Financial Intermediation
Charles M. Kahn

Markus K. Brunnermeier and Lasse Heje Pedersen,
"Market Liquidity and Funding Liquidity"
The Review of Financial Studies, Vol 22., No 6, Jun 2009, pp. 2201-2238

Since Hicks economists have been aware of a distinction between two fundamental views about liquidity: liquidity as a characteristic of an asset, and liquidity as a characteristic of the financial position of an economic agent. This paper explicitly links the two—"market liquidity" referring to the tradability of the asset in a market at a price close to its "fundamental" value, and "funding liquidity" referring to the ability of agents to obtain the financing to acquire positions in assets (short *or* long). These two notions interact and are mutually reinforcing: if participants have funding liquidity, then assets have a ready market; if assets are liquid, then their value as collateral makes it easier to fund their purchase.
The authors emphasize some basic facts about market liquidity: that it can suddenly dry up, that this drying up is correlated with market volatility—both cross sectionally and over time, and that when market liquidity dries up, we observe a "flight to quality"—a relatively high demand for the "best" assets. In the model they develop, all of these phenomena can be observed.
The model contains three types of agents:

1. Customers are risk averse. They have idiosyncratic, (non tradable) holdings, of some assets yielding a desire to trade among themselves to readjust the tradable portions of their portfolios.

2. Speculators (think of them as dealers or hedge funds) are risk neutral. They intervene in the market if customers are temporarily unavailable to hold the positions until the "right" customers finally show up. Their ability to smooth the market in this way is limited by limitations on their wealth, and the need to borrow fund to finance their trading activity; this financing entails what can be thought of as a margin requirement or collateral constraint.

3. Financiers. These agents do not directly participate in trading, but they provide the financing for the speculators to do so. Crucially, they determine the margin requirements under which speculators can receive financing.

A noteworthy feature of the model is the handling of the financial restrictions on speculators. Their wealth limitations mean that they are constrained in taking both long *and* short positions. The following inequality is key to the trading by speculators (since speculators are competitive we can treat them in aggregate)

$$\sum_j x_t^{j+} m_t^{j+} + x_t^{j-} m_t^{j-} \leq W_t \tag{1}$$

where W_t denotes speculator wealth at time t, x_t^{j+} and x_t^{j-} represent the long and short positions of a speculator in asset j at time t (So that the net position is $x_t^{j+} - x_t^{j-}$) and the corresponding m's represent the margin requirements imposed by the financiers on those trades. To understand why this is the natural restriction on speculator activity, consider first the case where a speculator is trying to establish a long positition of $x_t^{j+} > 0$ shares, which currently sell at a market price of $p_t^j = 100$. Rather than paying the full amount in cash, the speculator will prefer to purchase on margin, by borrowing and using the shares themselves as collateral. Suppose that the financiers are willing to allow him to borrow 90 per share. Then the margin or haircut is $100 - 90 = 10$. And the establishment of the position will require that the speculator use $10x_t^{j+}$ of his own capital.

The considerations are similar in the case of financing a short position. Suppose that the speculator wishes to establish a short position of $x_t^{j-} > 0$ shares. Then, simplifying slightly, he can do this by borrowing the securities from the financier in return for providing collateral (cash, say) and then selling the securities at the market price. Again suppose the market price is 100, and that the financier is willing to lend the securities to the speculator, provided that the speculator deposits 110 in collateral with the financier. Once he has the securities, he then sells them at the market price. Thus in this case $m_t^{j-1} = 110 - 100 = 10$, and the establishment of a short position has tied up $10x_t^{j-}$ of the speculator's capital.

(Clearly the speculator will not simultaneously take long and short positions in the same asset).

At this point in the paper, B&P incorporate a discussion of what parts of real world financial agents' balance sheets correspond to the notion of capital in this model. They argue that for limited liability corporations the right measure is equity reduced by assets that are not employable for the purpose of providing financial guarantees (i.e., goodwill, intangible property), while for limited liability partnerships the calculation will be more complex (the issue being "what is it that can be locked up" for security?) They argue that long term unsecured debt (lines of credit and bonds) should be included, but not short term debt.

They also note that constraint (1) oversimplifies the situation a bit, since it assumes that each position in the speculators' portfolio is margined separately and independently. In fact, we might expect that financiers would allow for a reduction in the margin requirement on portfolios which included natural hedges: if two assets were negatively correlated, then the risks would offset and the margin requirement for the portfolio could be reduced. They argue that nonetheless this simplification is a reasonable first step, and they note that the Basel regulatory standards do in some cases set up asset-by-asset requirements (although banks that use their own models for risk managing can cross-margin).

The model in greater detail There are four periods, $t = 0, 1, 2, 3$, and J assets (but we'll focus on a single asset j). The asset pays v^j at $t = 3$. There is no motivation to trade in this final period.

At each period, the asset's "fundamental value" denoted v_t^j, is the expectation of its final value:

$$v_t^j = E_t[v^j]$$

news arrives in each period to update this fundamental value, according to the following stochastic "ARCH" structure:

$$v_t^j = v_{t-1}^j + \sigma_t^j \varepsilon_t^j$$

where ε_t^j is the shock in period t, an i.i.d. normal variable with mean 0 and variance 1. The coefficient σ_t^j describes the volatility. It too is stochastic:

$$\sigma_{t+1}^j = \underline{\sigma}^j + \theta \left| \Delta v_t^j \right|$$

Where $\Delta v_t^j = v_t^j - v_{t-1}^j$, and $\underline{\sigma}^j$ and θ are parameters > 0. In other words, the volatility changes over time as well: If there was a big change in the fundamental value between period 0 and period 1 (positive or negative) a further big change is likely in the future (although positive and negative changes remain equally likely).

In a perfectly liquid market the price of the asset p_t^j would equal its fundamental value in every period. We will be examining a market which is less than perfectly liquid; as a consequence there will be a divergence between market price and fundamentals. Define

$$\Lambda_t^j \equiv p_t^j - v_t^j,$$

then $\left| \Lambda_t^j \right|$ becomes a measure of the degree of market illiquidity.

Customer demand Customers k maximize the expected utility of their terminal wealth. We assume they have CARA utility:

$$u(W_3^k) = -\exp(-\gamma W_3^k)$$

which keeps dynamics simple and works well with normal distributions.

Customers of type k have an endowment shock z^k to their terminal wealth. These shocks aggregate to zero across all customers. (Note, although the shock is to terminal wealth, customers know their own shock from the beginning). The following equation describes the dynamics of a customer's wealth:

$$W_{t+1}^k = W_t^k + (\mathbf{p}_{t+1} - \mathbf{p}_t)'(\mathbf{y}_t^k + \mathbf{z}^k)$$

where $\mathbf{y}_t^k$ represents the asset holdings at time t (and I am temporarily moving to vector notation to emphasize that this formula works perfectly well in the multiasset case).

The gimmick in the model is an assumption that customers may not show up in the market simultaneously. The specific assumption used is that with a probability $1 - a$ (in the examples a will be assumed to be very small) all the customers show up simultaneously at time 0. But with a small probability a, agents show up sequentially. Type $k = 0$ agents show up at time 0, type $k = 1$ agents show up at time 1 and type $k = 2$ agents show up at time 2. (Once an agent shows up at the market he remains there through the rest of time).

If all agents show up at the same time, then there is no role for the speculators; the agents trade among themselves to equalize their holdings (including the endowment shocks). But if the agents show up sequentially then the speculators have an economic role to play (and the possiblity of making some profit in playing that role). This is because the customors are risk averse. Suppose a customer with a positive endowment shock shows up at time 0, but other agents have not yet shown up. If he waits for the others to arrive before making any trades, he suffers the risk associated with the fundamental shocks in the meantime.

He therefore prefers offloading at least some of his position immediately to the risk-neutral speculators, who are interested in taking on the risk in return for suitable compensation.
To keep things very simple, think about the problem from a one-asset one-period perspective: The agent's goal is to pick a level of asset holding y today to maximize

$$\max_y -E(e^{-\gamma W})$$

where his wealth tomorrow W is related to his choice of asset holdings as follows:

$$W = W_0 + (v - p_0)(y + z) \tag{2}$$

where W_0 is today's wealth, and p_0 is the asset price today, and v is the (stochastic) fundamental value tomorrow. Since v is normally distributed, so is W, and so we can work with the mean and variance of W in the maximization problem:

$$\max_y -e^{-\gamma(E[W]-(\gamma/2)\text{var}(W))}$$

or equivalently

$$\max[E(W) - \frac{\gamma}{2}\text{var}(W)].$$

Now from (2) we have

$$\frac{\partial E(W)}{\partial y} = E(v) - p_0$$

and

$$\text{var}(W) = (y + z)^2\text{var}(v)$$

so that

$$\frac{\partial \text{var}(W)}{\partial y} = 2(y + z)\text{var}(v)$$

so y is defined by the first order condition

$$E(v) - p_0 - \gamma(y + z)\text{var}(v) = 0$$

or

$$y = \frac{E(v) - p_0}{\gamma\text{var}(v)} - z$$

In particular this defines exactly customers' demands for the asset in period 2 (when p_0 corresponds the the period 2 market price).
So since everybody is present in period 2, $p_2 = v_2$, in market clearing. No room for speculators; all customers instead trade so that $y = -z$, just offsetting their idiosyncratic holdings. (Notice as well, that the same thing happens in the state where everybody shows up early: $p_0 = v_0$, all customers trade $y = -z$ at time 0, and no further trade takes place, although prices fluctuate with fundamentals in subsequent periods).
Now consider the situation in period 1, if not everybody showed up early. The important fact to observe is that at time 2 there will be no uncertainty about utility going forward: when we substitute demand and the market clearing price back into the utility function it reduces to

$$-e^{-\gamma W_2^k}$$

but this means that the problem is of the same form from the point of view of period 1: for any customer present at period 1 ($k = 0, 1$)

$$y_1^k = \frac{v_1 - p_1}{\gamma \text{var}(v_2)} - z^k$$

(In this equation, v_1 is the expectation as of time 1 of the fundamental value at time 2, which in turn will equal the price at time 2). The process is more complicated to describe, but similar in principle for the first period of trading.

Role of the Speculator Note that as of time 2 there are no trades for the speculator to make, so his wealth will not vary from period 2 to period 3. However, in period 1, there will be room for him to trade, since $z^1 + z^0$ will not generally equal 0.
We need to think about the dynamics of his wealth as well:

$$W_{t+1} = W_t + (\mathbf{p}_{t+1} - \mathbf{p}_t)\mathbf{x}_t + \eta_t.$$

(Here $\mathbf{x}_t$ is the vector of net positions). B&P include the term η_t as an exogenous shock to wealth, because they are interested in examining what shocks to speculator wealth do to the market.
If amount $z^1 + z^0$ is sufficiently small, that it does not exhaust the capital constraint of the speculators, then they will compete the price down to the fundamental value: again $v_1 = p_1$ and the speculators hold the amount $z^1 + z^0$ in anticipation of the arrival in the second period of $k = 2$ type customers.
More generally, the demand by the speculators will depend on the comparison of v_1 and p_1:

$$\begin{aligned} x_t^1 &= W_t/m_t^+ \text{ if } p_t < v_t \\ &= -W_t/m_t^- \text{ if } p_t > v_t \\ &\in \left[-W_t/m_t^-, W_t/m_t^+\right] \text{ if } p_t = v_t \end{aligned}$$

Demand is "bang-bang": if $p_t < v_t$ then speculators want to hold as much of the asset as their wealth will allow, and want to short as much as possible if the inequality is reversed.
The description becomes more complicated if there are multiple assets in the market. Essentially the assets are ordered by the profitability per unit of capital used:

$$\frac{v_1^j - p_1^j}{m_1^{j+}}$$

(or the equivalent calculation on the short side). A speculator puts all of his capital into which ever asset gives the maximum value by this calculation. As a result, prices adjust, so that in equilibrium this value is the same for all assets on which speculators are taking positions. B&P rewrite the expression as follows for assets in which the speculator is taking a position

$$\left|\Lambda_1^j\right| = m_1^j(\phi_1 - 1) \tag{3}$$

(the m_1^j is chosen according to whether the position is long or short; the term in parentheses is just the shadow cost of capital to the speculators—zero if there is no shortage). The idea is this: when there is a constraint on speculators abilities to take positions, the degree to which an asset's price is distorted from the fundamental values is limited by the costs imposed on the speculators for taking the positions. (B&P do not provide a similar formula for period 0, but the idea is the same).

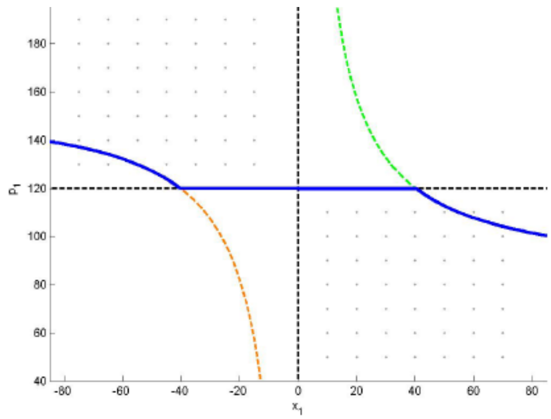

Figure 1:

Financiers' decisions Financiers essentially are assumed to set the margins based on value at risk. That is, they choose a level of protection to keep the probability that a price change will lead to a loss below some fixed threshold π :

$$\pi = \Pr(-\Delta p_2^j > m_1^{j+} | \text{information})$$

In the case of period 1 decisions, since period 2 price will equal fundamentals, as long as the finnanciers know the fundamentals (that is, as long as current values of v are in the information set) this can also be written as

$$\pi = 1 - \Phi \left(\frac{m_1^{j+} - \Lambda_1^j}{\sigma_2^j} \right)$$

The crucial part of the story is that the more prices are below fundamentals (and so the more likely prices are to rise in the future), the lower the margin required on a long position, and vice versa for a short position. In other words, margins are *stabilizing*, as illustrated in the picture below, showing in the solid line the net demand by speculators for the asset as a function of the price. The horizontal portion is the case where the price equals the fundamental value.

But B&P are looking for circumstances in which the result can be reversed. So they go on to consider situations in which financiers do *not* observe v itself.

Instead they only can base their margin requirements on their observations of prices. This is a messy problem. The price will depend on whether the state is a simultaneous arrival state or not, and if not simultaneous arrival, it will depend on the size of the idiosyncratic shocks z_0 and z_1. So B&P simplify the problem by assuming that a (the probability of the asynchronous arrival state) approaches 0. In this case, it is highly likely that $p_t^j = v_t^j$. That means that it is equally likely that prices rise or fall and so margins will be symmetric for short and long positions.

Now if prices moved a lot last period, then that means that expected fundamental volatility is large and so margins will increase. Moreover if prices swing away from fundamentals, this can also contribute to the observed volatility of prices and further increase margins—note however, that we are in a world where we are assuming this is very unlikely, so there is some

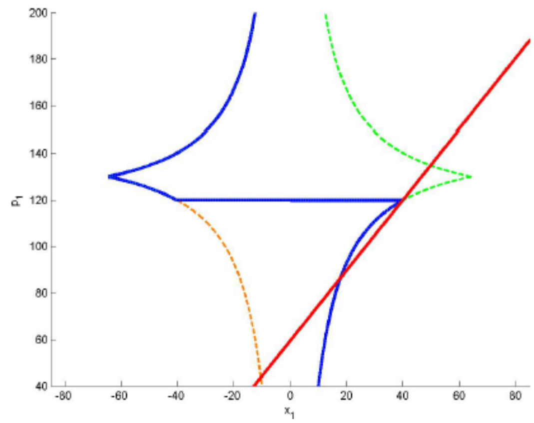

Figure 2:

tension here... Still, if an illiquidity shock and a fundamental shock go in the same direction, we have the possibility of destabilization from the margin setting.

Fragility For B&P, "financial fragility" means that equilibrium prices move discontinuously with small changes in the exogenous shocks—the shocks in question being the shock to speculator wealth and to the fundamentals of the asset value. This discontinuity will arise when there is a non-monotonic demand for assets. Customers' demands will always be monotonic; the non monotonicity will come from the speculators, and in particular from funding constraints on the speculators.
They claim that fragility can arise in the case of informed financiers (but do not illustrate the situation in detail). Instead they emphasize the case for uninformed financiers. The picture below illustrates the situation: customer supply of the asset is monotonic (indeed linear) in price. But the demand by the speculators is complicated by the margin requirements of the financiers. If prices are observed to be very high or very low, it is taken as indicative of future volatility, and so margins are tightened, both on the short and the long side. The picture is drawn with two equilibria (in general there would be three, with the middle one "unstable"; B&P focus on the other two). Small perturbations of the picture would make one of the equilibria disappear, so that prices might jump discontinuously.

Liquidity Spirals The diagram catches the essential idea of the liquidity spiral.
For this dynamic to operate, we have to be in a situation where prices are already away from fundamentals; otherwise small changes in speculator wealth will not have effects.

Commonality and Flight to Quality These results largely come from condition (3):

1. Market illiquidities of securities co-move. (As the shadow cost of capital moves, all Λ's move together.

2. Jumps in market liquidity occur simultaneously for all assets for which speculators are marginal investors

3. Lower fundamental volitility implies lower illiquidity

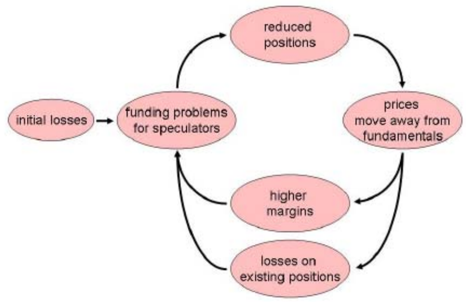

Figure 3:

4. Flight to quality. The liquidity differential between high and low fundamental volatility securities is bigger when speculator funding is tight. (Basically this is because in some regions the liquidity differential is proportional to the volatility).