 0 \quad \text{and} \quad \eta_I(t)I(t) = 0. \quad (25)$$

The ODE for the the multipliers $\lambda_\ell(t)$, $\lambda_h(t)$, and $\lambda_I(t)$ are

$$r\lambda_\ell(t) = 1 - \delta + \gamma_u(\lambda_h(t) - \lambda_\ell(t)) + \rho w_\ell(t) + \dot{\lambda}_\ell(t) \quad (26)$$

$$r\lambda_h(t) = 1 + \gamma_d(\lambda_\ell(t) - \lambda_h(t)) - \rho w_h(t) + \dot{\lambda}_h(t) \quad (27)$$

$$r\lambda_I(t) = \eta_I(t) + \dot{\lambda}_I(t), \quad (28)$$

respectively. For instance, (26) decomposes the flow value $r\lambda_\ell(t)$ of transferring an asset to a low-marginal-utility investor. The first term, $1 - \delta$, is the flow marginal utility of a low-marginal-utility investor holding one unit of the asset. The second term, $\gamma_u(\lambda_h(t) - \lambda_\ell(t))$, is the expected rate of net utility associated with a transition to high marginal utility. That is, with intensity γ_u , $\lambda_\ell(t)$ becomes the value $\lambda_h(t)$ of transferring an asset to a high-marginal-utility investor. The third term, $\rho w_\ell(t)$, is the expected rate of net utility of a contact between an ℓo investor and a marketmaker. The multipliers $(\lambda_\ell(t), \lambda_h(t), \lambda_I(t))$ must satisfy the following additional restrictions. First, they must satisfy the transversality conditions that $\lambda_\ell(t)e^{-rt}$, $\lambda_h(t)e^{-rt}$, and $\lambda_I(t)e^{-rt}$ go to zero as time goes to infinity. Second, the multipliers $\lambda_h(t)$ and $\lambda_\ell(t)$ are continuous. Because the control variable $u(t)$ does not appear in the short-selling constraint $I(t) \geq 0$, however, the multiplier $\lambda_I(t)$ might jump, with the restriction that

$$\lambda_I(t^+) - \lambda_I(t^-) \leq 0 \quad \text{if} \quad I(t) = 0. \quad (29)$$

In other words, the multiplier $\lambda_I(t)$ can jump down, but only when the short-selling constraint is binding. Intuitively, if $\lambda_I(t)$ were to jump up at t , a marketmaker could accumulate additional inventory shortly before t , say a quantity ε , improving the objective by $\varepsilon(\lambda_I(t^+) - \lambda_I(t^-))e^{-rt}$.¹³

The Socially Optimal Allocation

Appendix B guesses and verifies that the (essentially unique) socially optimal allocation is a buffer allocation. Namely, for a given buffer allocation, one constructs multipliers solving

¹³Because the ODEs for the state variables are linear, the first-order sufficient conditions impose no sign restriction on the multipliers $\lambda_h(t)$, $\lambda_\ell(t)$, and $\lambda_I(t)$ (see Section 3 in Part II of [Kamien and Schwartz \[1991\]](#)).

the first-order conditions (21) through (29). The restriction $w_\ell(t_1) = 0$ is used to find the breaking-times t_1 and t_2 .

Theorem 1 (Socially optimal Allocation). *There exists a socially optimal allocation $(\mu^*(t), I^*(t), u^*(t), t \geq 0)$. This allocation is a buffer allocation with breaking times (t_1^*, t_2^*) determined by*

$$e^{-\gamma t_1^*} = \left(1 - \frac{s}{y}\right) \frac{1 - e^{-\rho \Delta^*}}{\rho} \frac{\rho - \gamma}{e^{-\gamma \Delta^*} - e^{-\rho \Delta^*}} \quad (30)$$

$$t_2^* = t_1^* + \Delta^* \quad (31)$$

$$\Delta^* = \min \left\{ \frac{1}{r + \rho} \log \left(1 + \frac{\delta(r + \rho)}{\gamma_u + (1 - \delta)(r + \rho + \gamma_d)} \right), \bar{\Delta} \right\},$$

where $\bar{\Delta} \equiv \phi_1(\bar{m}) + \phi_2(\bar{m})$ and, if $\gamma = \rho$, one lets $(e^{-\gamma x} - e^{-\rho x})/(\rho - \gamma) \equiv x$ for all $x \in \mathbb{R}$.

If $\Delta^* = \bar{\Delta}$, the first breaking time is $t_1^* = 0$, meaning that a marketmaker starts accumulating inventory at the time of the “crash.”

The socially optimal allocation has three main features. First, it is optimal that a marketmaker provides some liquidity: From time t_1^* to time t_2^* , she builds up and unwinds a positive inventory position. Second, it is not necessarily optimal that a marketmaker provides liquidity at time zero, when the selling pressure is strongest. This suggests that, although a marketmaker should provide liquidity, she should not act as a “buyer of last resort.” Third, when the economy is close to its steady state, interpreted as a normal time, a marketmaker should act as a mere matchmaker, meaning that she should buy and sell instantly. Thus, the socially optimal allocation draws a sharp distinction between socially optimal marketmaking in a normal time of low selling pressure, versus a bad time of strong selling pressure.¹⁴

4 Market Equilibrium

This section studies marketmakers’ incentives to provide liquidity. I show that the socially optimal allocation can be implemented in a competitive equilibrium as long as marketmakers have access to sufficient capital.

4.1 Competitive Marketmakers

This subsection describes a competitive market structure that implements the socially optimal allocation.

¹⁴Weill [2006] shows that the above socially optimal allocation is unique and provides natural comparative statics. It also establishes that, as $\rho \rightarrow \infty$ and the trading frictions vanish, the socially optimal allocation converges to a Walrasian limit in which marketmakers do not hold any inventories.

It is assumed that a marketmaker has access to some bank account earning the constant interest rate $\bar{r} = r$. At each time t , she buys a flow $u_\ell(t) \in \mathbb{R}_+$ of assets, sells a flow $u_h(t) \in \mathbb{R}_+$, and consumes cash at the positive rate $c(t) \in \mathbb{R}_+$. She takes as given the asset price path $\{p(t), t \geq 0\}$. Hence, her bank account position $a(t)$ and her inventory position $I(t)$ evolve according to

$$\dot{a}(t) = ra(t) + p(t)(u_h(t) - u_\ell(t)) - c(t) \quad (32)$$

$$\dot{I}(t) = u_\ell(t) - u_h(t). \quad (33)$$

In addition, she faces the borrowing and short-selling constraints

$$a(t) \geq 0 \quad (34)$$

$$I(t) \geq 0. \quad (35)$$

Lastly, at time zero, a marketmaker holds no inventory ($I(0) = 0$) and maintains a strictly positive amount of capital, $a(0)$. This subsection restricts attention to some large $a(0)$, in the sense that the borrowing constraint (34) does not bind in equilibrium. (This statement is made precise by Theorem 2 and Proposition 3.) The marketmaker's objective is to maximize the present value

$$\int_0^{+\infty} e^{-rt} c(t) dt \quad (36)$$

of her consumption stream with respect to $\{a(t), I(t), u_\ell(t), u_h(t), c(t), t \geq 0\}$, subject to the constraints (32)-(35), and the constraint that $u_\ell(t)$ and $u_h(t)$ are piecewise continuous.

Let's turn to the investor's problem. An investor establishes contact with the marketplace at Poisson arrival times with intensity $\rho > 0$.¹⁵ Conditional on establishing contact at time t , he can buy or sell the asset at price $p(t)$. I solve the investor's problem using a "guess and verify" method. Specifically, I guess that, in equilibrium, an ℓo (hn) investor always finds it weakly optimal to sell (buy). If an ℓo (hn) investor is indifferent between selling and not selling, he might choose not to sell (buy). Lastly, I guess that investors of types ℓn and $h o$ never trade. The time- t continuation utility of an investor of type $\sigma \in \mathcal{T}$ who follows this policy is denoted $V_\sigma(t)$. Hence, a seller's reservation value is $\Delta V_\ell(t) \equiv V_{\ell o}(t) - V_{\ell n}(t)$, the net value of holding one asset rather than none, while following the candidate optimal trading strategy. Likewise, a buyer's reservation value is $\Delta V_h(t) \equiv V_{h o}(t) - V_{h n}(t)$. Appendix

¹⁵In the alternative market setting of Duffie, Gârleanu and Pedersen (2005), each investors bargain individually with a marketmaker who can trade assets instantly on a competitive inter-dealer market. This market setting is equivalent to the present one in the special case in which the bargaining strength of a marketmaker is equal to zero.

C provides ODEs for these continuation utilities and reservation values, as well as a precise definition of a competitive equilibrium. The main result of this subsection states that the optimal allocation can be implemented in some equilibrium:

Theorem 2 (Implementation). *There exists some $\bar{a}_0^* \in \mathbb{R}_+$ such that, for all $a(0) \geq \bar{a}_0^*$, there exists a competitive equilibrium whose allocation is the optimal allocation.*

The proof identifies the price and the reservation values with the Lagrange multipliers of the socially optimal allocation (see Table 3). For instance, the asset price $p(t)$ is equal to the multiplier $\lambda_I(t)$ for the ODE $\dot{I}(t) = u_\ell(t) - u_h(t)$, interpreted as the social value of increasing the inventory position of a marketmaker. Also, Table 3 and the complementary slackness condition (23) imply that, before the first breaking time t_1^* , $w_\ell(t) = p(t) - \Delta V_\ell(t) = 0$. In other words, because the socially optimal allocation prescribes that marketmakers do not accommodate the selling pressure, the social value $w_\ell(t)$ of increasing the rate of contact with sellers is zero. In equilibrium, this means that the price adjusts so that sellers are indifferent between selling and not selling.

Weill [2006] shows, without characterizing the equilibrium allocation, that the efficiency result of Theorem 2 generalizes to environments with aggregate uncertainty and non-linear utility flow for holding the assets as long as: i) agents have quasi-linear utility, meaning that they enjoy the consumption of some numéraire good with a constant marginal utility of 1, and ii) the probability distribution of an investor's contact times with the market does not depend on other agents' trading strategies.¹⁶

Equilibrium Price Path and Marketmakers' Incentive

Appendix B.2 derives closed-form solutions for the equilibrium price path $p(t)$ and the reservation values $\Delta V_\ell(t)$ and $\Delta V_h(t)$. The price path, shown in the lower panel of Figure 4, jumps down at time zero, then increases, and eventually stabilizes at its steady-state level.¹⁷ The price path reflects the three phases of the socially optimal allocation: before the first breaking time t_1^* , sellers are indifferent between selling or not selling, meaning that $p(t) = \Delta V_\ell(t) = \lambda_\ell(t)$. Moreover, equation (26) shows that the growth rate $\dot{p}(t)/p(t)$ of the price is strictly less than $r - (1 - \delta)/p(t)$. This is because a seller with marginal utility $1 - \delta$

¹⁶Assumption ii) would fail, for instance, in a model with congestions, when the intensity of contact with marketmakers is a decreasing function of the number of agents on the same side of the market. In that case, an optimal allocation would be implemented in a different competitive setting, in the spirit of Moen [1997] and Shimer [1995]'s competitive-search models.

¹⁷A simple way to construct the initial price jump is to start the economy in steady state at $t = 0$ and assume that agents anticipate a crash at some Poisson arrival time with intensity κ . One can show that the results of this paper would apply, provided that either κ is small enough or $t_1^* > 0$. For Figure 4, it is assumed that $\kappa = 0$.

Table 3: Identifying Prices with Multipliers.

Equilibrium Objects	Multipliers	Constraints
$p(t)$	$\lambda_I(t)$	$\dot{I}(t) = u_\ell(t) - u_h(t)$
$\Delta V_\ell(t)$	$\lambda_\ell(t)$	$\dot{\mu}_{\ell o}(t) = -u_\ell(t) \dots$
$p(t) - \Delta V_\ell(t)$	$w_\ell(t)$	$u_\ell(t) \leq \rho \mu_{\ell o}(t)$
$\Delta V_h(t)$	$\lambda_h(t)$	$\dot{\mu}_{hn}(t) = -u_h(t) \dots$
$\Delta V_h(t) - p(t)$	$w_h(t)$	$u_h(t) \leq \rho \mu_{hn}(t)$

In a given row, the equilibrium object in the first column is equal to the Lagrange multiplier in the second column. The third column describes the constraints associated with these multipliers. For instance, in the first row, the price $p(t)$ (first column) is equal to the multiplier $\lambda_I(t)$ (second column) of the ODE $\dot{I}(t) = u_\ell(t) - u_h(t)$ (third column).

does not need a large capital gain to be willing to hold the asset. By contrast, in between the two breaking times t_1^* and t_2^* , marketmakers accommodate all of the selling pressure. As a result, the “marginal investor” is a marketmaker and $p(t) > \Delta V_\ell(t)$, meaning that the liquidity provision of marketmakers raises the asset price above a seller’s reservation value. Moreover, equation (28) shows that the price grows at the higher rate $\dot{p}(t)/p(t) = r$, implying that the price recovers more quickly when marketmakers provide liquidity.

The capital gain during $[t_1^*, t_2^*]$ exactly compensates a marketmaker for the time value of cash spent on liquidity provision. In other words, a marketmaker is indifferent between i) investing cash in her bank account, and ii) buying assets after t_1^* and selling them before t_2^* . Before t_1^* and after t_2^* , however, the capital gain is strictly smaller than r , making it unprofitable for a marketmaker to buy the asset on her own account. Therefore, in equilibrium, a marketmaker’s intertemporal utility is equal to $a(0)$, the value of her time-zero capital. In other words, although a marketmaker buys low and sells high, competition drives the present value of her profit to zero.

The equilibrium of Theorem 2 implements a socially optimal allocation with a bid-ask spread of zero. Conversely, Weill [2006] shows that, if marketmakers could bargain individually with investors and charge a strictly positive bid-ask spread, then they would provide more liquidity than is socially optimal. However, this finding that efficiency requires a zero bid-ask spread crucially depends on the specification of the trading friction. For example, Weill [2006] shows that if marketmakers face a fixed upper limit on the total number of orders they can execute per unit of times then the equilibrium liquidity provision would continue

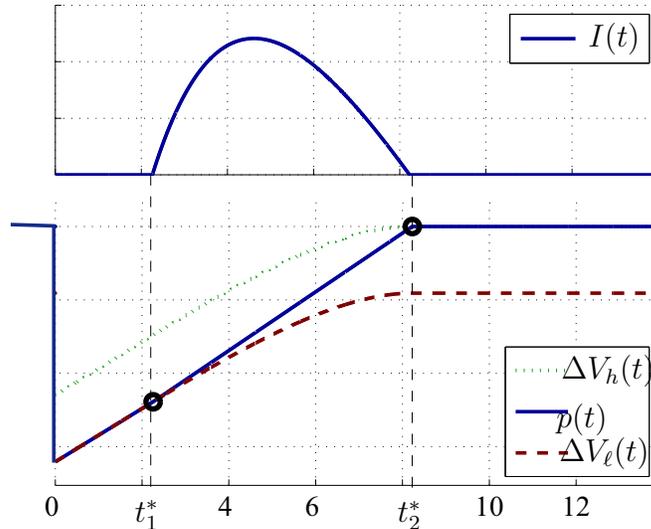


Figure 4: The Equilibrium Price Path.

to be socially optimal, but the equilibrium bid-ask spread would be strictly positive.

One might wonder whether the present results survive the introduction of a limit-order book. Indeed, although investors are not continuously in contact with the market, their limit orders could be continuously available for trading, and might substitute for marketmakers inventory accumulation. Weill [2006] addresses this issue in an extension of the model where the arrival of new information exposes limit orders to picking-off risk (see, among others, Copeland and Galai [1983]). It is shown that, if the picking-off risk is sufficiently large, then the limit-order book is empty, liquidity is only provided by marketmakers, and the asset allocation is the same as the one of Theorem 2.

4.2 Borrowing-Constrained Marketmakers

The implementation result of Theorem 2 relies on the assumption that the time-zero capital $a(0)$ is sufficiently large. This ensures that, in equilibrium, a marketmaker's borrowing constraint (34) never binds. There is, however, much anecdotal evidence suggesting that, during the October 1987 crash, specialists' and marketmakers' borrowing constraints were binding. Some market commentators have suggested that insufficient capital might have amplified the disruptions (see, among others, Brady [1988] and Bernanke [1990]). This subsection describes an amplification mechanism associated with insufficient capital and binding borrowing constraints. Specifically, the following proposition shows that if marketmakers are borrowing constrained during the crash and if their time-zero capital is small enough, then

they do not have enough purchasing power to absorb the selling pressure, and therefore fail to provide the optimal amount of liquidity.

Proposition 3 (Equilibrium with small capital). *There exists $\underline{a}_0^* \leq \bar{a}_0^*$ such that:*

- (i) *If marketmakers' aggregate capital is $a(0) \in [0, \underline{a}_0^*)$, there exists an equilibrium whose allocation is a buffer allocation with maximum inventory position $m \in [0, m^*)$.*
- (ii) *If marketmakers' aggregate capital is $a(0) \in [\underline{a}_0^*, \bar{a}_0^*)$, there exists an equilibrium whose allocation is a buffer allocation with maximum inventory position m^* .*

If $t_1^ > 0$, then $\underline{a}_0^* = \bar{a}_0^*$ and the interval $[\underline{a}_0^*, \bar{a}_0^*)$ is empty. Lastly, in all of the above, the equilibrium price path has a strictly positive jump at the time t_m such that $I(t_m) = m$ (that is, $t_m = \psi(m)$ for the function $\psi(\cdot)$ of Proposition 1).*

The equilibrium price and allocation are shown in Figures 5 and 6. The price jumps up at time t_m . It grows at a low rate $\dot{p}(t)/p(t) < r$ for $t \in [0, t_1)$, at a high rate $\dot{p}(t)/p(t) = r$ for $t \in (t_1, t_m) \cup (t_m, t_2)$, and at a zero rate after t_2 . Because of the price jump, a marketmaker can make positive profit: for instance, he can buy assets the last instant before the jump at a low price $p(t_m^-)$ and re-sell these assets the next instant after the jump at the strictly higher price $p(t_m^+)$. An optimal trading strategy maximizes the profit that a marketmaker extracts from the price jump, as follows: i) a marketmaker invests all of her capital at the risk-free rate during $t \in [0, t_1)$ in order to increase her buying power, ii) spends all her capital in order to buy assets before the jump, during $t \in (t_1, t_m)$, iii) re-sells all her assets after the jump, during (t_m, t_2) . A marketmaker does not hold any assets during $t \in [0, t_1)$ and $t \in (t_2, \infty)$ because the price grows at a rate strictly less than r . Because the price grows at rate r during (t_1, t_m) and (t_m, t_2) , a marketmaker is indifferent regarding the timing of her purchases and sales, as long as all assets are purchased during (t_1, t_m) , sold during (t_m, t_2) , and all capital is used up at time t_m .

The price jump at time t_m seems to suggest the following arbitrage: a utility-maximizing marketmaker would buy more assets shortly before t_m and sell them shortly after. This does not, in fact, truly represent an arbitrage, because a marketmaker runs out of capital precisely at the jump time t_m , so she cannot purchase more assets.

Perhaps the most surprising result is that, if $t_1^* = 0$, then there is a non-empty interval $[\underline{a}_0^*, \bar{a}_0^*)$ of time-zero capital such that the price has a positive jump and marketmakers accumulate the optimal amount m^* of inventories. If $t_1^* > 0$, then $\underline{a}_0^* = \bar{a}_0^*$ and the interval is empty. Intuitively, if the interval were not empty, then a small increase in time-zero capital combined with a positive price jump would give marketmakers incentive to provide more

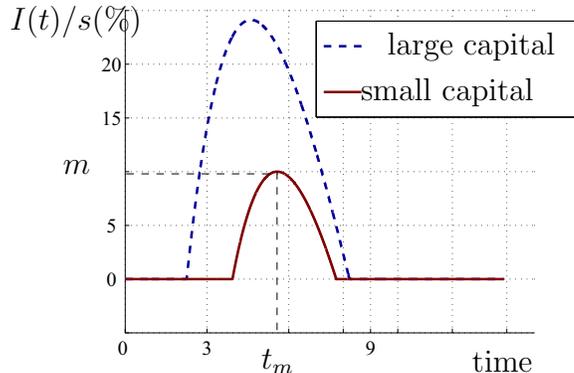


Figure 5: Equilibrium Allocations.

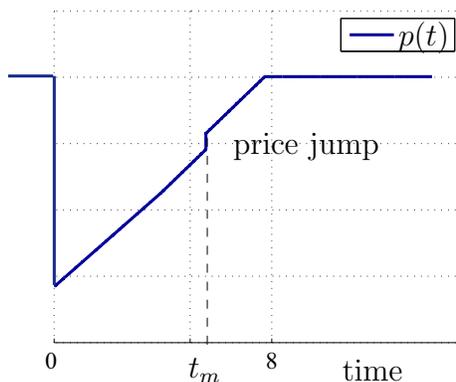


Figure 6: Equilibrium Price Path with a Small Marketmaking Capital.

liquidity. Hence, they would start buying assets at some time $t_1 < t_1^*$ and would end up accumulating more inventories than m^* , which would be a contradiction. If $t_1^* = 0$, this reasoning does not apply: indeed, an increase in time-zero capital cannot increase inventory accumulation because marketmakers cannot start accumulating inventories earlier than time zero.

Price Resilience

Proposition 3 illustrates the impact of marketmakers' liquidity provision on what Black [1971] called “price resiliency” – the speed with which an asset price recovers from a random shock. This speed can be measured by the time t_2 reaches its steady-state “fundamental value.” Note that this is also the time at which marketmakers are done unwinding their inventories. The proposition reveals that, when marketmakers provide more liquidity, the

market can appear less resilient. Indeed, increasing $a_0 \in [0, \underline{a}_0^*)$ means that marketmakers purchase more assets and, as a result, take longer to unwind their larger inventory position. This, in turn, increases the time t_2 at which the price recovers and reaches its steady state value.

Numerical calculations reported in Weill [2006] suggest that more liquidity can *lower* the price. To understand this somewhat counterintuitive finding, recall that the liquidity provision of marketmakers reduces the inter-temporal holding cost of the *average investor*¹⁸ by making marketmakers hold inventories for a longer time. Moreover, when they hold inventories, marketmakers become *marginal investors*. Therefore, the reduction in the inter-temporal holding cost of the average investor is achieved through an increase in the inter-temporal holding cost of the marginal investor. Because the asset liquidity discount capitalizes the inter-temporal holding of the marginal investor, more liquidity can lower the asset price.

5 Policy Implications

This section discusses some policy implications of this model of optimal liquidity provision.

Marketmaking Capital

The model suggests that, with perfect capital markets, competitive marketmakers would have enough incentive to raise sufficient capital. The intuition is that marketmakers will raise capital until their net profit is equal to zero, which precisely occurs when they provide optimal liquidity. For example, suppose that, at $t = 0$, wealthless marketmakers can borrow capital instantly on a competitive capital market. Then, for $t > 0$, the economic environment remains the one described in the present paper. If, at $t = 0$, a marketmaker borrows a quantity $a > 0$, then she has to repay $a \times e^{rT}$ at some time $T \geq t_2^*$. One can show that, with an optimal trading strategy (described at the end of Section 4.2), the net present value of her profit is

$$\left(\frac{p(t_m^+)}{p(t_m^-)} - 1 \right) a, \tag{37}$$

where the jump-size $(p(t_m^+)/p(t_m^-) - 1)$ depends implicitly on the time-zero aggregate mar-

¹⁸Indeed, equation (14) reveals that the planner's objective is to maximize the inter-temporal utility for the asset of the average investor or, equivalently to minimize his inter-temporal holding cost.

ketmaking capital.¹⁹ As long as the jump size is strictly positive, a marketmaker wants to borrow an infinite amount of capital. Therefore, in a capital-market equilibrium, a marketmaker’s net profit (37) must be zero, implying that $p(t_m^+)/p(t_m^-) = 1$ and $m = m^*$. This means that marketmakers borrow a sufficiently large amount of capital and provide the socially optimal amount of liquidity.

Lending capital to marketmakers, however, might be costly because of capital-market imperfections associated for example with moral hazard or adverse selection problems. In order to compensate for such lending costs, the net return $(p(t_m^+)/p(t_m^-) - 1)$ on marketmaking capital must be greater than zero. This would imply that, in an equilibrium, marketmakers do not raise sufficient capital. As a result, subsidizing loans to marketmakers may improve welfare.²⁰

During disruptions, some policy actions can be interpreted as bank-loan subsidization. For instance, during the October 1987 crash, the Federal Reserve lowered the fund rate, while encouraging commercial banks to lend generously to security dealers (Wigmore [1998]).

Price Continuity

It is often argued that marketmakers should provide liquidity in order to maintain price continuity and to smooth asset price movements.²¹ The present paper studies liquidity provision in terms of the Pareto criterion rather than in terms of some price smoothing objective. The results are evidence that Pareto optimality is consistent with a discrete price decline at the time of the crash. This suggests that requiring marketmakers to maintain price continuity at the time of the crash may result in a welfare loss.

A comparative static exercise suggests, however, that liquidity provision promotes some degree of price continuity. Namely, in an economy with no capital at time zero ($a(0) = 0$), no liquidity is provided and the price jumps up at time $t_s > 0$. In an economy with large time-zero capital, however, the price path is continuous at each time $t > 0$.

Marketmakers as Buyers of Last Resort

¹⁹The profit (37) is not discounted by e^{-rT} because a marketmaker can invest her capital at the risk free rate.

²⁰Weill [2006] provides an explicit model of marketmakers’ borrowing limits based on moral hazard. A different model of limited access to capital is due to Shleifer and Vishny [1997]. They show that capital constraints might be tighter when prices drop, due to a backward-looking, performance-based rule for allocating capital to arbitrage funds.

²¹For instance, Investor Relations, an advertising document for the specialist firm Fleet Meehan Specialist, argues that specialists “use their capital to fill temporary gaps in supply and demand. This can actually help to reduce short-term volatility by cushioning the intra-day price movements.”

A commonly held view is that marketmakers should not merely provide liquidity, they should also provide it promptly. In contrast with that view, the present model illustrates that prompt action is not necessarily consistent with efficiency. Namely, it is not always optimal that marketmakers start providing liquidity immediately at the time of the crash, when the selling pressure is strongest. For example, if the initial preference shock is very persistent, then marketmakers who buy asset immediately end up holding assets for a very long time. This cannot be efficient given that marketmakers are not the final holders of the asset. This suggests that requiring marketmakers to always buy assets immediately at the time of the crash can result in a welfare loss.

6 Conclusion

Although it focuses on liquidity provision during a crisis, the present paper may extend to less extreme situations. For instance, a similar model may help study the price impact of a large trader, such as a pension funds, selling its assets. Similarly, the model can help understand specialists' liquidity provision on individual stocks, in normal times (see the recent evidence of [Hendershott and Seasholes \[2007\]](#)).

One question that the paper leaves open is whether the welfare gains of marketmakers' liquidity provision can be quantitatively significant. Clearly, since marketmakers provide liquidity over relatively short time intervals, the welfare gains can only be significant if sellers find it very costly to hold assets during the crisis. Although the magnitude of price concessions sellers are willing to make during an actual crisis (e.g., the 23% price drop of the Dow Jones Industrial Average on October 19th, 1987) suggests that their holding costs are indeed very large, it can be quite difficult to rationalize these costs with standard models.

Taking a step back from the precise setup of the paper, one may also recall the commonly held view that liquidity provision creates large welfare gains, because it helps mitigate the risk of a meltdown of the financial system and may reduce the adverse impact of a disruption on the macro-economy. This view goes back, at least, to the famous Bagehot (1873) recommendation that the central bank provide liquidity on the money market during a crisis. Of course, the present paper does not explain the mechanism through which a failure to accommodate investors' liquidity needs may have an adverse impact on aggregate economic conditions. Instead, it takes these liquidity needs as given, and provides the optimal liquidity provision of marketmakers. Understanding precisely how financial market disruptions may propagate to the macro-economy remains an important open question for future research.

A Proof of Proposition 1

Hump Shape. Consider a buffer allocation (t_1, t_2) . One first shows that $I(t)$ is hump-shaped and that, given the first breaking time t_1 , the second breaking time t_2 is uniquely characterized. For $t \in [t_1, t_2]$, the inventory position $I(t)$ evolves according to $\dot{I}(t) = u_\ell(t) - u_h(t) = \rho(\mu_{\ell o}(t) - \mu_{h m}(t))$. With equation (3), this ODE can be written $\dot{I}(t) = -\rho I(t) + \rho(s - \mu_h(t))$. Together with the initial condition $I(t_1) = 0$, this implies that, for $t \in [t_1, t_2]$, $I(t) = H(t_1, t)$, where $H(t_1, t) \equiv \rho \int_{t_1}^t (s - \mu_h(z)) e^{\rho(z-t)} dz$. One has $\partial H/\partial t = -\rho H + \rho(s - \mu_h(t))$, implying that $\partial H/\partial t(t_1, t_1) = \rho(s - \mu_h(t_1)) \geq 0$ and $\partial H/\partial t(t_1, t_s) = -\rho^2 \int_{t_1}^{t_s} (s - \mu_h(z)) e^{\rho(z-t_s)} dz \leq 0$. Therefore, there exists $t_m \in [t_1, t_s]$ such that $\partial H/\partial t(t_1, t_m) = 0$. Moreover, $\partial H/\partial t = 0$ implies that $\partial^2 H/\partial t^2 = -\rho \dot{\mu}_h(t) - \rho \partial H/\partial t = -\rho \dot{\mu}_h(t) < 0$. This implies that t_m is unique, that $H(t_1, t)$ is strictly increasing for $t \in [t_1, t_m)$, and strictly decreasing for $t \in (t_m, \infty)$. Now, because $\mu_h(t) > s$ for t large enough, it follows that $H(t_1, t)$ is negative for t large enough. Therefore, given some $t_1 \in [0, t_s]$, there exists a unique $t_2 \in [t_s, +\infty)$ such that $H(t_1, t_2) = 0$.

Writing $\{t_1, t_m, t_2\}$ as a function of the maximum inventory position. The maximum inventory position of Proposition 1 is defined as $m \equiv I(t_m)$. One let $t_m \equiv \psi(m)$, for some function $\psi(\cdot)$ which can be written in closed form by substituting $\dot{I}(t_m) = 0$ and $I(t_m) = m$ in $\dot{I}(t) = -\rho I(t) + \rho(s - \mu_h(t))$:

$$\psi(m) = -\frac{1}{\gamma} \log \left(1 - \frac{s - m}{y} \right). \quad (38)$$

Now, solving the ODE $\dot{I}(t) = -\rho I(t) + \rho(s - \mu_h(t))$ with the initial condition $I(t_m) = m$, one finds

$$I(t) = m e^{-\rho(t-t_m)} + (s - y)(1 - e^{-\rho(t-t_m)}) + \rho y e^{-\gamma t_m} e^{-\rho(t-t_m)} \int_{t_m}^t e^{(\rho-\gamma)(u-t_m)} du. \quad (39)$$

Plugging (38) into (39), and making some algebraic manipulations, show that $I(t) = 0$ if and only if $t = t_m + z$, for some z solution of $G(m, z) = 1$, where

$$G(m, z) = \left(1 + \frac{m}{y - s} \right) e^{-\rho z} \left[1 + \rho \int_0^z e^{(\rho-\gamma)u} du \right]. \quad (40)$$

Let's define, for $x \in [0, +\infty)$, the two functions $g_i(m, x) = G(m, (-1)^i \sqrt{x})$, $i \in \{1, 2\}$. For $x > 0$, the partial derivatives of g_i with respect to x is

$$\begin{aligned} \frac{\partial g_i}{\partial x} &= \left(1 + \frac{m}{y - s} \right) \frac{(-1)^i \rho e^{-\rho(-1)^i \sqrt{x}}}{2\sqrt{x}} \\ &\quad \times \left[-1 - \rho \int_0^{(-1)^i \sqrt{x}} e^{(\rho-\gamma)u} du + e^{(\rho-\gamma)(-1)^i \sqrt{x}} \right]. \end{aligned} \quad (41)$$

One easily shows that this derivative can be extended by continuity at $x = 0$, with $\partial g_i/\partial x(m, 0) = -(1 + m/(y - s)) \rho \gamma/2$. The term in bracket in (41) is zero at $x = 0$, and is easily shown to be strictly increasing (decreasing) for $i = 1$ ($i = 2$). This shows that $g_i(m, \cdot)$ is strictly decreasing over $[0, +\infty)$. Moreover, for $m = 0$, $g_i(0, 0) = 1$. For $m > 0$, $g_i(m, 0) > 1$, $g_1(m, x) \rightarrow -\infty$ and $g_2(m, x) \rightarrow 0$ when $x \rightarrow +\infty$. This implies that, for any $m \geq 0$, there exists only one solution $x_i = \Phi_i(m)$ of $g_i(m, x) = 1$. An application of the Implicit Function Theorem (see

Taylor and Mann [1983], Chapter 12) shows that the function $\Phi_i(\cdot)$ is strictly increasing and continuously differentiable, and satisfies $\Phi_i(0) = 0$, $\Phi_i'(0) = 2/(\rho\gamma(y - s))$. Clearly, $G(m, z) = 0$ if and only if $z \in \{\phi_1(m), \phi_2(m)\}$, with $\phi_i(m) = (-1)^i \sqrt{\Phi_i(m)}$. Lastly, the restriction $t_1 \geq 0$ defines the domain $[0, \bar{m}]$ of the functions $\psi(\cdot)$, $\phi_1(\cdot)$, and $\phi_2(\cdot)$. Indeed $t_1 = \psi(m) - \phi_1(m)$, where the function $\psi(m) - \phi_1(m)$ is strictly decreasing, strictly positive at $m = 0$ and strictly negative for $m = s$. Hence, there exists a unique \bar{m} such that $\psi(\bar{m}) - \phi_1(\bar{m}) = 0$. By construction, the maximum inventory of a buffer allocation is less than \bar{m} .

B Socially Optimal Allocations

This appendix solves for socially optimal allocations. In order to prove the various results of Section 3 and 4, it is convenient to assume that, in addition to the trading technology and the short-selling constraint, the planner is also constrained by an inventory bound $I(t) \leq M$, for some $M \in [0, +\infty]$.

B.1 First-Order Sufficient Conditions

The current-value Lagrangian and the first-order conditions are the one of subsection 3.3, with an additional multiplier $\eta_M(t)$ for the inventory bound, that ends up being equal to zero. It is important to note that, because of the inventory bound, the multiplier $\lambda_I(t)$ can also jump up, with the restrictions $\lambda_I(t^+) - \lambda_I(t^-) \geq 0$ if $I(t) = M$. In what follows, it is convenient to eliminate $\lambda_h(t)$ and $\lambda_\ell(t)$ from the first-order conditions using (21) and (22). One obtains the reduced system

$$rw_\ell(t) = \delta - 1 - \gamma_u(w_h(t) + w_\ell(t)) - \rho w_\ell(t) + \eta_I(t) + \dot{w}_\ell(t) \quad (42)$$

$$rw_h(t) = 1 - \gamma_d(w_h(t) + w_\ell(t)) - \rho w_h(t) - \eta_I(t) + \dot{w}_h(t), \quad (43)$$

together with the ODE (28), the jump conditions (29),

$$\lambda_I(t^+) - \lambda_I(t^-) \geq 0 \text{ if } I(t) = M \quad (44)$$

$$\lambda_I(t^+) - \lambda_I(t^-) = w_\ell(t^+) - w_\ell(t^-) = -w_h(t^+) + w_h(t^-), \quad (45)$$

and the transversality conditions that $e^{-rt}\lambda_I(t)$, $e^{-rt}w_\ell(t)$, and $e^{-rt}w_h(t)$ go to zero as time goes to infinity. As before, the positivity restrictions and complementary slackness conditions are (23), (24), and (25). Note that, because the optimization problem is linear, there is no sign restrictions on the multiplier $\lambda_h(t)$ and $\lambda_\ell(t)$ (see section 3 in Part II of Kamien and Schwartz [1991]). As a result, the present reduced system of first-order condition is equivalent to the original system.

B.2 Multipliers for Buffer Allocations

Consider some feasible buffer allocation with breaking times (t_1, t_2) and a maximum inventory position $m \in [0, \min\{\bar{m}, M\}]$ reached at time t_m . This paragraph first constructs a collection $(w_h(t), w_\ell(t), \lambda_I(t), \eta_I(t))$ of multipliers solving the first-order sufficient conditions of Section B.1, but ignoring some of the positivity restrictions. These restrictions are imposed afterwards, when discussing the optimality of this allocation. First, summing equations (42) and (43), and using the transversality conditions shows that $w_\ell(t) + w_h(t) = \delta/(r + \rho + \gamma)$, for all $t \geq 0$. Then, one guesses that there are no jumps at t_1 and t_2 . With (45), this shows that $\lambda_I(t_i^+) - \lambda_I(t_i^-) = w_\ell(t_i^+) - w_\ell(t_i^-) = -w_h(t_i^+) + w_h(t_i^-) = 0$, for $i \in \{1, 2\}$. Now, one can solve for the multipliers, going backwards in time.

Time Interval $t \in [t_2, +\infty)$. Complementary slackness (24) implies that $w_h(t_2) = 0$. Plugging this into $w_h(t) + w_\ell(t) = \delta/(r + \rho + \gamma)$ gives that $w_\ell(t) = \delta/(r + \rho + \gamma)$. With (43), this also implies that $\eta_I(t) = -\gamma_d\delta/(r + \rho + \gamma)$. Lastly, together with the transversality condition, (28) implies that $r\lambda_I(t) = 1 - \delta\gamma_d/(r + \rho + \gamma)$.

Time Interval $t \in [t_m, t_2)$. First, because $I(t) > 0$, the complementary slackness condition (25) implies that $\eta_I(t) = 0$. Then, one solves the ODE (43) with the terminal condition $w_h(t_2) = 0$, and one finds that

$$w_h(t) = \frac{1}{r + \rho} \left(1 - \delta \frac{\gamma_d}{r + \rho + \gamma} \right) (1 - e^{(r+\rho)(t-t_2)}), \quad (46)$$

for $t \in [t_m, t_2)$. With $w_h(t) + w_\ell(t) = \delta/(r + \rho + \gamma)$, $w_\ell(t) = \delta/(r + \rho + \gamma) - w_h(t)$. Similarly, one can solve the ODE (28) with the terminal condition $r\lambda_I(t_2^-) = 1 - \gamma_d\delta/(r + \rho + \gamma)$, finding that $r\lambda_I(t) = (1 - \delta\gamma_d/(r + \rho + \gamma)) e^{r(t-t_2)}$.

Time Interval $t \in [t_1, t_m)$. In this time interval, $\eta_I(t) = 0$. One needs to consider two cases.

Case 1: $m < \bar{m}$. Complementary slackness at $t = t_1$ shows that $w_\ell(t_1) = 0$, implying that $w_h(t_1) = \delta/(r + \rho + \gamma)$. With this and (43), one finds

$$w_h(t) = \frac{1}{r + \rho} \left(1 - \delta \frac{\gamma_d}{r + \rho + \gamma} - \left(1 - \delta \frac{r + \rho + \gamma_d}{r + \rho + \gamma} \right) e^{(r+\rho)(t-t_1)} \right), \quad (47)$$

for $t \in [t_1, t_m)$. Given (46) and (47), the multiplier $w_h(t)$ is not necessarily continuous at time t_m . The size $w_h(t_m^-) - w_h(t_m^+)$ of the jump can be written as some function $b(\cdot)$ of the maximum inventory level m , where

$$b(m) \equiv \frac{1}{r + \rho} \left[\left(1 - \delta \frac{\gamma_d}{r + \rho + \gamma} \right) e^{-(r+\rho)\phi_2(m)} - \left(1 - \delta \frac{r + \rho + \gamma_d}{r + \rho + \gamma} \right) e^{(r+\rho)\phi_1(m)} \right]. \quad (48)$$

Equation (45) implies that $\lambda_I(t_m^+) - \lambda_I(t_m^-) = b(m)$. This and the ODE (28) show that $r\lambda_I(t) = (r\lambda_I(t_m^+) - rb(m)) e^{r(t-t_m)}$.

Case 2: $m = \bar{m}$. Then, by construction of \bar{m} , $t_1 = 0$. If $b(\bar{m}) < 0$, the multipliers are constructed as in Case 1. If, on the other hand, $b(\bar{m}) \geq 0$, then one constructs a set of multipliers by solving the ODEs (43) and (28) with terminal conditions $w_h(t_m^-) = w_h(t_m^+) + (1 - \alpha)b(\bar{m})$ and $\lambda_I(t_m^-) = \lambda_I(t_m^+) - (1 - \alpha)b(\bar{m})$, for all $\alpha \in [0, 1]$. That $b(\bar{m}) \geq 0$ implies that, for all $\alpha \in [0, 1]$, $w_h(t_1) = w_h(0) \in [0, \delta/(r + \rho + \gamma)]$. By construction, the $\alpha = 1$ multipliers do not jump at $t = t_m$.

Time Interval $t \in [0, t_1]$, $m < \bar{m}$ Complementary slackness shows that $w_\ell(t) = 0$, implying that $w_h(t) = \delta/(r + \rho + \gamma)$. With equation (42), this also implies that $\eta_I(t) = 1 - \delta(r + \rho + \gamma_d)/(r + \rho + \gamma) \geq 0$, and $r\lambda_I(t) = \eta_I(t) + e^{r(t-t_1)}(r\lambda_I(t_1) - \eta_I(t))$.

B.3 Proof of Proposition 2

Let us consider the multipliers $w_h^0(t), w_\ell^0(t), \eta_I^0(t), \lambda_I^0(t)$ associated with the no-inventory allocation, as constructed in Section B.2. Recall that $\lambda_I^0(t_s^+) - \lambda_I^0(t_s^-) = b(0) > 0$. It can be shown that, for

any buffer allocation (μ^m, I^m, u^m) ,

$$\begin{aligned} W(m) - W(0) &= - \int_0^{+\infty} e^{-rt} w_h^0(t) (\rho \mu_{hn}^m(t) - u_h^m(t)) dt - \int_0^{+\infty} e^{-rt} w_\ell^0(t) (\rho \mu_{\ell o}^m(t) - u_\ell^m(t)) dt \\ &\quad - \int_0^{+\infty} e^{-rt} \eta_I^0(t) I^m(t) dt + (\lambda_I^0(t_s^+) - \lambda_I^0(t_s^-)) I^m(t_s) e^{-rt_s}. \end{aligned} \quad (49)$$

Formula (49) follows from the standard comparison argument of Optimal Control (see, for example, Section 3 of Part II in [Kamien and Schwartz \[1991\]](#)), in the special case of a linear objective and linear constraints (see Corollary 2 in [Weill \[2006\]](#) for a derivation). The first two terms in equation (49) are zero because, for all $t \leq t_s$, $u_h^m(t) = \rho \mu_{hn}^m(t)$ and $w_\ell^0(t) = 0$, and, for all $t \geq t_s$, $u_\ell^m(t) = \rho \mu_{\ell o}^m(t)$ and $w_h^0(t) = 0$. The third term can be bounded by

$$0 \leq \int_{\psi(m) - \phi_1(m)}^{\psi(m) + \phi_2(m)} e^{-rt} \eta_I^0(t) I^m(t) dt \leq \left(1 - \gamma_d \frac{\delta}{r + \rho + \gamma} \right) m (\phi_2(m) + \phi_1(m)),$$

because $\eta_I^0(t) \leq 1 - \gamma_d \delta / (r + \rho + \gamma)$. Since $\lim_{m \rightarrow 0^+} \phi_i(m) = 0$, for $i \in \{1, 2\}$, this implies that, as $m \rightarrow 0^+$, $1/m \int_{\psi(m) - \phi_1(m)}^{\psi(m) + \phi_2(m)} e^{-rt} \eta_I(t) I^m(t) dt$ goes to zero. In order to study the last term of (49) note that equation (39) shows that $I^m(t_s) = 0$ at $m = 0$. Using the facts that $t_s = t_m + \psi(0) - \psi(m)$, and that $ye^{-\gamma t_m} = y - s + m$, differentiating (39) with respect to m shows that the derivative of $I^m(t_s)$ at $m = 0$ is equal to one. Therefore, $I^m(t_s)/m$ goes to 1, as m goes to zero, establishing Proposition 2.

B.4 Proof of Theorem 1

This paragraph verifies that some buffer allocation is constrained-optimal with inventory bound M . First, if some buffer allocation is constrained-optimal, it must satisfy the jump condition (44), meaning that $b(m) \geq 0$ and $b(m)(M - m) = 0$. In particular, if there is no inventory constraint, then the jump must be zero. One defines the maximum m such that the jump $b(m)$ is positive, $m^* = \sup\{m \in [0, \bar{m}] : b(m) \geq 0\}$. Furthermore, since $b(\cdot)$ is decreasing, $b(m) \geq 0$ for all $m \leq m^*$.

Proposition 4.. *For all $M \in [0, \infty]$, the buffer allocation with maximum inventory position $m = \min\{m^*, M\}$ is socially optimal with inventory bound M .*

In order to prove this Proposition, let's consider this allocation and its associated multipliers, constructed as in the previous subsection. Two optimality conditions remain to be verified: the jump conditions (44) and the positivity restrictions of (23) and (24). Because $m \leq m^*$, the jump condition (44) is satisfied. Also, because $w_h(t)$ is a decreasing function of time, $w_h(0) \in [0, \delta / (r + \rho + \gamma)]$ and $w_h(t_2) = 0$, it follows that, at each time, $w_h(t) \in [0, \delta / (r + \rho + \gamma)]$, and therefore that $w_\ell(t) \geq 0$. The inventory-accumulation period Δ^* and the breaking times (t_1^*, t_2^*) of Theorem 1 are found as follows. First, $\Delta^* = \phi_1(m^*) + \phi_2(m^*)$. Then, simple algebraic manipulations show that $b(m) \geq 0$ if and only if

$$e^{(r+\rho)(\phi_1(m)+\phi_2(m))} \leq 1 + \frac{\delta(r+\rho)}{\gamma_u + (1-\delta)(r+\rho+\gamma_d)}. \quad (50)$$

If $m^* < \bar{m}$, then (50) holds with equality at m^* , and if $m^* = \bar{m}$, it holds with inequality. This is equivalent to the formula of Theorem 1. Then, given Δ^* , the first breaking time t_1^* is a solution of $H(t_1^*, t_1^* + \Delta^*) = 0$, where $H(t_1, t)$ was defined in Section A. Direct integration shows that

$$H(t_1, t) = (s - y) \left(1 - e^{-\rho(t-t_1)} \right) + \rho y e^{-\gamma t_1} \frac{e^{-\gamma(t-t_1)} - e^{-\rho(t-t_1)}}{\rho - \gamma}, \quad (51)$$

where we let $(e^{-\gamma x} - e^{-\rho x})/(\rho - \gamma) = x$ if $\rho = \gamma$. Simple algebraic manipulation of (51) give the analytical solution of Theorem 1.

C Proofs of Theorem 2 and Proposition 3

Definition. We first provide a precise definition of a competitive equilibrium. First, investors' continuation utilities solve the ODE

$$rV_{\ell n}(t) = \gamma_u(V_{hn}(t) - V_{\ell n}(t)) + \dot{V}_{\ell n}(t) \quad (52)$$

$$rV_{\ell o}(t) = 1 - \delta + \gamma_u(V_{ho}(t) - V_{\ell o}(t)) + \rho(V_{\ell n}(t) - V_{\ell o}(t) + p(t)) + \dot{V}_{\ell o}(t) \quad (53)$$

$$rV_{hn}(t) = \gamma_d(V_{\ell n}(t) - V_{hn}(t)) + \rho(V_{ho}(t) - V_{hn}(t) - p(t)) + \dot{V}_{hn}(t) \quad (54)$$

$$rV_{ho}(t) = 1 + \gamma_d(V_{\ell o}(t) - V_{ho}(t)) + \dot{V}_{ho}(t), \quad (55)$$

where $\dot{V}_\sigma(t) \equiv dV_\sigma(t)/dt$. Hence, the reservation value $\Delta V_\ell(t)$ of a seller and of a buyer solve

$$r\Delta V_\ell(t) = 1 - \delta + \gamma_u(\Delta V_h(t) - \Delta V_\ell(t)) + \rho(p(t) - \Delta V_\ell(t)) + \Delta \dot{V}_\ell(t) \quad (56)$$

$$r\Delta V_h(t) = 1 + \gamma_d(\Delta V_\ell(t) - \Delta V_h(t)) - \rho(\Delta V_h(t) - p(t)) + \Delta \dot{V}_h(t). \quad (57)$$

Lastly, in order to complete the standard optimality verification argument, I impose the transversality conditions that both $\Delta V_h(t)e^{-rt}$ and $\Delta V_\ell(t)$ go to zero as time goes to infinity. Conversely, given the reservation values, one finds the continuation utilities by solving the ODE (52) and (54) for $V_{\ell n}$ and V_{hn} , and by letting $V_{j_o} = V_{j_n} + \Delta V_j$ for $j \in \{h, \ell\}$. Indeed, subtracting (52) from (54), integrating, and assuming transversality, one finds that $V_{hn}(t) - V_{\ell n}(t) = \int_t^{+\infty} e^{-(r+\gamma)(z-t)} \rho(\Delta V_h(z) - p(z)) dz$. And, replacing this expression in (52), that $V_{\ell n}(t) = \int_t^{+\infty} e^{-r(z-t)} \gamma_u(V_{hn}(z) - V_{\ell n}(z)) dz$. A *competitive equilibrium* is made up of a feasible allocation $(\mu(t), I(t), u(t))$, a price $p(t)$, a collection $(\Delta V_\ell(t), \Delta V_h(t))$ of reservation values, a consumption stream $c(t)$, and a bank account position $a(t)$ such that: i) given the price $p(t)$, $(I(t), u(t), c(t), a(t))$ solves the marketmaker's problem, and ii) given the price $p(t)$, the reservation values $(\Delta V_\ell(t), \Delta V_h(t))$ solve equations (56)-(57), satisfy the transversality conditions, and satisfy at each time,

$$p(t) - \Delta V_\ell(t) \geq 0 \quad (58)$$

$$\Delta V_h(t) - p(t) \geq 0 \quad (59)$$

$$(p(t) - \Delta V_\ell(t))(\rho\mu_{\ell o}(t) - u_\ell(t)) = 0 \quad (60)$$

$$(\Delta V_h(t) - p(t))(\rho\mu_{hn}(t) - u_h(t)) = 0. \quad (61)$$

Equations (58) through (61) verify the optimality of investors' policies. For instance, equation (58) means that the net utility of selling is positive, which verifies that a seller ℓo finds it weakly optimal to sell.²² Equation (60), on the other hand, verifies that a seller's trading decision is optimal. Namely, if the net utility $p(t) - \Delta V_\ell(t)$ of selling is strictly positive, then $u_\ell(t) = \rho\mu_{\ell o}(t)$, meaning that all ℓo investors in contact with marketmakers choose to sell. If, on the other hand, the net utility of selling is zero, then ℓo investors are indifferent between selling and not selling. As a result, $u_\ell(t) \leq \rho\mu_{\ell o}(t)$, meaning that some ℓo investors might choose not to sell.

Equilibrium: Solution method. The idea is to identify equilibrium objects with the multipliers of Appendix B.4, as in Table 3. For each $m \in [0, m^*]$, let $\Lambda(m)$ be the (set of) multipliers associated

²²Because of linear utility, it also shows that selling one share is always weakly preferred to selling a smaller quantity $q \in [0, 1]$.

with the buffer allocation m . If $m \in [0, m^*)$, or if $m = m^*$ and $m^* < \bar{m}$, then $\Lambda(m)$ is a singleton. If $m = m^*$ and $m^* = \bar{m}$, then $\Lambda(m^*)$ is a set (see Case 2 of Appendix B.4). Let $\bar{\lambda}$ (respectively $\underline{\lambda}$) be the element of $\Lambda(m^*)$ with largest (smallest) $\lambda_I(t_{m^*}^-)$. By construction $\bar{\lambda}$ has no jump at $t = t_m$. Lastly, one lets $\underline{a}_0^* \equiv \underline{\lambda}_I(t_{m^*}^-)e^{-rt_{m^*}}m^*$ and $\bar{a}_0^* \equiv \bar{\lambda}_I(t_{m^*}^-)e^{-rt_{m^*}}m^*$. By construction, $\underline{a}_0^* = \lim_{m \rightarrow m^*} \lambda_I(t_m)e^{-rt_m}m$. Moreover, $\underline{a}_0^* \leq \bar{a}_0^*$, with an equality if $m^* < \bar{m}$. Now, one can construct a competitive equilibrium using the following “backsolving” method. First, one picks some buffer allocation $m \in [0, m^*]$ and multipliers $\lambda \in \Lambda(m)$. Then, given λ , one guesses that price and values are given as in Table 3. If $m < m^*$, or if $m = m^*$ and $\lambda \neq \bar{\lambda}$, one takes time-zero capital to be $a(0) = \lambda_I(t_m)e^{-rt_m}m$. If $m = m^*$ and $\lambda = \bar{\lambda}$, one can take any $a(0) \in [\bar{a}_0^*, \infty)$. In the next paragraph, I verify that, given this time-zero capital, the buffer allocation, the price, and the values are the basis of a competitive equilibrium.

Conversely, let’s consider any $a(0) \in [0, \underline{a}_0^*)$. Given that $m \mapsto \lambda_I(t_m^-)e^{-rt_m}m$ is continuous and is zero at $m = 0$, the construction of the previous paragraph implies that there exists a competitive equilibrium implementing a buffer allocation with some inventory bound $m \in [0, m^*)$. For any $a(0) \in [\underline{a}_0^*, \infty)$, the previous paragraph implies that there exists a competitive equilibrium implementing the buffer allocation with maximum inventory m^* . In particular, if $a(0) \geq \bar{a}_0^*$, the multipliers do not jump at time t_{m^*} . This establishes Proposition 3 and Theorem 2.

Equilibrium: Verification. The current value Lagrangian for the representative marketmaker’s problem is

$$\begin{aligned} \mathcal{L}(t) = & c(t) + \hat{\lambda}_I(t)(u_\ell(t) - u_h(t)) + \hat{\lambda}_a(t)(ra(t) + p(t)(u_h(t) - u_\ell(t)) - c(t)) \\ & + \hat{\eta}_I(t)I(t) + \hat{\eta}_a(t)a(t) + \hat{w}_c(t)c(t). \end{aligned}$$

The first-order sufficient conditions are

$$1 + \hat{w}_c(t) = \hat{\lambda}_a(t) \tag{62}$$

$$\hat{\lambda}_I(t) = \hat{\lambda}_a(t)p(t) \tag{63}$$

$$r\hat{\lambda}_I(t) = \hat{\eta}_I(t) + \dot{\hat{\lambda}}_I(t) \tag{64}$$

$$\dot{\hat{\lambda}}_a(t) = -\hat{\eta}_a(t) \tag{65}$$

$$\hat{w}_c(t) \geq 0 \quad \text{and} \quad \hat{w}_c(t)c(t) = 0 \tag{66}$$

$$\hat{\eta}_I(t) \geq 0 \quad \text{and} \quad \hat{\eta}_I(t)I(t) = 0 \tag{67}$$

$$\hat{\eta}_a(t) \geq 0 \quad \text{and} \quad \hat{\eta}_a(t)a(t) = 0 \tag{68}$$

$$\hat{\lambda}_a(t^+) - \hat{\lambda}_a(t^-) \leq 0 \quad \text{if} \quad a(t) = 0, \tag{69}$$

together with the transversality conditions that $\lambda_x(t)x(t)e^{-rt}$ go to zero as t goes to infinity, for for $x \in \{I, a\}$. The Bellman equations and optimality conditions for the investors are (56)-(61). Direct comparison shows that a solution of the system (62)-(68), (56)-(61) of equilibrium equations (with transversality) is $p(t) = \lambda_I(t)$, $\hat{\lambda}_a(t) = 1 + (\lambda_I(t_m^+)/\lambda_I(t_m^-) - 1)$ for $t < t_m$, $\hat{\lambda}_a(t) = 1$, for $t \geq t_m$, $\hat{\eta}_a(t) = 0$, $\hat{\lambda}_I(t) = \hat{\lambda}_a(t)\lambda_I(t)$, $\hat{\eta}_I(t) = \hat{\lambda}_a(t)\eta_I(t)$, $\hat{w}_c(t) = \hat{\lambda}_a(t) - 1$, $\Delta V_\ell(t) = \lambda_\ell(t)$, and $\Delta V_h(t) = \lambda_h(t)$, together with the corresponding inventory-constrained allocation, and some consumption process $c(t)$ such that $c(t) = 0$ for $t \leq t_2$ and $c(t) = ra(t_2)$, for $t > t_2$. To conclude the optimality verification argument for a marketmaker, one needs to check the jump condition (69), and that $a(t) \geq 0$ for all $t \geq 0$. To that end, one notes that, for $t \in [t_1, t_2]$, $d/dt(a(t)e^{-rt}) = -p(t)\dot{I}(t)$ and $\dot{I}(t_m) = 0$. This implies that that $a(t)e^{-rt}$ is continuously differentiable and achieves its minimum at $t = t_m$. By construction $a(t_m) = 0$.

References

- Franklin Allen and Douglas Gale. From cash-in-the-market pricing to financial fragility, working paper. 2004. [35](#)
- Ben S. Bernanke. Clearing and settlement during the crash. *Review of Financial Studies*, 3:133–151, 1990. [21](#)
- Antonio E. Bernardo and Ivo Welch. Liquidity and financial market runs. *Quarterly Journal of Economics*, 2004. [5](#)
- Bruno Biais, Larry B. Glosten, and Chester Spatt. Market microstructure: A survey of micro-foundations, empirical results, and policy implications. *Journal of Financial Markets*, 8:217–264, 2005. [8](#)
- Fischer Black. Towards a fully automated exchange, part i. *Financial Analyst Journal*, 27:29–34, 1971. [23](#), [33](#)
- Nicholas F. Brady. *The Presidential Task Force on Market Mechanisms*. US Government Printing Office, Washington DC, 1988. [2](#), [3](#), [9](#), [21](#)
- Pierre Brémaud. *Point Processes and Queues*. Springer-Verlag, New-York, 1981. [23](#), [49](#), [56](#), [58](#)
- Thomas E. Copeland and Dan Galai. Information effects on the bid-ask spread. *Journal of Finance*, 38:1457–1469, 1983. [21](#), [37](#)
- Harold Demsetz. The cost of transacting. *The Quarterly Journal of Economics*, 82:33–53, 1968. [3](#)
- Douglas W. Diamond and Philip H. Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91:401–417, 1983. [5](#)
- Darrell Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, 2001. [58](#)
- Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Over-the-counter markets. *Econometrica*, 73:1815–1847, 2005. [5](#), [31](#)
- Michael J. Fleming and Kenneth D. Garbade. When the back-office moved to the front burner: Settlement fails in the treasury market after 9/11. *Federal Reserve Bank of New-York Economic Policy Review*, November:35–57, 2002. [2](#)
- Douglas F. Foster and S. Viswanathan. Strategic trading when agents forecast the forecasts of others. *Journal of Finance*, 51:1437–1478, 1996. [33](#)
- Michael A. Goldstein and Kenneth A. Kavajecz. Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*, 7, 2004. [37](#)
- Maurice Greenberg. Shake up the nyse specialist system or drop it. *Financial Times*, October 10, 2003. [3](#)
- Bruce Greenwald and Jeremy Stein. The task force report: The reasoning behind the recommendations. *Journal of Economic Perspective*, 2:3–23, 1988. [3](#)

- Bruce Greenwald and Jeremy Stein. Transactional risk, market crashes, and the role of circuit breakers. *Journal of Business*, 64:443–462, 1991. 5
- Sanford J. Grossman and Merton H. Miller. Liquidity and market structure. *Journal of Finance*, 43:617–637, 1988. 5, 6, 33
- George Hall and John Rust. Middlemen versus market makers: A theory of competitive exchange. *Journal of Political Economy*, 111:353–403, 2003. 5
- Terry Hendershott and Mark S. Seasholes. Marketmaker inventories and stock prices. *Forthcoming, American Economic Review*, 2007. 26
- Craig W. Holden and Avanidhar Subrahmanyam. Long-lived private information and imperfect competition. *Journal of Finance*, 47:247–270, 1992. 33
- Arthur J. Hosios. On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies*, 57(1):279–298, 1990. 29
- Morton I. Kamien and Nancy L. Schwartz. *Dynamic Optimization: the Calculus of Variation and Optimal Control Theory in Economics and Management*. North-Holland, New-York, 1991. 15, 16, 28, 30, 3, 49
- Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1336, 1985. 33
- Ricardo Lagos. Asset prices and liquidity in an exchange economy. Working Paper, New-York University, 2006. 5
- Ricardo Lagos, Guillaume Rocheteau, and Pierre-Olivier Weill. Crashes and recoveries in illiquid markets. Working Paper, New York University, Federal Reserve Bank of Cleveland, and University of California, Los Angeles, 2006. 31
- Yiting Li. Middlemen and private information. *Journal of Monetary Economics*, 42:131–159, 1998. 5
- Francis Longstaff. The flight to liquidity premium in u.s. treasury bond prices. *Journal of Business*, 77:511–526, 2004. 2
- Arnaud Marès. Market liquidity and the role of public policy. *Bureau of International Settlement Papers*, 12:385–401, 2001. 3
- Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, Oxford, 1995. 18, 21
- Adrian Masters. Efficiency of intermediation in search equilibrium. Working paper, 2004. 5
- James J. McAndrews and Simon M. Potter. Liquidity effects of the events of september 11, 2001. *FRBNY Economic Policy Review*, November:59–79, 2002. 2
- Espen R. Moen. Competitive search equilibrium. *Journal of Political Economy*, 105:385–411, 1997. 19, 18, 23, 25
- Maureen O’Hara. *Market Microstructure Theory*. Blackwell, Cambridge, 1995. 4

- Robert T. Parry. The october '87 crash ten years later. *Federal Reserve Bank of San Francisco Economic Letter*, 97-32, 1997. 4, 45
- Philip Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, New York, 1990. 6
- Ariel Rubinstein and Asher Wolinsky. Middlemen. *Quarterly Journal of Economics*, pages 581–594, 1987. 5
- William G. Schwert. Stock market volatility: Ten years after the crash. *Brookings-Wharton Papers on Financial Services*, 1:65–114, 1998. 37
- Atle Seierstad and Knut Sydsæter. Sufficient conditions in optimal control theory. *International Economic Review*, 18:367–391, 1977. 15, 49
- Andrei Shevchenko. Middlemen. *International Economic Review*, 45:1–25, 2004. 5
- Robert Shimer. Contract in a frictional labor market, working paper. MIT, 1995. 19, 18, 23, 25
- Andrei Shleifer and Robert W. Vishny. The limits of arbitrage. *Journal of Finance*, 52:35, 1997. 25
- Daniel F. Spulber. Marketmaking by price-setting firms. *Review of Economic Studies*, 63:559–580, 1996. 5
- Yeneng Sun. On the sample measurability problem in modeling individual risks, working paper. National University of Singapore, 2000. 7
- Angus E. Taylor and Robert W. Mann. *Advanced Calculus*. Wiley, John and Sons, New-York, 1983. 28
- Dimitri Vayanos. Strategic trading in a dynamic noisy market. *Journal of Finance*, 56, 2001. 33
- Dimitri Vayanos and Minh Chau. Strong-form efficiency with monopolistic insiders. *Review of Financial Studies*, 2006. Forthcoming. 33
- Dimitri Vayanos and Tan Wang. Search and endogenous concentration of liquidity in asset markets. *Journal of Economic Theory*, 2006. Forthcoming. 5
- Dimitri Vayanos and Pierre-Olivier Weill. A search-based model of the on-the-run phenomenon. Working Paper, London School of Economics and UCLA, 2006. 5
- Pierre-Olivier Weill. Liquidity premia in dynamic bargaining markets. Working Paper, Department of Economics, UCLA, 2004. 5
- Pierre-Olivier Weill. Addenda to “leaning against the wind”. Working Paper, Department of Economics, UCLA, 2006. 17, 19, 20, 21, 24, 25, 30
- Barrie A. Wigmore. Revisiting the october 1987 crash. *Financial Analysts Journal*, 54:36–48, 1998. 4, 25, 45