NBER WORKING PAPER SERIES

FINANCIAL MARKET RUNS

Antonio E. Bernardo
Ivo Welch

Working Paper 9251
http://www.nber.org/papers/w9251

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2002

Financial Market Runs
Antonio E. Bernardo and Ivo Welch
NBER Working Paper No. 9251
October 2002
JEL No. G2, G1, G21, E44, N2

ABSTRACT

Our paper offers a minimalist model of a run on a financial market. The prime ingredient is that each risk-neutral investor fears having to liquidate *after* a run, but *befor*e prices can recover back to fundamental values. During the urn, only the risk-averse market-making sector is willing to absorb shares. To avoid having to possibly liquidate shares at the *marginal* post-run price–in which case the market-making sector will already hold a lot of share inventory and thus be more reluctant to absorb additional shares–all investors may prefer selling their shares into the market today at the *average* run price, thereby causing the run itself. Consequently, stock prices are low and risk is allocated inefficiently. Liquidity runs and crises are not caused by liquidity shocks per se, but by the *fear of future* liquidity shocks.

Antonio E. Bernardo
The Anderson School
UCLA
110 Westwood Plaza
Los Angeles, CA 90095

Ivo Welch
Yale School of Management
46 Hillhouse Avenue
New Haven, CT 06520
and NBER
ivo.welch@yale.edu

In contrast to the *financial institutions* literature (e.g., Diamond and Dybvig (1983)), runs on *financial markets* have not been a prime subject of inquiry. Our paper offers a minimalist model of a run on a financial market. The main ingredient of our model is that investors fear (but do not necessarily experience) future liquidity shocks. This creates two scenarios.

In the good scenario, a risk-neutral public holds most of the risky shares. Investors hit by a liquidity shock in the future will sell to the risk-averse market-making sector at a "low-inventory price," which will be close to the risk-neutral value of the asset. In the good scenario, the market-making sector provides the public with low-cost insurance against liquidity shocks.

In the bad scenario, every investor conjectures that other investors intend to sell today, thus causing a "run." By joining the pool of selling requests today, an individual investor can expect to receive the average price that is necessary to induce the market-making sector to absorb all tendered shares today. The investor's alternative is to not enter the pool and instead to hold onto the shares. In making this decision, this investor is better off *if* he can wait out the storm and realize the eventual expected asset value. However, if he were randomly hit by the possible liquidity shock, this investor would need to sell his shares *behind* the rest of the public. But, with the market-making sector already holding the shares of other tendering investors, this post-run price will be worse than the average in-run price today. If the average in-run price is greater than the expected payoff achieved by waiting, this investor will join the herd and also sell into the run. If other investors act alike, the conjecture that other investors sell today ends up being verified. In the bad scenario, the market-making sector holds too many shares and provides the public with high-cost insurance against liquidity shocks.

Our bad scenario relies either on random or batch execution. However, if execution is sequential, investors cannot expect to avoid a later place in line by joining the selling pool. Thus, the last investors (who now know they are last) are better off waiting rather than joining the herd and the bad scenario unravels. In reality, financial markets lack perfectly sequential execution in at least three circumstances. First, there is often no sequential execution after a market closure: for example, at the stock market opening or after a trading halt, markets are often conducted in a "batch" mode where all orders are crossed at the same price—and, indeed, fears of stock market runs seem higher around the NYSE opening period.[1] Second, even during normal trading, sequential execution may break down under the load of orders flowing in, and investors' order executions could become random. There is a lot of anecdotal evidence that sequential execution broke down in the 1987 stock market crash. Greenwald and Stein (1988, p15f) note that "investors cannot know with any precision at what prices their orders are executed...trades consummated only minutes apart were executed at wildly different prices, so that an investor submitting a market order had virtually no idea where it would be completed." A tendering investor, not knowing his place in the queue, would expect to receive some average price[2]—and the chain of perfect sequentiality may not just be broken on the exchange itself, but also in the communication of brokers with the exchanges and with their investors. This institutional breakdown could lead to an immediate transition from a situation in which liquidity shocks are not a major concern (as in a sequential market) to a situation in which they become paramount (as in a batch market). Third, in many over-the-counter financial markets, counterparties need

[1] Examining market-making inventories at the reopening of the stock markets on September 17, 2002 would allow for an interesting test of our model.

[2] Yet another interpretation would have a seller be unaware whether he received information about overvaluation/undervaluation before or after other investors, as in Abreu and Brunnermeier (2001).

to be found, and when multiple sellers are searching for counterparties, there is randomness as to who will find the potential buyers first.

It is important to point out that our model is not driven by the liquidity shocks themselves. *Instead, prices and market-making inventories are driven by the fear of future liquidity shocks.* Thus, the liquidity shocks might loom in the *future* and cause a run *today*. If underlying exogenous parameters change, high volatility and runs (low prices, high market-making inventory) can appear and disappear many times before the liquidity shocks themselves. An empiricist might not even necessarily recognize the relevance of actual liquidity constraints.

In a sense, the outcome of our model is perplexing. There are no transaction or search costs or asymmetric information. Investors are numerous, risk-neutral, and homogeneous. The market-making sector can be very deep with only slight risk aversion (small discounts to absorb liquidity). Unlike much of the feedback trading literature, liquidity shocks can loom in the distance and need not be correlated among investors. Unlike in the financial institutions run literature, in our financial markets setting, there are no sequential service constraints, no productive inefficiencies, and no need for investors to join in a run in order to get anything. Rather, our investors can attempt to "wait out the storm," and thereby perhaps do better. And, yet, our financial market can produce outcomes in which every investor wants to sell to avoid selling behind the average investor. Consequently, inefficient bearers of risk (the market-making sector) hold too much of the risky asset. Moreover, there is an accelerator effect whereby small changes in the likelihood of a liquidity shock can have big effects on the allocation of risk and the equilibrium price. The intuition that runs can be driven by investors fearing "to come in last" is solid, and resonates with many who witnessed the 1987 crash.

The main assumptions and insights of the model seem both realistic and robust. Indeed, the analytics of the model are simple, relying only on situations in which sequential execution breaks down, and on some split of participants into a (potentially only slightly) risk-averse market-making sector and an outside sector living in fear of potential *future* liquidity shocks. Not requiring much machinery, the model hints that run equilibria may not be esoteric but intrinsic to financial markets with capacity limits (just as they are intrinsic for financial institutions). This is not to argue that runs are frequent (indeed, they are very rare!), but that their occurrance is not logically far-fetched.

After working with exogenous liquidity shocks in Section II, we introduce margin constraints in Section III. These constraints endogenize the probability of future investor liquidity concerns, so that a price drop can quickly trigger investors' fear of a liquidity run, which in turn can trigger a further price decline, further margin constraints, etc. In this circumstance, the liquidity run becomes a short-term high-frequency phenomenon, and normal transaction channels may be quickly overwhelmed.

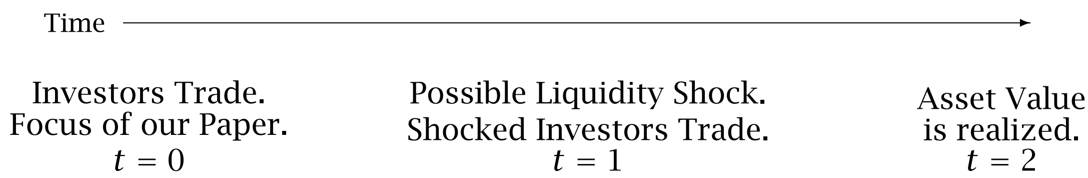
Our models work even if investor shocks are independent; however, if investor shocks are independent, the question of why unaffected individuals do not join the market-making sector becomes pertinent. Our paper contains a long discussion thereof on page 18, but we believe this is indeed how liquidity runs remain limited and how they come to an end: unaffected investors eventually join the market-making sector, earning a positive rate of return for doing so, which compensates them for their residual risk-aversion. With buy and sell orders flowing in more smoothly, sequential order is restored again. In one sense, the question as to why the market-making sector does not expand is similar to why banks in a Diamond and

Dybvig (1983) do not find themselves quickly additional backers to avoid inefficient liquidation.[3] Such backers could earn positive expected rates of returns. Unlike in the financial intermediation sector, we believe that this process can occur faster in a financial market sector—and it is this process that naturally limits the depth and duration of liquidity runs. Taken together—widespread and positively correlated endogenous liquidity fears and runs interacting with margin constraints—allow our model to capture at least some of the causal dynamics during a stock market crash.

Our paper now proceeds as follows: Section I lays out the model. The model's emphasis is on simplicity. Section II describes the equilibrium under CARA and CRRA market-making utility. Although we solve the model under perfectly correlated liquidity shocks, we then show that the intuition of our model survives even if shocks are uncorrelated. (This section also offers an equilibrium model for the market-making sectors' inventory in ordinary times.) Section III adds margin constraints to our model, which endogenizes the liquidation probability. Section IV discusses the economics of the equilibrium. Section V relates our work to earlier papers, particularly the bank-run literature (Diamond and Dybvig (1983)). And Section VI concludes.

[3] After all, banks have existing financial links. A similar question is why banks do not simply sell their loans for (more) cash to other banks. The answer is probably that time and imperfect information about asset quality play a role—but ultimately, this is how bank runs can come to an end—and with a net profit for some willing outsiders.

I The Basic Model Setup



We consider a model with three dates ($t = 0, 1, 2$) and two assets: a risk-free bond in infinitely elastic supply with a gross payoff of \$1 at date two and a risky asset (henceforth, "stock") with gross random payoff of $\tilde{Z}$ at date two. For simplicity, we normalize the date 0 and date 1 price of the bond to be \$1. The date 0 and date 1 price of the stock is determined endogenously.

Shares in the stock trade at date 0 and date 1. There are two types of traders in our market: atomistic individual investors and a competitive market-making sector.

Market-makers constitute an entire sector which encompasses not just the specialist, but all traders willing to absorb shares upon demand, i.e., regardless of the (fear of) liquidity shocks. Still, it is reasonable to attribute a finite risk absorption capacity to this sector and thus we assume the market-making sector is risk-averse in aggregate. For example, many institutions and traders do not seem willing to absorb shares during a financial markets crash, and instead prefer to wait it out.[4] We also assume that the market-making sector is competitive and is characterized

[4] Gammill and Marsh (1988) describe the broader market-making sector during the 1987 crash in great detail. Dennis and Strickland (2002) find that institutions are more likely than retail investors to sell into a dropping market. Indeed, portfolio insurers even precommit to such strategies. Amihud, Mendelson, and Wood (1990) document the liquidity decline during the 1987 crash, and describe that "Orders could not be executed, and information on market conditions and on order execution was delayed. Consequently, much of the burden of responding to the unexpected order flow fell on the exchange specialists, market makers, and other traders with immediate access to the trading floor." Thus, the market-making sector may be smaller than often assumed, and suffer a steeper and quicker price drop than sketched by our model.

by a "representative" market-maker with date 0 wealth W_0 and date 0 inventory of zero shares.

Individual investors are identical and endowed with shares which sum to the total supply of shares (normalized to one). Individual investors are assumed to be risk-neutral. Importantly, individual investors face a potential liquidity shock at date 1. We model liquidity shocks in various ways. In Section II, we assume that each individual investor may be forced to liquidate her shares at date 1 with an exogenous probability s. After showing that this leads to a non-zero market-making inventory at date 0, we parameterize the liquidity shocks and market-making utility function. The liquidity shocks are assumed to be perfectly correlated across investors in Subsection II.A and independent across investors in Subsections II.B and II.C. In Section III, we endogenize the date 1 liquidation probability s to depend on the date 0 stock price by introducing margin constraints. Each of these liquidity assumptions has been employed in related literature, and each offers its own trade off of realism and model cleanness.

To recap, there are two important differences between the market-making sector and individual investors. First, individual investors are assumed to be risk-neutral and the market-making sector is risk-averse (e.g., Diamond and Verrecchia (1991)). The former assumption is not crucial to the analysis but captures the fact that the investing public has considerably more risk absorption capacity than the market-making sector, and that, in a Pareto efficient outcome, shares should be held by the investing public (the most efficient bearers of risk). Second, only individual investors face a potential liquidity shock at date 1. This provides the motivation for trade between individual investors and the market-making sector at both dates 0 and 1. Moreover, this ensures that only the market-making sector (in equilibrium) is willing to buy shares in a run situation, at a price that depends on the market-making sector's risk-tolerance. An investor who learns that he is not subject to any future forced liquidation can join and thereby deepen the market-making sector.[5]

Individual investors are assumed to submit market orders. Limit orders could potentially deepen the market-making sector, and are thus not fully considered in our model. Greenwald and Stein (1988, footnote 16) also note that "limit orders do not represent an especially attractive alternative under the conditions of October 19th and 20th. An investor's threshold price should depend on his most current information, which includes the current market price. Under very volatile conditions, this can mean resubmitting limit orders on an almost continuous basis, which would have been extremely difficult to accomplish."

To close the model, we assume that the equilibrium price is determined by a zero-utility condition on the representative market-maker. Specifically, in our batch execution model, we assume that the price of the stock at each date is set so that the representative market-maker is indifferent between buying the entire batch and holding his inventory. Because the representative market-maker is risk averse, this price is typically decreasing in his inventory. Our zero-utility condition is analogous to Kyle's zero-expected-profits condition for a risk-neutral market-making sector and can be justified by the joint assumptions that the market-making sector is competitive and market-makers are free to enter or exit after each date.

There are two equivalent interpretations to the market microstructure which determines the equilibrium price—and both of them can be shown to lead to the same market-maker demand function and thus identical solutions for our model. One

[5] Naturally, this investor must have some risk-aversion or limited capital. Otherwise, the market-making sector itself becomes infinitely deep when the first investor appears. Section IV.A discusses what happens when the aggregate risk tolerance changes in a crisis.

interpretation is that all orders are batched and executed at an identical "average" price. For simplicity of exposition, we proceed using this specific assumption only. Sell orders from individual investors at each date are batched and then executed at an average price that yields zero utility for the representative market maker. Another interpretation is that orders are executed sequentially (with lower prices for subsequent trades) but in random order. In both cases the investor submits a market order and is unsure of the exact price at which her shares will be executed. In the first interpretation, we must impose a zero-utility condition in each period while in the second interpretation we must impose a zero-utility condition on each trade. Thus, our model applies equally to batch auction markets (e.g., at the NYSE stock market opening and after a trading halt); and to over-the-counter markets and to stock market crashes, when limited communication lines to the market-making sector can change the typical deterministic sequential execution into random execution.[6]

II Equilibrium With Exogenous Liquidity Shocks

In what follows, we assume that individual investors are endowed with the entire supply of shares at date 0 (consistent with an efficient allocation of risk). We analyze only the situations in which individual investors may wish to sell shares at date 0 due to the fear of a liquidity shock at date 1. These situations are the most

[6] An alternative setup is provided in Diamond and Verrecchia (1991). Their price setting mechanism differs from ours in two respects: market makers earn surplus on inframarginal trades and they set prices by solving a dynamic optimization problem. The latter implies that market makers forecast future buys and sells when setting today's price. Whether or not market-makers earn surplus is not important for the qualitative results of our model. Solving the market-makers' price function in a dynamic optimization problem is not tractable in our setup. Nonetheless, our qualitative results would still obtain, because the important feature of our model is that the price at $t = 1$ in the event of a liquidity shock is lower than the price at $t = 0$.

interesting because we intend to demonstrate that the fear of liquidity shocks can lead to substantial market-maker inventories and thus inefficient allocations of risk. Therefore, we ignore situations in which individual investors want to buy shares at date 0.

Consider an individual investor who conjectures that a total of α shares will be sold by individual investors to the market-making sector at date 0 and let $p_0(\alpha)$ denote the date-0 price set by the market makers when α sell orders arrive at date 0. If this investor also sells her shares at date 0, she will expect to receive the price $p_0(\alpha)$. However, if this investor chooses not to sell her shares at date 0 then either (i) she will be forced to liquidate her shares with probability s at date 1 or (ii) she will not be forced to liquidate her shares with probability $1 - s$ at date 1 and will optimally wait to receive the expected value of the stock, μ, at date 2. If liquidity shocks are perfectly correlated (as in Subsection A), the remaining proportion $(1-\alpha)$ of shares will be liquidated at date 1 if the liquidity shock occurs. If liquidity shocks are independent (as in Subsection B), the *Law of Large Numbers* ensures that the proportion $(1-\alpha)\cdot s$ of shares will be liquidated at date 1. Let $p_1(q_1(\alpha);\alpha)$ denote the date-1 price set by the market-makers when they hold α shares of inventory and $q_1(\alpha)$ new sell orders arrive at date 1. If this investor does not sell at date 0 she will expect to receive $s\cdot p_1(q_1(\alpha);\alpha) + (1-s)\cdot\mu$. Thus, it will be optimal for this investor to sell if and only if

$$p_0(\alpha) \geq s\cdot p_1(q_1(\alpha);\alpha) + (1-s)\cdot\mu. \tag{1}$$

Notice that if an investor is forced to liquidate at date 1, she receives a lower selling price than if she had sold at date 0. A risk-averse market-making sector implies that $p'(\cdot) < 0$, i.e., the market-making sector will require a lower price (greater risk premium) if it has to buy a greater number of shares. However, if she is not forced to liquidate at date 1, she receives the expected value of the stock which is greater than the selling price at date 0. The decision to sell at date 0 depends critically on the investor's beliefs about whether other investors will choose to sell at date 0.

We consider only symmetric Nash equilibria.

Definition 1 *Let $F(\alpha)$ denote the expected net benefit of selling shares at date 0 (compared to not selling) when the investor conjectures that α shares will be sold at date 0. If liquidity shocks will be perfectly correlated across investors,*

$$F(\alpha) = \overbrace{p_0(\alpha)}^{\text{if tender today}} - \overbrace{s \cdot p_1(q_1(\alpha); \alpha)}^{\text{if forced to liquidate tomorrow}} - \overbrace{(1-s) \cdot \mu}^{\substack{\text{if liquidation} \\ \text{not necessary}}} . \tag{2}$$

where $q_1(\alpha) = (1-\alpha)$ in the perfectly correlated shock case, and $q_1(\alpha) = s \cdot (1-\alpha)$ in the independent shock case. Then (i) waiting ($\alpha^\star = 0$) is a pure strategy Nash equilibrium iff $F(0) \le 0$; (ii) selling ($\alpha^\star = 1$) is a pure strategy Nash equilibrium iff $F(1) \ge 0$; and (iii) $\alpha^\star \in (0,1)$ is a mixed strategy Nash equilibrium iff $F(\alpha^\star) = 0$.

We can immediately demonstrate that ($\alpha^\star = 0$) is not a symmetric Nash equilibrium if the probability of a liquidity shock is positive.

Theorem 1 *Although market-makers are risk-averse and investors are risk-neutral and not yet liquidity-shocked, the market-making sector holds inventory at date 0 if there is a positive probability of a liquidity shock (s) at date 1.*

Proof: For $s > 0$, $F(0) > 0$, so $\alpha^\star = 0$ is not an equilibrium. □

The intuition is that if the market-making sector holds zero inventory, it would be willing to accept the first share at its risk-neutral valuation today. Thus, the first seller would avoid the liquidation risk tomorrow without any price penalty today.

Unfortunately, there is little more we can say without parameterizing returns and the representative market-making sector's utility function in order to determine the $p_t(\cdot)$ functions. Thus, we now consider two cases. In the first case, we assume that the market-making sector has constant absolute risk aversion (CARA) preferences and the stock payoff is distributed normal. We solve for the symmetric Nash equilibria with perfectly correlated liquidity shocks in Subsection A and independent liquidity shocks in Subsection B. The CARA plus normality assumptions allow us to obtain simple closed-form solutions but at the expense of rich comparative statics. In the second case, we assume that the market-making sector has constant relative risk aversion (CRRA) preferences and the stock payoff is distributed binomial. (For brevity, we solve only for independent liquidity shocks in the CRRA case.) The CRRA example yields richer comparative statics, but it can only be solved numerically.

A Example 1: CARA utility, normally distributed payoffs, and perfectly correlated liquidity shocks

In this example, we assume that (i) the stock payoff $\tilde{Z}$ is normally distributed with mean μ and variance σ^2, (ii) the market-making sector has the negative exponential utility function $u(w) = -e^{-\gamma \cdot w}$ where γ is the coefficient of absolute risk aversion, and (iii) liquidity shocks are perfectly correlated.

First, we derive the equilibrium pricing function, $p_0(\alpha)$. By assumption, this is the price at which the representative market-maker is indifferent between buying α

shares at date 0 and maintaining zero inventory of shares. Under the assumptions of CARA preferences and normally distributed stock payoffs, the price $p_0(\alpha)$ solves

$$\mathrm{E}[-e^{-\gamma\cdot\tilde{W}_2}] = \mathrm{E}[e^{-\gamma\cdot W_0}] \qquad \Longrightarrow \qquad \mathrm{E}[\tilde{W}_2] - \gamma\cdot\mathrm{Var}[\tilde{W}_2]/2 = W_0 \qquad . \tag{3}$$

where $\tilde{W}_2 \equiv W_0 + \alpha\cdot(\tilde{Z} - p_0)$. Solving yields $p_0(\alpha) = \mu - \gamma\cdot\sigma^2\cdot\alpha/2$.

Now we derive the price that obtains at date 1 in the event of a liquidity shock. Let $p_1(1-\alpha;\alpha)$ denote the price at which the representative market-maker is indifferent between buying $(1-\alpha)$ new shares at date 1 and maintaining an inventory of α shares. Under the assumptions of CARA preferences and normally distributed stock payoffs, the price $p_1(1-\alpha;\alpha)$ solves

$$\begin{aligned} &\mathrm{E}[\tilde{W}_2 + (1-\alpha)\cdot(\tilde{Z} - p_1)] - \gamma\cdot\mathrm{Var}[\tilde{W}_2 + (1-\alpha)\cdot(\tilde{Z} - p_1)]/2 \\ &= \mathrm{E}[\tilde{W}_2] - \gamma\cdot\mathrm{Var}[\tilde{W}_2]/2 \\ \Longrightarrow \quad &p_1(1-\alpha;\alpha) = \mu - (1+\alpha)\cdot\gamma\cdot\sigma^2/2 \qquad . \end{aligned} \tag{4}$$

Substituting $p_0(\alpha)$ and $p_1(1-\alpha;\alpha)$ into our definition of $F(\alpha)$ yields the following result:

Theorem 2 *If liquidity shocks are perfectly correlated across investors there is a unique symmetric Nash equilibrium with*

$$\alpha^\star = \begin{cases} \left(\frac{s}{1-s}\right) & \textit{if } s \le 1/2 \\ 1 & \textit{if } s > 1/2 \end{cases} \qquad . \tag{5}$$

Proof: Substitute the pricing functions into equation 2. Note that $F(0) > 0$ for all $s > 0$ and $F_{\alpha^\star}$, the derivative of F with respect to $\alpha^\star$, is negative. Thus, there are two possibilities. If $F(1) \ge 0$ then there is a unique pure strategy equilibrium $\alpha^\star = 1$ and if $F(1) < 0$ there is a unique mixed strategy, $\alpha^\star$, where $F(\alpha^\star) = 0$. For $s > 1/2$, $F(1) > 0$ thus $\alpha^\star = 1$. For $s \leq 1/2$ solving for $\alpha^\star$ yields the result. □

Equilibrium market-maker inventory increases in the probability of a liquidity shock. The efficient outcome would be for market makers to hold no inventory at date zero, but the desire of investors to preempt other investors forces the risk-averse market-making sector to inefficiently hold shares. This inefficient allocation of risk is reflected in a lower equilibrium price for the stock. Moreover, the market-maker inventory is convex in the liquidation probability. This is an "accelerator" effect: fear of other investors liquidating has an immediate influence on each investor's own decision to liquidate. For very small values of s, i.e., very little chance of future liquidity shocks, an investor sees other investors waiting and thus does not mind waiting herself. The market-making sector needs to hold almost no shares today ($\alpha^\star$ close to zero) and the outcome is close to the Pareto-optimum. With increasing s, the fraction of tendering investors rises ever more quickly since the first derivative of $\alpha^\star$ with respect to s is $\partial\alpha^\star/\partial s = 1/(1-s)^2$ which is increasing in s. In fact, even if there is "only" a 50-50 chance of investors facing a future liquidity shock, and even if the market-making sector is extremely risk-averse ($\gamma \to \infty$), risk-neutral investors find themselves unwilling to hold *any* stock today. Naturally, this is an extremely inefficient outcome.

Although these are not distinct equilibria, there is a flavor of two distinct scenarios here: a good scenario, in which the probability of liquidation is low, and the market-making sector is not holding much inventory; and a bad (or run) scenario, in which the probability of individual liquidation is average, and the risk-averse market-making sector has to absorb *all* shares in the economy.[7]

Interestingly, with CARA utility, the risk-absorption capacity of the market-making sector (γ) and the riskiness of the stock (σ) play no role in the equilibrium outcome ($\alpha^\star$). Expanding the market-making sector in both good and bad times would not solve the allocation problem created by the fear of facing a liquidity shock.[8] The reason is that there are two countervailing forces when the market-making sector is deep (or payoff variance is low): On the one hand, the average in-run price is higher because the market-making sector is close to risk neutral. On the other hand, the marginal price obtained after the run is also higher. In the case of constant absolute risk aversion preferences, these two effects exactly offset each other in the investors' selling decision. With CARA preferences, the market-making price is linear in inventory. Although risk aversion and payoff variance affect the slope of the linear demand curve, they do not affect the relation between average and marginal prices. Thus, the tradeoff between tendering today and waiting is independent of these parameters. The prime ingredient in this version of our model is investors' *fear* of future liquidation, s.

---

[7] It is straightforward to show if $s \leq 1/2$ the volatility of stock returns ($\tilde{R}_{0,1} \equiv {(p_1 - p_0)}/{p_0}$) is given by

$$\sigma(\tilde{R}_{0,1}) = \frac{\gamma \cdot \sigma^2 \cdot s}{2 \cdot (1-s) \cdot \mu - \gamma \cdot \sigma^2 \cdot s} \quad . \tag{6}$$

As expected, the underlying value volatility σ^2 is distinct from the stock return volatility. The volatility of stock returns increases in s, γ, and σ^2 and decreases in μ. Moreover, the change in volatility increases in s. Thus, seemingly small changes in liquidation probability s can significantly change market volatility.

[8] Of course, when the market-making sector is deep, prices are close to risk-neutral even if no risk-neutral investor is willing to hold shares and thus the welfare loss is small.

B Example 2: CARA utility, normally distributed payoffs, and independent liquidity shocks

The key difference between the perfectly correlated and independent liquidity shock cases is that in the former case, *all* investors who did not sell at date 0 must liquidate with probability s at date 1 whereas in the latter case, proportion s of investors who did not sell at date 0 must liquidate with probability 1 at date 1.

The derivation of the equilibrium price function at date 0 is the same in both cases. Thus, as we demonstrated above, if an individual investor conjectures that a total of α shares will be sold by individual investors to the market-making sector at date 0, she will still receive the price $p_0(\alpha) = \mu - \gamma \cdot \sigma^2 \cdot \alpha/2$ if she sells at date 0. However, as stated earlier, because liquidity shocks are independent across individual investors we know with probability one (by the *Law of Large Numbers*) that a proportion $(1-\alpha)\cdot s$ of shares will be liquidated at date 1. Let $p_1((1-\alpha)\cdot s; \alpha)$ denote the date-1 price set by the market makers when $(1-\alpha)\cdot s$ new sell orders arrive at date 1 and the market-making sector already held α shares. By assumption, $p_1((1-\alpha)\cdot s; \alpha)$ is the price at which the representative market-maker is indifferent between buying $(1-\alpha)\cdot s$ new shares at date 1 and maintaining an inventory of α shares. Under the assumptions of CARA preferences and normally distributed stock payoffs, the price $p_1((1-\alpha)\cdot s; \alpha)$ solves

$$
\begin{aligned}
& \mathrm{E}[\tilde{W}_2 + s\cdot(1-\alpha)\cdot(\tilde{Z} - p_1)] - \gamma\cdot\mathrm{Var}[\tilde{W}_2 + s\cdot(1-\alpha)\cdot(\tilde{Z} - p_1)]/2 \\
& = \mathrm{E}[\tilde{W}_2] - \gamma\cdot\mathrm{Var}[\tilde{W}_2]/2 \\
\Rightarrow \quad & p_1(s\cdot(1-\alpha); \alpha) = \mu - [2\cdot\alpha + (1-\alpha)\cdot s]\cdot\gamma\cdot\sigma^2/2 \quad . \qquad (7)
\end{aligned}
$$

Replacing the perfect correlation shocks $p_1((1-\alpha);\alpha)$ from the previous section with its independent shocks equivalent $p_1((1-\alpha)\cdot s;\alpha)$ in our definition of $F(\alpha)$ yields the following result:

Theorem 3 *If liquidity shocks are independent across investors there is a unique symmetric Nash equilibrium with*

$$\alpha^\star = \begin{cases} \left(\frac{s}{1-s}\right)^2 & \text{if } s \leq 1/2 \\ 1 & \text{if } s > 1/2 \end{cases} . \tag{8}$$

Proof: Substitute the pricing functions into equation 2, with $q_1(\alpha) = s\cdot(1-\alpha)$. Note that $F(0) > 0$ for all $s > 0$ and $F_{\alpha^\star}$, the derivative of F with respect to $\alpha^\star$, is negative. Thus, there are two possibilities. If $F(1) \geq 0$ then there is a unique pure strategy equilibrium $\alpha^\star = 1$ and if $F(1) < 0$ there is a unique mixed strategy, $\alpha^\star$, where $F(\alpha^\star) = 0$. For $s > 1/2$, $F(1) > 0$ thus $\alpha^\star = 1$. For $s \leq 1/2$ solving for $\alpha^\star$ yields the result. □

Again, market-maker inventory increases in the probability of a liquidity shock, and the model features the "accelerator" effect: fear of other investors liquidating has an immediate influence on each investor's own decision to liquidate. The first derivative of $\alpha^\star$ with respect to s is $\partial\alpha^\star/\partial s = 2\cdot s/(1-s)^3$. Thus, around $s = 0.23$, the fraction of investors that unload their shares onto the market-making sector changes one-to-one with changes in s. Above $s = 0.23$, even small changes in the perceived fraction of investors can cause large changes in market-making inventory and equilibrium pricing. Again, even if there is "only" a 50-50 chance of investors facing a future liquidity shock, and even if the market-making sector is extremely risk-averse ($\gamma \to \infty$), risk-neutral investors find themselves unwilling to hold *any* stock today.

Unlike in the perfectly correlated shocks version of our model, in this independent liquidity shock version of our model, we know with certainty that the price at date 1 will be lower than the price at date 0. This creates an arbitrage opportunity. One might reasonably ask why a market-maker would buy shares at date 0 when she knows for sure that the price will be lower at date 1. But this is "just" a model artifact caused by our law-of-large number assumption, though: In our model, there is no uncertainty about the number of liquidity shocked investors next period. In reality, however, there are many sources of uncertainty that make it possible that the price at date 1 will be higher than the price at date 0. For example, we could introduce some uncertainty about whether *any* liquidity shocks will appear. Or, a change in the environment which make liquidity shocks unlikely for some investors will cause them to buy the stock at date 1 and drive up its price. One might also reasonably ask why investors do not short the stock at date 0 and buy it back at date 1. Shorting stock, however, might be extremely difficult during such runs. Alternatively, uncertainty about the timing of the stock price bounce back can potentially introduce a source of risk (costly margin calls) that limits the aggressiveness of short positions at date 0 (see, e.g., Liu and Longstaff (2000)).[9]

A more pertinent question—because it is not an artifact of the lack of uncertainty in the number of liquidity shocked investors—is why noone simply waits to be a standby investor to buy only at the bottom of the crash. But, this question is bigger than just our model. What prevented an investor from becoming rich during the 1987 (or any other) crash? *Recent U.S. crashes and mini-crashes indeed showed immediate bouncebacks,* and it is these temporary liquidity and price drop phenomena that require (at least a partial) explanation.

[9]Rational expectations models relying on information are similarly concerned with breaking backward induction arbitrage links include Dow and Gorton (1994) and Allen, Morris, and Shin (2002). See also Shleifer and Vishny (1997).

There are a number of answers. First, there may be (unmodelled) uncertainty about the timing of the end of the crash and about a simultaneous revision in the expected return of the assets. Indeed, we sometimes see persistent market calamities following financial market crashes, we sometimes see sharp drops followed by immediate reversals—presumably with investors recognizing that their liquidity constraints will not bind, that they can safely join the market-making sector, and that the good state is about to return (see footnote 4). Such investors should do well. Indeed, we believe this is how temporary crashes ultimately end: with enough investors moving from the liquidity-fearing sector to the market-making sector. Second, execution in the final stage, the termination of the run, may be as uncertain as it is in the run initiation. A standby investor may try to wait until the bottom of the market to buy shares for a song, but he may not be sure whether his buy orders will be executed at the immediate last-investor run price or whether he may miss this opportunity altogether. After all, the price in our model drops sharply and then rebounds sharply back to μ. To the extent that other investors also wish to jump back in, and to the extent that sequential execution is fragile or non-existent, the spirit of our model is not incongruent with the fact that neither of the authors ended up excessively wealthy after the last stock market crash. Third, one may wonder why in a more dynamic context, there is not enough buffer stock to prevent such liquidity runs in the first place. But it is costly to create stand-by liquidity. For example, it would be costly for a large investor to carry zero-inventory most of the time (so as to be almost risk-neutral) and who is lurking around only for the opportunities presented in a crash. If crashes are rare, this may not be a profitable use of resources (Greenwald and Stein (1988, p.19)). Similarly, individuals may not find it in their interests to maintain a buffer stock of very liquid assets that could ensure them against the rare probabilities of liquidity shocks.

In sum, we believe liquidity runs and crashes to be sufficiently rare phenomena that moderating market forces may not be sufficiently profitable to grip instantaneously, but may require a short period of time.

C Example 3: CRRA utility

Although the CARA equilibrium illustrates the importance of the fear of liquidity shocks, the linearity of the market-maker's demand function reduces the richness of its comparative statics. To obtain a non-linear demand curve, we now assume that the representative market-maker has constant relative risk aversion (CRRA) preferences of the form $u(w) = \left(w^{1-\gamma}/1-\gamma\right)$ where γ is the coefficient of relative risk aversion. We further assume that the stock payoff $\tilde{Z}$ takes on one of two values: U with probability π and D with probability $1-\pi$. Absence of arbitrage requires $U > p_t > D$.

Again, the equilibrium pricing function, $p_0(\alpha)$, ensures that the representative market-maker is indifferent between buying α shares at date 0 and maintaining zero inventory of shares. Let $W_2(z) \equiv W_0 + \alpha \cdot (z - p_0)$ for $z = U, D$. In this example, the price $p_0(\alpha)$ solves

$$\left(\frac{1}{1-\gamma}\right) \cdot \left\{\pi \cdot [W_2(U)]^{1-\gamma} + (1-\pi) \cdot [W_2(D)]^{1-\gamma}\right\} = \left(\frac{1}{1-\gamma}\right) W_0^{1-\gamma}. \tag{9}$$

This price function is the same whether liquidity shocks are perfectly correlated or independent across investors.

If a liquidity shock occurs and they are perfectly correlated across investors the price $p_1(\,1-\alpha;\alpha\,)$ solves

$$\begin{aligned}
& \left(\frac{1}{1-\gamma}\right)\cdot\left\{\pi\cdot[W_2(U)+(1-\alpha)\cdot(U-p_1)]^{1-\gamma}+(1-\pi)\cdot[W_2(D)+(1-\alpha)\cdot(D-p_1)]^{1-\gamma}\right\} \\
= & \left(\frac{1}{1-\gamma}\right)\cdot\left\{\pi\cdot[W_2(U)]^{1-\gamma}+(1-\pi)\cdot[W_2(D)]^{1-\gamma}\right\}. \qquad (10)
\end{aligned}$$

If a liquidity shock occurs and they are independent across investors the price $p_1(\,(1-\alpha)\cdot s;\alpha\,)$ solves

$$\begin{aligned}
& \left(\frac{1}{1-\gamma}\right)\cdot\left\{\pi\cdot[W_2(U)+s\cdot(1-\alpha)\cdot(U-p_1)]^{1-\gamma}+(1-\pi)\cdot[W_2(D)+s\cdot(1-\alpha)\cdot(D-p_1)]^{1-\gamma}\right\} \\
= & \left(\frac{1}{1-\gamma}\right)\cdot\left\{\pi\cdot[W_2(U)]^{1-\gamma}+(1-\pi)\cdot[W_2(D)]^{1-\gamma}\right\}. \qquad (11)
\end{aligned}$$

Unfortunately, there are no closed-form expressions for the price functions, $p_0(\,\alpha\,)$, $p_1(\,(1-\alpha);\alpha\,)$, and $p_1(\,(1-\alpha)\cdot s;\alpha\,)$. However, a simple numerical algorithm can find them exactly. We can then use a simple search algorithm to find the equilibrium inventory of the market-making sector ($\alpha^\star$).

Although we do not have a general proof of uniqueness for $\alpha^\star$ nor for the monotonicity of the comparative statics for $\alpha^\star$ with respect to important parameters of the model, an examination of a large region of the relevant parameter space yields consistent results. For the numerical examples that follow, we have chosen reasonable base-case parameters to represent market-maker wealth (roughly 1/10 of the value of the risky asset) and risk-aversion ($\gamma = 3$). For all cases, it can be shown that $\alpha^\star$ is unique. Figure 1 graphs the market-making sector's equilibrium holdings ($\alpha^\star$) as a function of exogenous parameters for the case of independent liquidity shocks across investors. The numerical results are qualitatively similar when liquidity shocks are perfectly correlated across investors. Typically, we find that the market-making sector holds more inventory ($\alpha^\star$)

- when the market-making sector has greater wealth;
- when the market-making sector has greater risk-absorption capacity (risk-aversion coefficient γ is lower);
- when the asset is less risky ($U - D$ is smaller holding the mean payoff $\pi U + (1-\pi)D$ constant);
- when the probability of a liquidity shock (s) is higher.

We already know from the CARA case that it is not the steepness of the demand curve itself (i.e., the "depth") that matters to market-making inventory: higher risk-capacity for the market-making sector does not only allow investors to unload shares at an attractive price in a run, but it also allows them to enjoy a better price after a run. Instead, what matters is the second derivative of the demand curve. Just as in the CARA case, fear of liquidity can cause the run, and the sensitivity of α to s increases in s. But, in the CRRA case, the other parameters (such as wealth, risk-aversion, and riskiness) matter for the relative share allocations to the extent that they bend the market-making demand function.

III Equilibrium With Margin Constraints

Margin calls, which force investors to sell more shares if the share price declines, are well-known to be important during financial market crashes (see, e.g., Chowdhry and Nanda (1998)). Our model does not require margin calls, but margin calls can endogenize liquidity constraints, and can produce very high-frequency "phase transitions" (as well as multiple equilibria). They can amplify the already existing liquidity run accelerator effect.

We now sketch a simple model of margin constraints. As before, the stock payoff is normally distributed with mean μ and variance σ^2. The representative market-maker has CARA utility with risk aversion parameter γ. Suppose that every investor has financed the purchase of the stock with margin. We assume that the investor will face a margin call, and will be forced to liquidate if her wealth (including the stock) falls too much. If an investor cannot meet the margin requirement, she will be forced to liquidate her share at date 1. If she can meet the margin requirement, it will be optimal for her to do so and hold the stock to date 2. Below, we propose a reasonable specification contrasting the component of external income that can be allocated to covering margin constraints to the amount of margin that investors have to provide.

Suppose, for example, a representative investor had purchased shares at price p (determined outside the model), and let $m \in [0, 1]$ represent the proportion of the investment financed with margin. Thus, if the price falls from p to p_0, this investor would need to come up with cash of $m \cdot (p - p_0)$ in order to hold onto the shares until the final period. One could then interpret the margin constraints to be triggered by a decline in price from the purchase price p to p_0 (a return), rather than just being driven by a low price level per se.

Each investor has an external source of income at date 0. This external income $\tilde{W}$ is assumed to be distributed uniformly over the interval $[0, B]$ and is independent of any stock price movements. Therefore, the endogenous probability of liquidation is given by

$$\begin{aligned} s(\alpha) &= \text{Prob}[\tilde{W} < m\cdot(p - p_0)] = \frac{m\cdot(p-p_0)}{B} = \frac{m\cdot(p-\mu)}{B} + \left(\frac{m\cdot\kappa}{B}\right)\cdot\alpha \\ &\equiv c_0 + c_1\cdot\alpha \qquad , \end{aligned} \tag{12}$$

where $\kappa \equiv \gamma\cdot\sigma^2/2$, $c_0 \equiv {}^{m\cdot(p-\mu)}/_{B}$, and $c_1 \equiv {}^{m\cdot\kappa}/_{B}$ (> 0). We consider only the case where the income shocks are perfectly correlated, i.e., *all* or *no* investors face a liquidity shock at date 1 depending on the realization of $\tilde{W}$.[10]

Theorem 4 *If (i) $B > p > \mu > \kappa$ where $\kappa \equiv \gamma\cdot\sigma^2/2$; (ii) B is sufficiently large; and (iii) income shocks are perfectly correlated, then there is a unique tendering equilibrium, $\alpha^* \in (0,1)$, in which α^* increases in m, κ, and $(p-\mu)$ and decreases in B.*[11]

Proof: In the perfectly correlated income shocks case we have $p_0 = \mu - \kappa\cdot\alpha$ if α proportion tender at date 0, and $p_1 = \mu - \kappa\cdot(1+\alpha)$ if all remaining investors are forced to liquidate at date 1. Substituting these two price functions and the liquidation probability (12) into (2) yields

$$F(\alpha) \equiv \kappa\cdot[(c_0 + c_1\cdot\alpha)\cdot(1+\alpha) - \alpha] \qquad . \tag{13}$$

First, note that $F(0) = \kappa\cdot c_0 > 0$, so $\alpha = 0$ is not an equilibrium. Second, note that $F(1) = \kappa\cdot[2\cdot(c_0 + c_1) - 1] < 0$ by assumptions (i) and (ii), so $\alpha = 1$ is not an equilibrium. Finally, $F_\alpha = \kappa\cdot[c_0 + c_1 + 2\cdot c_0\cdot\alpha - 1] < 0$ by assumptions (i) and (ii), so there is a unique $\alpha^* \in (0,1)$ such that $F(\alpha^*) = 0$.

The comparative statics results follow from the facts that (i) $F_{c_0} = \kappa\cdot(1+\alpha) > 0$; (ii) $F_{c_1} = \kappa\cdot\alpha\cdot(1+\alpha) > 0$; (iii) c_0 is increasing in m and p and decreasing in B and μ; (iv) c_1 is increasing in m and κ and decreasing in B; and (v) $F_\kappa = [(c_0 + c_1\cdot\alpha)\cdot(1+\alpha) - \alpha] + \kappa\cdot\alpha\cdot(1+\alpha)\cdot\left(\frac{\partial c_1}{\partial\kappa}\right) > 0$ when B is sufficiently large.

[10] The independent liquidity shocks case solution involves cubic equations. It is difficult to find the parameter restrictions ensuring a unique equilibrium so that we can do the comparative statics. However, there is no reason why the intuition of the equilibrium discussed in this section would not carry over to the independent liquidity shock scenario.

[11] The restrictions (i) $B > p > \mu > \kappa$ and (ii) B sufficiently large are made for the following reasons: The assumption $\mu > \kappa$ ensures that $p_0 > 0$ so we get a meaningful margin requirement. The assumption $p > \mu$ ensures that we are considering cases where the price of the stock has fallen from the time it was purchased so that we can consider the effect of margin calls. Finally, the assumption B sufficiently large ensures an interior solution for α^*. If B is not large enough the probability of not meeting the margin call at date 1 is very high and the only equilibrium we get at date 0 is everyone tendering (which has no interesting comparative statics).

The intuition for the comparative statics are straightforward:

Margin Constraint (m): the more investors can borrow, the greater will be the tendering proportion at date 0, because it is less likely that investors will be able to meet margin calls at date 1.

Original Purchase Price (p): Similarly, the more investors paid relative to the current mean ($p - \mu$), the greater the tendering proportion at date 0, because it is less likely investors will be able to meet the margin call at date 1. One way to interpret $p - \mu$ is the innovation in beliefs from purchase of the stock to now. This states that if there is a big negative shock to beliefs, then margin calls exacerbate the price move through early liquidation: this is overreaction to bad news. However, there is no overreaction to good news, because there is not a margin call in that case!

Effective Variance (κ): The greater the effective variance, the more likely investors face margin calls at date 1, because the market-makers' demand curve is more steeply sloped. Thus, the higher is the tendering proportion at date 0.

Expected Income (B): The higher external expected income is likely to be (to meet future margin calls), the smaller is the tendering proportion at date 0.

Our particular underlying margin assumptions have produced the particular linear mapping of price declines into liquidation probabilities in (12). However, this equation should only be considered an illustrative sketch *for brevity of exposition*: in the real world, different investors would purchase at different price points and face different margin constraints. The important aspect of our sketch is the presence of some positive feedback trading in the event of a crash, where a lower price can also increase the probability of future liquidation needs, which gave us an explicit specification for $s(p_0)$ and therefore $s(\alpha)$. The liquidity run phenomenon then interacts with and rationally amplifies this feedback trading (Shleifer (2000)). One could also entertain altogether different mechanisms that accomplish the link from stock prices to liquidation probabilities. For example, one could write down a model in which risk management systems, principal-agent problems, or limited horizons induce investors to be more likely to liquidate when stock prices fall. Or, one could estimate an empirical relation on the data itself without writing down a specific model. Thus, one could entertain other reduced-form price-liquidation probability linkages. Indeed, for the special case $c_0 = 0$ and $c_1 = 1$, there are three equilibria: one stable one in which no investor tenders and therefore no investor is afraid of liquidation; a stable one in which every investor tenders *because* every investor tenders; and one in which there is an interior tendering equilibrium. In this case, one might even observe a sudden equilibrium switch, where one moment noone was afraid of a liquidity shock and the next moment everyone is afraid of liquidity shock. The relationships between prices and liquidation probabilities need not be linear either. However, further such modeling could detract from the main intuition of our paper: when there are some feedback effects from the underlying stock price to liquidation probabilities, the fear of future needs to liquidate can cause potentially rapid and violent liquidity runs.

IV Discussion, Extensions, and Welfare

A Preventing Runs and Front-Running: Time-Varying Market Depth

The obvious question is what mechanisms could prevent the need for the market-making sector to absorb run inventory from the public.

The first answer lies in the enforcement of perfect sequentiality. With sequential execution the last investors (who now know they are the last investors!) would be better off just waiting it out instead of being the last in-the-run investors. This can unravel the tendering equilibrium. In response to the 1987 crash, the NYSE massively expanded its communication infrastructure, a mechanism to prevent the conversion of the sequential market into a random-execution market in times of declines.

Interestingly, a belief that one can front-run others (get their share sales executed with higher priority) can encourage run equilibria because it increases the expected payoff to tendering early.[12] Naturally, in an equilibrium with homogeneous agents, noone can expect to front-run anyone else. However, in a real-world context, some heterogeneous investors may rationally or irrationally believe in their ability to front-run. Portfolio insurance may be an example of a strategy that attempts to precommit to withdraw funds in the case of large moves, which will thus worsen the liquidity effects described in our own paper. Leland and Rubinstein (1988, p.46f) describe some possible front-running in 1987: "With the sudden fall in the market during the last half hour of trading on October 16, many insurers found themselves with an overhang of unfilled sell orders going into Monday. In addition, several

[12]Within the context of the informational cascades literature (Bikhchandani, Hirshleifer, and Welch (1992), Welch (1992)), Chen (1995b) has modelled such informational interactions in a banking run context.

smart institutional traders knew about this overhang and tried to exit the market early Monday before the insurers could complete their trades."

The second answer lies in providing liquidity during runs. If the market-making sector were to expand *only* in "bad" situations but offer lousy execution in "good" situations, then each investor would be relatively better off not trading into the market today, and instead would be relatively more eager to try to wait it out. It is straightforward to solve a model that proves this point. For simplicity, we analyze the case in which the representative market-maker has CARA utility and the stock payoff is normally distributed. We also assume that the liquidity shocks are perfectly correlated across investors (the independent liquidity shocks case yields similar qualitative results). In this example, we allow the market-making sector to be deeper at date 1 than at date 0 (i.e., $\gamma_{t=1} \leq \gamma_{t=0}$). Again, we assume that prices are set at each date by a zero-utility condition. However, the new prices reflect the different market-making depth at each date. Thus, $p_0(\alpha) = \mu - \gamma_0\sigma^2\alpha/2$ and $p_1(1-\alpha;\alpha) = \mu - (1+\alpha)\gamma_1\sigma^2/2$. Substituting $p_0(\alpha)$ and $p_1(1-\alpha;\alpha)$ into our definition of $F(\alpha)$ yields the following result:

Theorem 5 *If the market-making sector is deeper at date 1 (i.e., $\gamma_1 < \gamma_0$) and liquidity shocks are perfectly correlated there is a unique symmetric Nash equilibrium with*

$$\alpha^\star = \begin{cases} \left(\frac{s}{\gamma_0/\gamma_1 - s}\right) & \text{if } s \leq \gamma_0/(2\cdot\gamma_1) \\ 1 & \text{if } s > \gamma_0/(2\cdot\gamma_1) \end{cases} . \tag{14}$$

Proof: Substitute the pricing functions into equation 2. Note that $F(0) > 0$ for all $s > 0$ and $F_{\alpha^\star}$, the derivative of F with respect to $\alpha^\star$, is negative since $\gamma_1 < \gamma_0$. Thus, there are two possibilities. If $F(1) \geq 0$ then there is a unique pure strategy equilibrium $\alpha^\star = 1$ and if $F(1) < 0$ there is a unique mixed strategy, $\alpha^\star$, where $F(\alpha^\star) = 0$. For $s > \gamma_0/(2\cdot\gamma_1)$, $F(1) > 0$ thus $\alpha^\star = 1$. For $s \leq \gamma_0/(2\cdot\gamma_1)$ solving for $\alpha^\star$ yields the result. □

Market-making inventory $\alpha^\star$ decreases in γ_0 and increases in γ_1. If early on, market makers are more risk-averse (γ_0 is high), investors are less eager to tender to market-makers at time 0 ($\alpha^\star$ is low) and more inclined to take the chance of being forced to sell if personally hit by a subsequent liquidity shock. Conversely, if the subsequent "standby liquidity" in a crisis is low, because the market-making risk aversion γ_1 is then unusually high, the post-run price will be lower, which prompts investors to be more eager to sell at date 0. Casual empiricism suggests that, if anything, the market-making sector becomes intrinsically *more* risk-averse during runs than it is in ordinary times. Thus, government intervention which commits to provide market-depth in "bad" *but not in "good" times* might usefully mitigate run inefficiencies. This gives a natural interpretation to government intervention: if correctly done, standby liquidity in market runs could help prevent runs in the first place. (It is unlikely that the private sector could provide unusually good liquidity *only* in bad scenarios, but not in good scenarios.) Interestingly, non-intervention in good markets is as important as intervention in bad markets! Indeed, this is the equivalent of the national petroleum reserves, which are rarely released, but whose presence may in itself prevent runs. Greenwald and Stein (1988, p.19) discuss an alternative mechanism, in which large financial insurers would agree to cover some of the losses of market-makers if the market drops significantly. However, there is anecdotal evidence that many such institutions often run portfolio-insurance schemes and tend to sell more into a crash rather than against a crash (footnote 4).

B A Multi-Period View

The single-period setting of our model is important in one sense but not another. Our model requires that uncertainty is resolved while market-makers hold shares that investors—fearing liquidity shocks—have offloaded on them. Consequently, we do not believe our model applies to situations in which market-makers can leisurely unravel their holdings with little risk (e.g., over many periods without uncertainty). However, our model is robust to multiple trading periods *before* a potential liquidity shock can come about. That is, given many trading opportunities prior to a potential liquidity shock, every investor would want to offload shares immediately to avoid having to trade behind other investors. In periods between this first period and the period of the potential future shock, investors voluntarily do not trade. Thus, a liquidity run can occur even if the liquidity shock is far away.

This is easy to show. Suppose investors now have two opportunities to sell, denoted date 0 and date 1, prior to the occurance of a liquidity shock. Now, an equilibrium is a pair (α_0, α_1), for which—given that α_0 proportion sell at date 0—it is optimal for an α_1 proportion to sell at date 1, and vice-versa. One condition for optimality is that someone who sells at one date does not have the incentive to deviate and sell at the other date. But there is only one case for which this is true: α proportion sell at date 0 and no one sells at date 1! In this case, the date 0 price exceeds the date 1 price so no one has an incentive to deviate and sell at date 1. Moreover, the possibility of a liquidity shock in the future makes an investor indifferent between selling at date 0 and waiting if she conjectures that $\alpha^\star$ (as in our earlier model) proportion of investors will sell at date 0. Note that the opposite is not an equilibrium, i.e. no one sells at date 0 and α proportion sell at date 1 because the date 0 price will be higher, permitting investors to front-run and thus profit by selling at date 0. In sum, the only equilibrium is the same $\alpha^\star$ as in our earlier model: first, active selling at date 0, followed by no more selling until the date at which the liquidity shock occurs.[13]

However, the investor liquidity shocks *must* be simultaneous: it is the period with the highest probability of liquidity shocks that would matter, not the average or cumulative probability of a liquidity shock. Luckily, many constraints seem to appear at roughly the same relevant time for many participants: a high-frequency financial market drop or a low-frequency economic depression may force many individuals to seek liquidity at the same time.

Naturally, changes in model parameters might cause some readjustments as time goes by. We have already sketched the influence of time-varying market depths in Subsection A. Similarly, one could imagine time-varying probability assessments of *future* liquidity shocks, which could lead to active trading and time-varying market-making inventory adjustments, even in the absence of any *current* liquidity shocks. (Incidentally, such a model can easily explain relatively high trading volume in the presence of only mild news.)

C The Social Cost of Investor Fear

In our model, there is no asymmetric information or trading costs–and yet the market outcome can be significantly worse than the Pareto-optimal allocation. In what follows, we analyze the social cost of investor fear when market-makers have CARA utility and the stock payoff is normally distributed. Our analysis considers exogenous liquidity shocks which are either perfectly correlated or independent across investors.

[13] Note that in this independent shocks version of our model, we must again resort to a competitive market-making sector to prevent an individual market-maker to exploit to wait for the vulture opportunity.

C.1 Perfectly correlated liquidity shocks

In what follows, our benchmark is *not* a price of μ, but a requirement that risk-neutral investors must not sell at date 0 (similar to the analysis in Diamond and Dybvig (1983)). In this Pareto-optimal outcome, the risk-neutral investors hold all the shares at date 0 and sell to the market-making sector at date 1 *only if* they are actually hit by a liquidity shock. Consequently, every investor would sell shares with probability s at a price $p_1 = \mu - {}^{\gamma \cdot \sigma^2}/_2$ (assuming that the market-maker sector executes these sell orders at a price that yields no utility gain for them) and would retain shares with probability $1 - s$ (with expected value μ). Thus, investors' utility would be

$$\mu - \frac{\gamma \cdot s \cdot \sigma^2}{2} \quad . \tag{15}$$

In the batch-execution model, risk-neutral investor sell with probability $\alpha^\star$ at date 0 at the average price $p_0 = \mu - (\gamma \cdot \alpha^\star \cdot \sigma^2)/2$, liquidate with probability $(1 - \alpha^\star) \cdot s$ at date 1 at the average price $p_1 = \mu - (1 + \alpha^\star) \cdot \gamma \cdot \sigma^2]/2$, and retain shares with probability $(1 - s) \cdot (1 - \alpha^\star)$ at expected value μ. Thus, investors' utility is

$$\begin{cases} \mu - \frac{\gamma \cdot \sigma^2}{2} \cdot \frac{s}{(1-s)} & \text{if } s \leq 1/2 \\ \mu - \frac{\gamma \cdot \sigma^2}{2} & \text{if } s > 1/2 \end{cases} \tag{16}$$

By assumption, the market making sector has zero expected utility gain thus a total welfare comparison only requires a comparison of the investors' utility. Simple algebra shows that equilibrium welfare (expected selling price) is below the Pareto-optimal level of welfare by the amount:

$$\begin{cases} \frac{\gamma \cdot \sigma^2}{2} \cdot \frac{s^2}{(1-s)} & \text{if } s \leq 1/2 \\ \frac{\gamma \cdot \sigma^2}{2} \cdot (1-s) & \text{if } s > 1/2 \end{cases} \tag{17}$$

The Pareto-inferior outcome is caused by a prisoner's dilemma among risk-neutral investors that cannot easily be overcome. The welfare loss is increasing in the market-maker's risk-aversion and the payoff variance σ^2 since inefficient risk-sharing is exacerbated. Finally, the welfare loss is greatest when $s = 1/2$ because the market-making sector must absorb *all* shares in this case, not just those of the liquidity-shocked individuals. Because $\alpha^\star$ increases at a faster rate as s approaches $1/2$, the welfare loss increases in s for $s \in [0, 0.5)$. However, because (i) $\alpha^\star = 1$ for all $s \geq 1/2$ and (ii) as s increases the market makers would hold an increasing proportion of shares in the Pareto-optimal outcome, the welfare loss decreases in s for $s \in (0.5, 1]$.

C.2 Independent liquidity shocks

In the Pareto-optimal outcome, the risk-neutral investors hold all the shares and sell to the market-making sector *only if* they are actually hit by a liquidity shock. In this case, investors sell shares with probability s at a price $p_1 = \mu - {\gamma \cdot s\sigma^2}/2$ (assuming that the market-maker sector executes these sell orders at a price that yields no utility gain for them) and would retain shares with probability $1 - s$ (with expected value μ). Thus, investors' utility would be

$$\mu - \frac{\gamma \cdot s^2 \cdot \sigma^2}{2} \quad . \tag{18}$$

In the batch-execution model, risk-neutral investor sell with probability $\alpha^\star$ at date 0 at the average price $p_0 = \mu - (\gamma \cdot \alpha^\star \cdot \sigma^2)/2$, liquidate with probability $(1 - \alpha^\star) \cdot s$ at date 1 at the average price $p_1 = \mu - [2 \cdot \alpha^\star + (1 - \alpha^\star) \cdot s] \gamma \cdot \sigma^2 / 2$, and retain shares with probability $(1 - s) \cdot (1 - \alpha^\star)$ at expected value μ. Thus, investors' utility is

$$\begin{cases} \mu - \frac{\gamma \cdot \sigma^2}{2} \cdot \frac{s^2}{(1-s)^2} & \text{if } s \le 1/2 \\ \mu - \frac{\gamma \cdot \sigma^2}{2} & \text{if } s > 1/2 \end{cases} \tag{19}$$

As in the perfectly correlated case, the welfare loss is increasing in γ and σ^2 and the welfare cost is greatest when $s = 1/2$.

D Some Final Thoughts

Contagion effects across investors fall naturally out of the model. In the bad scenario, there are spillovers in the decisions of investors to sell their shares. This causes each individual investor to fear that he may have to sell (for exogenous reasons) *behind* every other investor. If selling late, he will get only the marginal price *after* everyone else has already sold to the market-making sector, which—already being burdened with the inventory of all other investors—can only offer a very low price.

The negative payoff externalities among investors causes an accelerator effect, in which just small increases in the probability of future liquidity shocks cause a large layoff of risky shares onto the risk-averse market-making sector. Again, the accelerator effect does not amplify the effects of the actual liquidity shock! It amplifies the extent to which one investor's *fear* of a future liquidity shock has a negative spillover on other investors' fears.

The idea of financial contagion and spillover effects across markets and institutions, as modelled, e.g., in Allen and Gale (2000), also apply naturally to our scenario. By expanding the domain of liquidity runs from financial institutions to financial markets, our model suggests that cross-liquidity constraints could be more important than previously thought. Financial crises could transmit not only from one institution to another, but across both financial institutions and financial markets.

We have repeatedly pointed out that runs are not caused by liquidity shocks themselves, *but by fears of future liquidity shocks*. The probability of a future liquidity shock may constantly fluctuate, even though the liquidity shock itself can be off on the horizon. Consequently, an empiricist can observe dramatic price movements and market-making inventory changes without observing any actual liquidity shocks. And, for the rare empiricist able to measure the *fear* of liquidity shocks (s), depending on its value, seemingly small changes can cause large sudden changes in the desire of investors to unload shares onto the market-making sector.[14]

V Related Literature

Our *financial markets* runs model has both similarities and differences to the *financial intermediation* runs models, foremost Diamond and Dybvig (1983). Our model is also driven by investor liquidity shocks and payoff externalities. However, our model does not require a sequential service constraint, productive inefficiencies, or a total loss if an investor fails to join a run.[15] Indeed, working out the endogenous

[14]Although our model has emphasized purely rational behavior, where the fear of liquidity shocks is rationally assessed or derived from margin constraints, our equilibrium could also be embedded in a world of "non-rational behavioral economics," if the fear of a liquidity shock (the need to terminate an investment early during a market run) were itself non-rational.

[15]Allen and Gale (2000) and similar financial contagion models, though quite different, build on the Diamond and Dybvig framework and retain these two assumptions. The same can be stated

pricing and market inventory is a major focus of our paper. Also, in both cases, a "lender of last resort" can prevent the run.

There is also related literature on stock market crashes. Grossman and Miller (1988) present a two trading period model in which all traders are not simultaneously present in the market. In the first period, there is a temporary order imbalance which must be absorbed by market makers. Between the first and second period, new information arrives about the security so the market-making sector is exposed to risk. However, the market-making sector is small and has low risk absorption capacity. Thus, the equilibrium price falls more than if all traders were available to absorb the imbalance. In the second period, the remaining traders arrive to buy some of the market makers inventory and the price rises. The key feature of their model which produces crashes is the asynchronous arrival of traders in the market, combined with the limited risk-bearing capacity of market makers. Greenwald and Stein (1991) extend the Grossman and Miller (1988) analysis by assuming that traders can only submit market orders in the second period of trade. This introduces transactional risk (uncertainty about the price at which their trades will execute) which reduces the willingness of buyers to absorb the market makers inventory in the second period. Knowing this, the market makers demand a larger risk premium in the first period to absorb the temporary order imbalance, which causes prices to fall even further than in the Grossman and Miller analysis. Like Greenwald and Stein (1991), we permit only market orders and have an uncertain execution price.

Although our model is closest in spirit Greenwald and Stein (1991), it is quite different. First and foremost, runs in our model occur when investors think that

---

for the liquidity crisis and international runs on currency reserves literature, e.g., Caballero and Krishnamurthy (2001). Geanakoplos (2001) embeds collateral crises into a broader model.

others will tender and therefore choose to tender themselves. In Greenwald and Stein, there are no runs (defined as "I tender because I think you will tender") in the spirit of bank runs. Rather, Greenwald and Stein (1991) have crashes in price because the uncertain execution price with batch orders introduces an extra risk which makes the price fall further than if investors could submit limit orders. Second, our financial market runs are endogenous and are not driven by asynchronous trading arrivals or exogenous supply shocks. Indeed, the Greenwald and Stein (1991) model is driven by the uncertainty in the number of arriving value traders, an uncertainty which does not even exist in our model. Third, the negative externality in their model is that value buyers may destroy the value opportunities for other buyers. In our model, the negative externality derives from investor selling, not buying. Further, this ever-present negative externality forces market-makers to hold a socially suboptimal inventory of shares. And finally, they argue that circuit-breakers might help: in our model, circuit-breakers are counterproductive.

There are also other areas of research more distant in spirit, but which also explain facets of financial market crashes. There is a large literature examining the impact of portfolio insurance (e.g., Grossman (1988), Brennan and Schwartz (1989), Genotte and Leland (1990), Jacklin, Kleidon, and Pfleiderer (1992), Donaldson and Uhlig (1993), Grossman and Zhou (1994), Basak (1995)). Portfolio insurers are usually modelled as agents who display positive feedback trading (of an accelerating kind) for exogenous (often assumed) reasons. This literature's primary goal is to show that portfolio insurers can exacerbate crashes (discontinuous movements). Models differ in choosing discrete single-shot vs. continuous time modelling techniques, implications on what portfolio insurance does for general price volatility in ordinary markets, and in how asymmetric information matters. Our own model differs from this literature in that the reason for selling is not the (usually exogenous) consumption motive, but the direct negative externality arising endogenously from other investors' trading.

Other papers have also presented ingenious mechanisms that can elicit large price changes. In Madrigal and Scheinkman (1997), an informed strategic market-maker attempting to control both the order flow she receives and the information revealed to the market by the prices she sets may choose an equilibrium price schedule that is discontinuous in order flow thus prompting large changes in price for arbitrarily small changes in market conditions. In Romer (1993), uncertainty about the quality of others' information is revealed by trading, and large price movements, such as the October 1987 crash, may be caused not by news about fundamentals but rather by the trading process itself. In Sandroni (1998), market crashes can be a self-fulfilling prophecy when agents have different discount rates and different beliefs about the likelihood of rare events (even if these beliefs converge in the limit). Barlevi and Veronesi (2001) present a model in which uninformed traders precipitate a stock price crash because as prices fall they rationally infer that informed traders have negative information which leads them to reduce their demand for the stock and drive its price even lower. The key feature of their model is that the uninformed traders have locally upward sloping demand curves which, when combined with the informed's downward sloping demands, can generate an equilibrium price function discontinuous in fundamentals.

Finally, we are not the first to employ margin constraints to generate (multiple) equilibria. In Chowdhry and Nanda (1998), perhaps the paper most similar to our own endogenous liquidity constraint section, some investors engage in margin borrowing to obtain their desired investment portfolio. Because shares can be used as collateral there is a link between the price of the stock and the capacity to invest in it which introduces the possibility of multiple equilibria. For example, lower (higher) stock prices can be a self-fulfilling equilibrium because it diminishes (increases) the capacity for levered investors to purchases their desired amount of stock which in turn makes the price fall (rise) rational.[16]

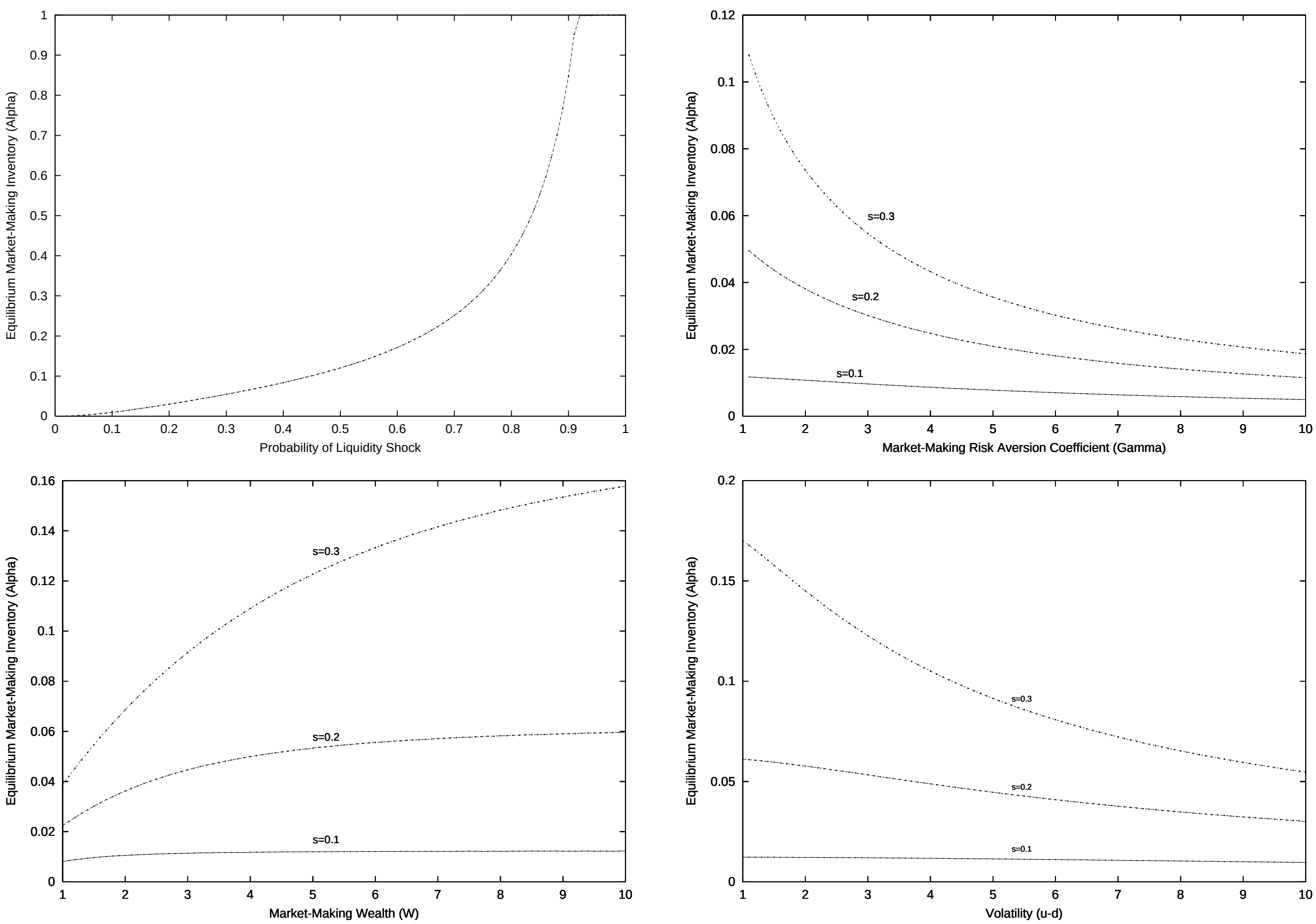
VI Conclusion

Our paper has developed a theory of financial market runs: socially inefficiently large market-making inventory in batch or random-ordering financial markets. Batch markets are the standard stock market opening mechanism and auction mechanism on some foreign exchanges. Random-ordering markets are common in many over-the-counter markets. They also can occur (infrequently) after a large price drop, when limited communication channels between investors and the financial system fail and break down the perfect sequentiality of execution. In such cases, investors' *fears* of *future* liquidity constraints can cause a prisoner's dilemma among investors *today*. This destroys efficient risk-sharing and aggravates any fundamental price drops.

Aside from sequential execution and reasonable *fear* of liquidity shocks (but not necessarily actual liquidity shocks), our model required very little machinery. Thus, it is the (presumably rare) combination of breakdown of sequential execution and a common fear of liquidity shocks, perhaps caused by or related to margin constraints, that facilitates a run on a financial market.

[16]For a more recent example, Yuan (2000) demonstrates that margin constraints can be beneficial because they may apply to informed investors and thus reduce the adverse selection problem with uninformed investors.

Figure 1. Comparative Statics under Market-Making CRRA Utility



Comparative statics when investors face independent liquidity shocks. Our base parameters are a down-stock-value of $D = 10$ and an up-stock-value $U = 20$ with equal probability $\pi = 0.5$, a risk aversion coefficient of $\gamma = 3$, and market-making wealth of $W = 1.5$ (i.e., roughly $1/10$ of the value of the financial market).

References

Abreu, D., and M. K. Brunnermeier, 2001, "Bubbles and Crashes," working paper, Princeton University.

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

Allen, F., S. Morris, and H. S. Shin, 2002, "Beauty Contests, Bubbles and Iterated Expectations in Asset Markets," working paper, University of Pennsylvania and Yale University and London School of Economics.

Amihud, Y., H. Mendelson, and R. A. Wood, 1990, "Liquidity and the 1987 Stock Market Crash," *Journal of Portfolio Management*, 16(3), 65–69.

Barlevi, G., and P. Veronesi, 2001, "Rational Panics and Stock Market Crashes," working paper, University of Chicago.

Basak, S., 1995, "A General Equilibrium Model of Portfolio Insurance," *Review of Financial Studies*, 8(4), 1059–1090.

Bikhchandani, S., D. Hirshleifer, and I. Welch, 1992, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100(5), 992–1026.

Brennan, M. J., and E. Schwartz, 1989, "Portfolio Insurance and Financial Market Equilibrium," *Journal of Business*, 62, 455–472.

Caballero, R. J., and A. Krishnamurthy, 2001, "International Liquidity Illustion: On the risks of sterilization," working paper, MIT and Northwestern.

Chen, Y., 1995b, "Banking Panics: the role of the first-come, first-served rule and informational externalities," working paper, UCLA.

Chowdhry, B., and V. Nanda, 1998, "Leverage and Market Stability : The Role of Margin Rules and Price Limits," *Journal of Business*, 71(2), 179–210.

Dennis, P. J., and D. Strickland, 2002, "Who blinks in Volatile Markets, Individuals or Institutions," *The Journal of Finance*, ?, forthcoming.

Diamond, D., and P. Dybvig, 1983, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 91, 401–419.

Diamond, D. W., and R. E. Verrecchia, 1991, "Disclosure, Liquidity, and the Cost of Capital," *The Journal of Finance*, 46(4).

Donaldson, R. G., and H. Uhlig, 1993, "The Impact of Large Portfolio Insurers on Asset Prices," *The Journal of Finance*, 48(5), 1943–1953.

Dow, J., and G. Gorton, 1994, "Arbitrage Chains," *The Journal of Finance*, 49(3), 819–849.

Gammill, J. F., and T. Marsh, 1988, "Trading Activity and Price Behavior in the Stock and Stock Index Futures Markets in October 1987," *Journal of Economic Perspectives*, 2(3), 25–44.

Geanakoplos, J., 2001, "Liquidity, Default and Crashes: Endogenous Contracts in General Equilibrium," working paper, Yale University, Cowles Foundation.

Genotte, G., and H. Leland, 1990, "Market Liquidity, Hedging, and Crashes," *American Economic Review*, 80, 999–1021.

Greenwald, B., and J. Stein, 1988, "The Task Force Report: The Reasoning Behind the Recommendations," *Journal of Economic Perspectives*, 2(3), 3–23.

——— , 1991, "Transactional Risk, Market Crashes, and the Role of Circuit Breakers," *Journal of Business*, 64, 443–462.

Grossman, S., and M. Miller, 1988, "Liquidity and Market Structure," *Journal of Finance*, 43, 617–633.

Grossman, S. J., 1988, "An Analysis of the Implications for Stock and Futurs Price Volatility of Program Trading and Dynamic Trading Strategies," *Journal of Business*, 60, 275–298.

Grossman, S. J., and Z. Zhou, 1994, "Equilibrium Analysis of Portfolio Insurance," *The Journal of Finance*, 51(4), 1379–1403.

Jacklin, C. J., A. W. Kleidon, and P. Pfleiderer, 1992, "Understimation of Portfolio Insurance and the Crash of October 1987," *Review of Financial Studies*, 5(1), 35–63.

Leland, H., and M. Rubinstein, 1988, "Comments on the Market Crash: Six Months After," *Journal of Economic Perspectives*, 2(3), 45–50.

Liu, J., and F. A. Longstaff, 2000, "Losing Money on Arbitrages: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities," working paper, UCLA.

Madrigal, V., and J. A. Scheinkman, 1997, "Price Crashes, Information Aggregation, and Market-Making," *Journal of Economic Theory*, 75(1), 16–63.

Romer, D., 1993, "Rational Asset-Price Movements without News," *American Economic Review*, 83(5), 1112–1130.

Sandroni, A., 1998, "Learning, Rare Events, and Recurrent Market Crashes in Frictionless Economies without Intrinsic Uncertainty," *Journal of Economic Theory*, 82(1), 1–18.

Shleifer, A., 2000, *Inefficient markets - An Introduction to Behavioral Finance*. Oxford University Press, Oxford, United Kingdom.

Shleifer, A., and R. W. Vishny, 1997, "The Limits of Arbitrage," *The Journal of Finance*, 52(1), 35–55.

Welch, I., 1992, "Sequential Sales, Learning, and Cascades," *Journal of Finance*, 47(2), 695–732.

Yuan, K., 2000, "Asymmetric Price Movements and Borrowing Constraints," working paper, Massachusetts Institute of Technology.