Extrapolative Expectations: Implications for Volatility and Liquidity*

Prachi Deuskar†
University of Illinois at Urbana-Champaign

August 2007

Abstract

This paper presents a model of liquidity and volatility in which investors extrapolate recent price movements to forecast the volatility of a risky asset. High perceived volatility leads to high risk premium, low current return, low risk-free rate and illiquid markets. Illiquidity amplifies supply shocks, increasing realized volatility of prices, which feeds into subsequent volatility forecasts. As a result, clustering of volatility and liquidity arises endogenously. The model helps to unify several known facts about liquidity and volatility, and I find support for its new prediction which links misperception of volatility to liquidity.

*I would like to thank Viral Acharya, Yakov Amihud, Menachem Brenner, Rob Engle, Joel Hasbrouck, Yang Lu, Amrut Nashikkar, Lasse Pedersen, Allen Poteshman, Jono Spitzer, Marti Subrahmanyam, Dimitri Vayanos, Josh White, Jeff Wurgler and seminar participants at the NYU Finance Department, Board of Governors of the Federal Reserve System, Fordham University, University of Illinois Urbana-Champaign, Santa Clara University and AFA Meetings 2007. All errors are my own.

†Department of Finance, College of Business, University of Illinois at Urbana-Champaign, 304C David Kinley Hall, 1407 West Gregory Drive, Urbana, IL 61801. Ph: (217) 244-0604, Fax: (217) 244-9867, E-mail: pdeuskar@uiuc.edu.
Recent research has uncovered a number of robust empirical facts about stock market liquidity. There is a strong systematic and persistent component to liquidity across stocks, which varies sharply over time. Markets are more liquid when the current stock return is higher and when investor sentiment is higher. Behavior of stock market volatility also shows dramatic patterns of time-variation, clustering and asymmetry with respect to stock returns.

Market participants expectations regarding volatility are an important determinant of the aggressiveness of their actions, and hence of liquidity and volatility. This paper presents a unified explanation for behavior of volatility and liquidity using time-varying forecasts of volatility by investors. In a simple, integrated framework, the model explains time-varying and persistent liquidity, time-varying and persistent volatility, and the co-movement of liquidity with the risk premium, realized returns, risk-free rate, and realized volatility. Empirical analysis using perception of volatility from the options market and proxies for liquidity in the stock market supports the new prediction of the model, and links misperception of volatility to liquidity.

More specifically, in my model a group of investors extrapolates recent price movements to forecast volatility for the next period. These investors act as liquidity providers for a second group of investors who demand liquidity. When liquidity providers believe that the risky asset is less risky,
they are willing to hold more of it, thereby pushing down its risk premium and pushing up its current return. During this period they charge a lower premium to accommodate demands by the liquidity traders. This results in better liquidity. Conversely, when these investors perceive the asset to be more risky, the risk premium rises, the current return falls, and the market is more illiquid. During this period investors put more funds in the safe asset, pushing down the risk-free rate, as in a flight to safety.

An important feature of the model is that forecasts of volatility also affect realized volatility. A forecast of higher volatility results in greater illiquidity. As a result, the price of the risky assets moves more in response to liquidity demands and the realized volatility of the price is higher. This higher realized volatility feeds into the extrapolative volatility forecast for the next period. Thus, there is persistence in forecasts of volatility that leads to persistence in liquidity and clustering of realized volatility. Therefore, even though fundamentals (dividends) in the model are homoskedastic, prices are heteroskedastic. In the real world, it is hard to find evidence for or against heteroskedasticity in the fundamentals, such as cash flows or earnings, as these variables are observed at a low frequency. However, there is clearly strong evidence of time-varying and persistent volatility in returns. The model presented here illustrates how even without fundamental heteroskedasticity, return heteroskedasticity can arise.

As Fama (1998) points out, it is not enough for a new theory to explain existing facts. It must also generate new, rejectable hypotheses. The model in this paper, in addition to explaining a number of known stylized patterns about the behavior of liquidity and volatility with a single behavioral assumption, this model makes a new prediction. It predicts that when misperceived volatility is
high, liquidity is low. To test this prediction I construct measures of liquidity and misperception of volatility. As a proxy for misperception of volatility, I use the difference between the volatility index (VIX), created by the Chicago Board of Options Exchange (CBOE) and based on the implied volatility of S&P 100 index options, and the expected realized volatility of the index returns. I measure market-wide liquidity using daily data on returns and volume. As predicted, I find that when misperception of volatility is high, liquidity is low, even after controlling for actual volatility. The results are robust when using different models for extracting implied volatility and calculating the expected realized volatility. I find similar support for the models predictions when I measure volatility misperceptions and stock illiquidity at the individual stock level using trade and quote data. The illiquidity of individual stocks also moves positively with market-wide misperception of volatility, and more so for more volatile stocks. It appears that when the investors perceive the markets to be more volatile, they in particular withdraw liquidity from more volatile stocks.

The key assumption of the model is that the investors make estimations of future volatility based on a short history of prices. This behavior is consistent with research in psychology that has established that individuals use certain heuristics when forming judgments under uncertainty. [See Tversky and Kahneman (1982).] People tend to generalize from a few experiences (representativeness heuristic). They suffer from local representativeness; i.e., they see patterns in truly random data. They are also affected by the availability heuristic; i.e., they associate ease of recall with frequency of occurrence. In most cases, more recent events are more easily recalled. Thus, taken together, these three heuristics generally point toward extrapolation based on recent instances when forming judgments. Most importantly for this paper, the perception of risk is also affected by these heuristics; Slovic,
Fischhoff, and Lichtenstein (1982) find that judgment about risk is disproportionately influenced by recent experiences.

In the financial markets as well, anecdotal evidence suggests that investors' judgments about risk seem to be highly affected by recent experiences. Salient events such as the crash of 1987 and the LTCM crisis of 1998 are believed to have compelled investors to be temporarily overcautious. As is well known, implied volatilities of index options reached record levels immediately after the October 1987 crash.

More rigorous findings from the options market support this anecdotal and experimental evidence. Stein (1989) finds that the implied volatility of medium term options on the S&P 100 overreacts to the changes in the implied volatility of short term options. That is, while forming expectations of the volatility over the medium term, the market places too much emphasis on recent shocks to volatility, and does not pay enough attention to the mean-reverting nature of volatility.

Poteshman (2001) documents that option-implied volatilities underreact to changes in instantaneous volatility that are preceded by changes of the opposite sign. On the other hand, it is shown that a string of similar changes in instantaneous volatility causes an overreaction in the option-implied volatilities. Thus the history of volatility influences forecasts of volatility.

These studies suggest that, to some extent, investors extrapolate on the basis of a short time period when forecasting volatility. Even standard practitioner guidelines for risk management advocate volatility forecasts that effectively use a limited history of price movements. Widely-used models of volatility forecasting such as ARCH (suggested in Engle (1982)), and its various generalizations,
also put a higher weight on the most recent observation, giving them a flavor of extrapolation. In summary, a behavioral assumption that investors base forecasts of volatility on extrapolation is supported by anecdotes, empirical finance research, and industry practice.

The assumption that naive investors believe the volatility to be serially correlated is used in Frankel (2005) to explain crashes. Though the core assumption in that study is similar to the one in this paper, the focus there is on explaining why crashes are more likely to occur than booms. Danielsson, Shin, and Zigrand (2004) investigate the asset-price dynamics when the investors use a backward-looking volatility estimate as an input for risk management. My paper uses this assumption to explain time-variation in liquidity and realized volatility.

Spiegel (1998) uses a model with overlapping generations of short-lived investors to show that there can be an equilibrium with high volatility of asset prices even though dividends are constant. However, there is no implication for time-varying and persistent volatility of prices in that paper. This paper, using the assumption of extrapolative expectations, shows that even when dividends are homoskedastic, volatility of prices can be time-varying and persistent.

This paper contributes to a rapidly growing literature that examines properties of liquidity and its role in asset pricing. The empirical side of this literature has documented that aggregate liquidity is time-varying [see, for example, Chordia, Roll, and Subrahmanyam (2000), Hasbrouck and Seppi (2001), and Huberman and Halka (2001)] and that it matters for asset pricing [see, for example, Amihud (2002), Pastor and Stambaugh (2003), and Acharya and Pedersen (2005)]. The theoretical side of this literature has derived implications of substantial magnitude for asset pricing given
Some recent working papers focus on related issues. Brunnermeier and Pedersen (2005) build a model that links market liquidity to the funding capital of the market-makers. Carlin, Lobo, and Viswanathan (2006) explain liquidity crises using a model of cooperative and predatory trading among market participants. Both papers attempt to explain variation in liquidity with special attention to episodic liquidity dry-ups. My paper complements their work by shedding light on a different channel for variation in liquidity—variation during both normal periods and “crises,” and persistence in liquidity.

Another closely related paper, Vayanos (2004), examines flight to quality and pricing of risk in an economy with multiple assets. It provides an explanation for various observed phenomena, such as flight to quality during volatile times, and time-variation in risk premiums, illiquidity premiums and correlations across assets. In that paper, the illiquidity or transaction costs of the risky assets are exogenous and constant over time. Also, the stochastic volatility of the risky assets is exogenous. In contrast, this paper generates time-varying illiquidity as well time-varying realized volatility endogenously.

Baker and Stein (2004) present a model of aggregate liquidity and investor sentiment in which sentiment arises out of beliefs about the mean value of the risky asset. This paper provides a different channel through which investor beliefs influence liquidity; in my model, the investor sentiment
reflects beliefs about the volatility of the risky asset. This allows me to integrate the volatility, liquidity and sentiment stories. Another advantage of my model is that volatility sentiment feeds on itself, giving rise to persistence in volatility, consistent with observation.

The rest of the paper is organized as follows. Section I presents the theoretical model. Empirical results are in section II. The last section concludes. Proofs are given in the appendix.

I A Model of Extrapolative Volatility Expectations

This section presents a simple model that simultaneously captures several empirical facts about liquidity and volatility. The basic mechanism of the model is similar to Grossman and Miller (1988), where risk-averse liquidity providers accommodate liquidity demands in exchange for some compensation.

A Basic Model

In this economy, there is one risky asset that pays a perpetuity of dividends $d_t$ which are distributed independently and identically each period with mean $\bar{d}$ and variance $\sigma_d^2$. The supply of the risky asset is normalized to 1. There is a riskless asset with perfectly elastic supply. It pays $1 + r$ at $t + 1$ for every dollar invested at $t$. The riskless rate is exogenous, for now. It is endogenized in the next subsection.

These two assets are held by overlapping generations of short-lived investors. The investors born
at time $t$ invest in the safe and the risky asset at $t$. They sell their holdings to the next generation and consume at $t+1$. The investors are mean-variance optimizers. Specifically, investors born at time $t$ solve

$$\max_{y_t} \quad E_t[W_{t+1}] - \frac{\gamma}{2} Var_t[W_{t+1}]$$

$$s.t. \quad W_{t+1} = (W^i - y_t P_t)(1 + r) + y_t (P_{t+1} + d_{t+1})$$

where $W^i$ is the initial wealth of each generation, $y_t$ is the quantity of the risky asset demanded at time $t$ and $P_t$ is the price of the risky asset at time $t$. The total number of these optimizing investors is normalized to 1.

The superscript $i$ for expectation and variance indicates expectations taken using beliefs of the investors. At time $t$ the investors believe that the variance of $P_{t+1}$ is $\theta_t \sigma^2_P$, where $\sigma^2_P$ is the unconditional variance of the price. $\theta_t$ can take one of two values, $\theta_H$ or $\theta_L$, where $\theta_H \geq \theta_L$. The investors believe that $\theta_{t+1}$ and $\theta$ in all the subsequent periods is going to be $\theta_H$ or $\theta_L$ with equal probability, which is the unconditional probability of these two states.

In addition to the investors described above, there are liquidity traders. Each period a new generation of liquidity traders is born with initial wealth $W^l$. They place a demand for $z_t$ units of the risky assets, which is distributed identically every period and is independent of all other variables. $z_t$ has mean 0 and variance $\sigma^2_z$. The liquidity traders invest the balance of their wealth $(W^l - z_t P_t)$ in the risk-free asset.
Evolution of $\theta_t$

The investors are backward-looking when it comes to forming forecasts about volatility. They extrapolate from the recent history to estimate volatility in the near future. Specifically, at $t$ they observe $P_{t-1}$ and form beliefs about $\theta_t$. If at $t-1$ the deviation of the price from its mean is large, the investors put a high value on $\theta_t$. Thus, if the price in the past was very volatile, they believe that the price in the next period is going to be very volatile. In particular,

$$\text{If } |P_{t-1} - E[P_{t-1}|P_{t-2}, P_{t-3}, \cdots]| > M, \text{ then } \theta_t = \theta_H$$

and

$$\text{if } |P_{t-1} - E[P_{t-1}|P_{t-2}, P_{t-3}, \cdots]| \leq M, \text{ then } \theta_t = \theta_L$$

where $M > 0$ is the critical cut-off.

Equilibrium

In equilibrium, the market for the risky asset clears. The quantity of the risky asset demanded by the liquidity providers is given by

$$y_t = \frac{\bar{d} + E_t[P_{t+1}] - \gamma(\theta_t \sigma^2 + \sigma_d^2)(1 + r)P_t}{\gamma(\theta_t \sigma^2 + \sigma_d^2)}$$  \hspace{1cm} (3)

By the market clearing condition for the risky asset, $y_t + z_t$ equals 1. Solving for the price of the risky asset, we get

$$P_t = \frac{\bar{d} + E_t[P_{t+1}] - \gamma(\theta_t \sigma^2 + \sigma_d^2) + \gamma(\theta_t \sigma^2 + \sigma_d^2)z_t}{1 + r}$$  \hspace{1cm} (4)
Define \( \bar{\theta} \) as the unconditional mean of \( \theta_t \); i.e., \( 0.5\theta_H + 0.5\theta_L \). Iterating forward gives the price of the risky asset as

\[
P_t = \frac{\bar{d}}{r} - \left( \frac{\gamma(\bar{\theta}\sigma_P^2 + \sigma_d^2)}{r} + \frac{\gamma(\theta_t - \bar{\theta})\sigma_P^2}{1 + r} \right) + \gamma(\theta_t\sigma_P^2 + \sigma_d^2) \frac{1}{1 + r} z_t
\]  

(5)

\( \sigma_P^2 \) is the unconditional variance of the price obtained in a steady-state equilibrium. The expression for \( \sigma_P^2 \) is provided in the appendix.

The price of the risky asset can be decomposed into the risk-neutral valuation, the risk premium and the total price impact of trade by the liquidity traders,

\[
P_t = P^{RN} - RP_t + \lambda_t z_t
\]  

(6)

where

\[
P^{RN} = \text{Risk-neutral valuation} = \frac{\bar{d}}{r}
\]

\[
RP_t = \text{Risk premium at time } t = \frac{\gamma(\bar{\theta}\sigma_P^2 + \sigma_d^2)}{r} + \frac{\gamma(\theta_t - \bar{\theta})\sigma_P^2}{1 + r}
\]

\[
\lambda_t = \text{Price impact per share (illiquidity) at time } t = \frac{\gamma(\theta_t\sigma_P^2 + \sigma_d^2)}{1 + r}
\]

\[
\lambda_t z_t = \text{Total price impact of trade}
\]

If investors are risk-neutral, i.e., if \( \gamma \) is 0, the price collapses to the risk-neutral valuation. If the liquidity demand \( z_t \) is 0, the price is just the risk-neutral valuation less the conditional risk premium. This is determined by the optimal portfolio choice of the risk-averse agents. However, the liquidity traders place a demand for the risky asset for exogenous reasons. Thus, the optimizing agents must hold a quantity of risky asset that is different from what they would hold otherwise. They do this only if the price is adjusted in the appropriate direction, and thus the price of the
risky asset moves in response to \( z_t \), the liquidity demand. Using market microstructure language, \( P^{RN} - RP_t \) can be thought of as the mid-price and \( \lambda_t \) can be thought of as the half-spread. The risk-averse investors jointly act as market-makers as in Grossman and Miller (1988). They are being compensated to accommodate the demand of the liquidity traders. In particular, they are willing to hold an additional unit of the risky asset at a bid of \( P^{RN} - RP_t - \lambda_t \) and are willing to sell one unit of the risky asset at an offer of \( P^{RN} - RP_t + \lambda_t \). Thus \( \lambda_t \) is an indicator of illiquidity, and represents the response of price to the order-flow and the cost of trading for the liquidity traders.\(^1\) \( \lambda_t \) can also be thought as the slope of the supply curve of the liquidity providing agents. A higher \( \lambda_t \) translates into a steeper supply curve and thus the price moves more in response to the liquidity demand, \( z_t \).

**Properties of Liquidity and Volatility**

Since the per unit price impact, \( \lambda_t \), is the compensation for the risk-averse investors, it has some intuitive properties, including the effects of volatility of fundamentals and the risk-aversion of the investors.

**Proposition 1** *Price impact depends on the primitive parameters as follows:*

(i) *The higher the fundamental risk, the lower is liquidity.* \( \frac{d\lambda_t}{d\sigma_f^2} > 0 \).

(ii) *The higher the risk-aversion, the lower is liquidity.* \( \frac{d\lambda_t}{d\gamma} > 0 \).

(iii) *The higher the risk of liquidity demand, the lower is liquidity.* \( \frac{d\lambda_t}{d\sigma_z^2} > 0 \).
In this model, the volatility of the price of the risky asset is time-varying. I define conditional volatility of the price of the risky asset as

$$ Vol_t = (\text{Var}(P_t|P_{t-1}, P_{t-2}, \ldots))^{1/2} = \frac{\gamma(\theta_t \sigma^2 + \sigma_z^2)}{1 + r} \sigma_z $$

This is the volatility of the price conditional on past prices, similar to the ARCH / GARCH models. Note that the volatility of the price at time $t$ is determined by $\theta_t$. As $\theta_t$ changes over time, the \textit{realized} volatility of prices, $Vol_t$, also changes. If at time $t$ the investors believe that volatility of the future price is going to be high, the price impact of trade, $\lambda_t$, is high. High $\lambda_t$ amplifies the effect of volatility on $z_t$, to determine the volatility of the current price. Thus, beliefs about high (low) future volatility, $\theta_H$ ($\theta_L$), result into high (low) current \textit{realized} volatility.

The model also captures the empirically observed relationships between liquidity, on one hand, and expected returns, realized returns, and realized volatility on the other. The following proposition summarizes these relationships.

\textbf{Proposition 2} Price impact covaries with other variables of interest as follows:

(i) Illiquidity and the risk premium are positively correlated. $\text{Cov}(\lambda_t, RP_t) \geq 0$.

(ii) Illiquidity and contemporaneous volatility are positively correlated. $\text{Cov}(\lambda_t, Vol_t) \geq 0$.

(iii) Illiquidity and contemporaneous price are negatively correlated. $\text{Cov}(\lambda_t, P_t) \leq 0$.

Amihud (2002) finds that illiquidity and the risk premium (expected return) of the market are positively correlated. Proposition 2(i) captures this effect. Chordia, Roll, and Subrahmanyam
(2000) provide evidence of low liquidity in volatile and down markets. Proposition 2(ii) and (iii) support these results.

A new prediction of the model that I test in the empirical section, is the relationship between volatility misperception and liquidity. This is presented in the following proposition.

**Proposition 3** Illiquidity and misperception of volatility are positively correlated.

$$\text{Cov}(\lambda_t, \theta_t) \geq 0.$$  

When the investors believe that volatility is going to be high in the future, they demand higher compensation per unit to provide liquidity. Thus, when misperception of volatility, $\theta$ is high, price impact $\lambda$ is also high.

Another important aspect of the model is the evolution of liquidity and realized volatility of prices over time. As stated in (2), if the price is away from its conditional mean by more than a critical amount $M$, the investors expect volatility to be higher next period. Using the equilibrium expression for price, this rule translates into following conditions.

$$\text{If } |\frac{\gamma(\theta_t \sigma_P^2 + \sigma_d^2)}{1 + r} z_t| > M, \text{ then } \theta_{t+1} = \theta_H$$

and if $$|\frac{\gamma(\theta_t \sigma_P^2 + \sigma_d^2)}{1 + r} z_t| \leq M, \text{ then } \theta_{t+1} = \theta_L$$  

(8)

As discussed earlier, $\theta$ determines the current realized volatility of the price. Realized volatility at time $t$ then feeds into the beliefs at time $t + 1$. This feedback effect is illustrated in Figure 1. If
at $t$ the perceived volatility is higher ($\theta_t = \theta_H$), the risk-averse investors charge more per unit to accommodate the liquidity traders. Thus, illiquidity ($\lambda_t$) is higher. Higher illiquidity amplifies the volatility of the liquidity demand to generate more volatile price of the risky asset at $t$. Higher realized volatility at $t$ increases the likelihood of higher perceived volatility at $t + 1$. Similarly, if perceived volatility at $t$ is lower, it is more likely to have lower perceived volatility at $t + 1$. Thus a belief in higher or lower volatility propagates itself. Persistence in beliefs translates into persistence in liquidity and realized volatility as stated in the following proposition.

**Proposition 4**

(i) Illiquidity is persistent. $\text{Cov}(\lambda_t, \lambda_{t+1}) \geq 0$.

(ii) Realized volatility of the risky asset price is persistent. $\text{Cov}(\text{Vol}_t, \text{Vol}_{t+1}) \geq 0$.

Persistence in liquidity is noted in several empirical investigations. [For example, see Amihud (2002) and Pastor and Stambaugh (2003)]. Persistence in volatility of returns has been established using various extensions of the ARCH model in Engle (1982). There is strong evidence of volatility clustering in returns at almost all frequencies. However, there is no evidence for or against volatility clustering in the fundamentals, such as cash flows or dividends. Hence, it is not clear what the source of heteroskedasticity in returns is. In the model in this paper, the dividend process is homoskedastic. However, the volatility of the prices turns out to be time-varying and persistent. Thus the model provides a channel for endogenous volatility clustering; even when the fundamentals are heteroskedastic, this channel will amplify the effect to provide a more pronounced pattern in the volatility of prices.
It is important to emphasize that it is not just the volatility estimates of irrational investors that show persistence. The realized volatility shows clustering as well. As discussed above, this happens because of the actions of irrational traders. A higher estimate of volatility results in a greater price impact of trade, which in turn amplifies the effect of volatility of liquidity demand to generate more volatile prices. These volatile prices feed into the forecast for next period, creating persistence in liquidity as well as realized volatility. Thus, to a certain extent, the investors bring their beliefs into reality through their own actions.

Given that the investors are risk-averse, a time-varying coefficient of risk-aversion $\gamma$ would also imply a time-varying price impact $\lambda$ and time-varying realized volatility of prices. However, this channel could only generate persistence in volatility and liquidity if the risk-aversion changes in a particular manner. More specifically, we would need higher risk-aversion in periods following higher realized volatility. Also, it is more conceivable that risk-aversion changes at low frequencies, for example with business cycles, than that it changes at high frequencies. However, time-variation and persistence in liquidity and realized volatility are observed empirically at high frequencies as well as low frequencies. So although changing risk-aversion is almost certainly a part of the story, it cannot be the entire story.

B Endogenous Risk-Free Rate and Flight to Safety

In this subsection, the set up is exactly the same as before except that the supply of the riskless asset is fixed at $S$. Thus the demand for the riskless asset at time $t$ determines the riskless rate $r_t$ between $t$ and $t + 1$. 

15
As before, the optimizing investors have total wealth $W^i$. They buy $y_t$ units of the risky asset and put the balance of their wealth in the riskless asset. The investors solve the optimization problem as in subsection A and demand $y_t$ units of the risky asset.

$$y_t = \frac{d + E_t[\Delta_{t+1}] - (1 + r_t)P_t}{\gamma(Va_1^i[\Delta_{t+1}] + \sigma_d^2)}$$ (9)

The superscript $i$ indicates expectation and variance based on the beliefs of the irrational investors.

It is useful to recall that at time $t$ the investors think that the variance of price of the risky asset is going to be $\theta_t\sigma_P^2$. $\sigma_P^2$ is the unconditional variance of the price of the risky asset. $^2\theta_t$ can be high ($\theta_H$) or low ($\theta_L$). The evolution of $\theta_t$ happens as described in the earlier subsection, i.e., if price in the previous period is too volatile, $\theta_t$ takes the value $\theta_H$, otherwise investors think it is $\theta_L$. Unconditionally, both values of $\theta_t$ are equally likely.

As before, liquidity traders have wealth $W^l$. They demand $z_t$ units of the risky asset and put the balance of their wealth in the riskless asset.

**Equilibrium**

Market clearing gives the price of the risky asset as

$$P_t = \frac{\bar{d}}{1 + r_t} + \frac{E_t[\Delta_{t+1}] - \gamma(\theta_t\sigma_P^2 + \sigma_d^2)}{1 + r_t} + \frac{\gamma(\theta_t\sigma_P^2 + \sigma_d^2)}{(1 + r_t)} z_t$$ (10)

Define total initial wealth in the economy, $W$, as $W^i + W^l$, the sum of the wealth of the two groups of agents. The budget constraints of the agent coupled with the market clearing for the safe asset
give

\[ S = (W - P_t)(1 + r_t) \]  (11)

Substituting for \( P_t \) from (10) and rearranging, we get the price of the safe asset as

\[ P^s_t \equiv \frac{1}{1 + r_t} = \frac{W}{S + \bar{d} + E[P^s_{t+1}] - \gamma(\theta_t \sigma^2_P + \sigma^2_d) + \gamma(\theta_t \sigma^2_P + \sigma^2_d)z_t} \]  (12)

Iterating forward the expression in (10), we get the price of the risky asset as

\[ P_t = \left( \frac{\bar{d}}{1 - \bar{P}^s} - \frac{\bar{R}P - \gamma \theta_t \sigma^2_P + \gamma(\theta_t \sigma^2_P + \sigma^2_d)z_t}{\gamma(\theta_t \sigma^2_P + \sigma^2_d)} \right) P^s_t \]  (13)

\( \bar{P}^s \) is the unconditional expectation of the price of the riskless asset. \( \bar{R}P \) is the constant risk premium, the expression for which is provided in the appendix. We need one additional restriction in this set-up. From expression (12), it is clear that if \( z_t \) is too low, the gross interest rate \( 1 + r_t \) would be negative. To avoid such situations, we put a lower bound on \( z_t \).

\[ z_t \geq 1 - \frac{(S + \bar{d}/(1 - \bar{P}^s) - \bar{R}P + \gamma \sigma^2_d)}{\gamma(\theta_t \sigma^2_P + \sigma^2_d)} \]  (14)

**Flight to Safety**

When the risk-free rate is determined in equilibrium, the price of the risky asset is no longer linear in liquidity demand \( z_t \). The price impact of trade, \( \lambda_t \), is shown in the appendix to be

\[ \lambda_t \equiv \frac{dP_t}{dz_t} = \frac{S}{W} \gamma(\theta_t \sigma^2_P + \sigma^2_d)(P^s_t)^2 \]  (15)

There are now two channels through which \( z_t \) affects the price. The first is, as before, the direct price impact as investors charge a premium to accommodate the liquidity demand. Higher \( z_t \) pushes
up the price because a lower expected return is needed to make the investors hold less of the risky asset to accommodate greater demand by the liquidity traders. The second and new channel is through $P_t^a$. Higher $z_t$ pushes up the interest rate, resulting in more discounting and thus lower $P_t$. The result is $\lambda_t$ decreasing in $z_t$ (rather than a constant with reference to $z_t$ as in subsection A).

Even when the risk-free rate is determined endogenously, the positive relationship between liquidity and misperception is retained. Other properties of liquidity as described earlier also hold. A new implication of this extension is that the relationship between liquidity and the risk-free rate illustrates “flight to safety behavior. The following proposition formally states this relationship.

**Proposition 5** In periods of greater price impact, i.e., low liquidity, the riskless interest rate is lower.

$$\text{Cov}(\lambda_t, r_t) \leq 0$$

The flight to safety behavior takes place because of volatility misperception $\theta_t$. Higher $\theta_t$ means investors are more nervous about the volatility of the asset and are unwilling to provide liquidity without higher compensation. This lowers the liquidity of the risky asset. This is also the period when these risk-averse investors are willing to hold less of the risky asset and hence put more money in the safe asset. This pushes down the interest rate. As a result, during periods of low liquidity in the risky asset, investors flock to the risk-free asset—essentially the flight to safety pattern.

In a recent working paper, Caballero and Krishnamurthy (2007) model flight to safety behavior as a result of uncertainty aversion. However, they focus on what happens when the perceived uncertainty by the investors is high, and not on explaining what might lead to this perception.
Vayanos (2004) also provides an explanation for flight to safety as a result of exogenously changing volatility. This paper models how a flight to safety can result from high perceived volatility.

II Empirical Results

The theoretical model presented in this paper ties together a number of known empirical facts about liquidity and volatility. Hence, the brief empirical analysis here focuses only on the new implications. In particular, a central and untested prediction of the model is the relationship between the price impact of trade and the misperception of volatility, as stated in proposition 3. In this section, I test this prediction by constructing proxies for misperception of volatility and price impact, and then examining the relationship between the two.

A Measuring Volatility Misperception

One convenient way to assess the market perception of volatility is to use the implied volatility of options. The CBOE has created a well-known volatility index, VIX, which measures the market expectation of near term volatility conveyed by stock index option prices. VIX is constructed using the implied volatilities of eight different OEX option series (options on the S&P 100 index) so that, at any given time, it represents the implied volatility of a hypothetical at-the-money OEX option with exactly 30 days until expiration. I obtain a monthly time series of VIX from the CBOE website for a period from January 1986 to December 2004.

Thus, at any given point VIX is the market’s estimate of the volatility of S&P 100 returns over
the next month. I calculate the standard deviation of S&P 100 daily returns for each month. I annualize the standard deviation to be comparable with VIX, and call this variable RV (realized volatility). I then measure the misperception of market-wide volatility as

\[ Mis_{m,t} = VIX_t - E_t[RV_{t+1}] \] (16)

where \[ RV_{t+1} = \alpha_0 + \alpha_1 RV_t + \alpha_2 RV_{t-1} + \alpha_3 VIX_t + \epsilon_{t+1} \]

Thus the rational expectation of \( RV_{t+1} \) is modeled as a function of two of its own lags and the level of VIX.\(^5\) In this specification, \( E_t[RV_{t+1}] \) can be thought of as the econometrician’s estimate of volatility, whereas \( VIX_t \) is the market’s estimate of volatility. Thus \( Mis_t \) is the misperception of volatility by the market, and is based on time \( t \) information. It is not the forecast error relative to subsequently realized volatility, which can only be obtained at time \( t + 1 \). There has been a lot of research into the information content of the implied volatility of options. For example, see Canina and Figlewski (1993) and Christensen and Prabhala (1998), among many others. In particular, Christensen and Prabhala (1998) find that the implied volatility of S&P 100 index options is very informative in predicting the subsequent realized volatility. That is why I model \( E_t[RV_{t+1}] \) as a function of \( VIX_t \). Thus the econometrician is not overlooking the information in the implied volatility.

Table I presents the descriptive statistics for \( Mis_{m,t} \) for the period from February 1986 to December 2004. On average, the market overestimates the volatility of S&P 100 index return by 5.65 percentage points. This misperception moves around quite a bit, with a monthly standard deviation of 2.21 percentage points.
B Measuring Price Impact

To test the predictions of the model, a measure of liquidity (or illiquidity) is needed. I use $Illiq$ suggested in Amihud (2002). For each stock $i$ for each month, $Illiq$ is calculated as follows.

$$Illiq_{i,t} = \frac{1}{D} \sum_{d=1}^{D} \frac{|R_{i,d,t}|}{Volume_{i,d,t}}$$  \hspace{1cm} (17)

$R_{i,d,t}$ is the return on stock $i$ on day $d$ during month $t$. $Volume_{i,d,t}$ is the dollar volume for the same stock on the same day in thousands of dollars. $D$ is the number of days in month $t$ for which the trading volume is nonzero. $Illiq$ captures the average movement in the return per dollar volume traded. Hasbrouck (2006) finds that among various price impact measures, $Illiq$ has the highest correlation with the measures of price impact constructed from high-frequency data. An aggregate measure of illiquidity is formed as follows.

$$Illiq_t = \left( \frac{1}{N} \sum_{i=1}^{N} Illiq_{i,t} \right) \frac{MktCap_{t-1}}{MktCap_0}$$ \hspace{1cm} (18)

Thus, the aggregate price impact is the equal-weighted average of the price impacts of individual stocks. $N$ is the number of stocks in month $t$. Equal-weighting avoids overrepresentation of large stocks and follows most studies of aggregate liquidity [for example, see Amihud (2002), Pastor and Stambaugh (2003), Acharya and Pedersen (2005)]. The average is calculated separately for NYSE / Amex and Nasdaq.

A volume of one thousand dollars in 1985 is not comparable to a volume of one thousand dollars in 2004. Scaling by market capitalization produces a price impact measure for volume that aligns with the size of the stock market. $MktCap_{t-1}$ is the market capitalization at end of the month $t-1$.
of all the stocks eligible to be included for the average in month \( t \). \( \text{MktCap}_0 \) is the corresponding figure for December 1985. Thus \( \text{Illiq}_t \) is the cost of trading a volume of one thousand 1985 stock market dollars. A similar adjustment to the aggregate liquidity measure is used in Pastor and Stambaugh (2003) and Acharya and Pedersen (2005).

Table I gives the mean and standard deviation of price impact coefficients for NYSE / Amex and Nasdaq stocks. The cost of trading on Nasdaq is 37 basis points per thousand dollars (1985 stock market dollars) versus 7 basis points for NYSE / Amex. The standard deviation of price impact for Nasdaq is also much higher than that of NYSE / Amex.

C Relationship between Volatility Misperception and Liquidity

The first elements to examine are the raw correlations between misperception and illiquidity. The correlation between \( \text{Mis}_{m,t-1} \) and \( \text{Illiq}_t \) is 0.40 for NYSE / Amex and 0.16 for the Nasdaq, both statistically significant. The positive and significant correlation is in line with Proposition 3; when the investors perceive volatility to be low, they provide more liquidity.

Figure 2 gives a pictorial representation of the relationship between illiquidity and volatility misperception. As is clearly shown, aggregate \( \text{Illiq} \) is positively correlated with \( \text{Mis}_m \).
Given that both these variables are very persistent, it is appropriate to control for autocorrelation in these variables as well as for other important effects. I explore the relationship between liquidity and $M_i$ in more depth using multivariate regressions. Chordia, Roll, and Subrahmanyam (2001) document a significant relationship between liquidity and volatility. Volatility is also known to explain cross-sectional variation in liquidity. Comparative statics in Proposition 1 and results in Proposition 2 also point toward the dependence of liquidity on volatility; thus it is appropriate to include a volatility variable. I use a technique from Campbell, Lettau, Malkiel, and Xu (2001) to construct measures of aggregate idiosyncratic volatility $IVol_t$ and aggregate systematic volatility $SVol_t$. For every month $t$,

$$(IVol_t)^2 = \frac{1}{\sum_{i=1}^{N} MktCap_{i,t-1}} \sum_{i=1}^{N} MktCap_{i,t-1} Var_t (R_{i,d,t} - R_{m,d,t})$$ (19)

where

$MktCap_{i,t-1} = \text{Market capitalization of stock } i \text{ at the end of month } t - 1.$

$R_{i,d,t} = \text{Return on stock } i \text{ for day } d \text{ in month } t$

$R_{m,d,t} = \text{Market return (CRSP value-weighted return) for day } d \text{ in month } t$

$Var_t = \text{Variance of the daily observations over month } t$

Thus $(IVol_t)^2$ is a value-weighted average of the variance of the daily excess return over the market for each stock. $SVol_t$ is the volatility of the CRSP value weighted market return.

$$(SVol_t)^2 = Var_t (R_{m,d,t})$$ (20)

Idiosyncratic variance captures the potential asymmetric information for a stock. A market-maker worries about information that an insider has because it affects her profits negatively. As demon-
strated in Kyle (1985), the larger the potential for inside information, the greater the price impact. Thus the relation between price impact and $IVol_t$ is expected to be positive. Risk-averse market-makers would be worried about the systematic risk of the stock they are asked to take positions in. Thus with higher systematic risk, liquidity should be lower. So the relation between price impact and $SVol_t$ is also expected to be positive.

Another purpose of including the volatility variable is to control for the possibility that the model for $E_t[RV_{t+1}]$ does not capture the true expectation of future volatility and is biased in some way. Controlling for actual volatility ensures that any results regarding the relationship between liquidity and volatility misperception are not affected by this bias. In addition, there may be other factors which contribute to variation in return volatility, and hence liquidity. Including realized volatility controls for these other channels affecting liquidity. Finally, Chordia, Roll, and Subrahmanyam (2001) find that liquidity is lower in down markets. To control for this effect, I include the contemporaneous market return in the regression. Again, this is necessary if there are other channels through which the relationship between liquidity and market return comes about.

Thus the regression specification is as follows:

$$Illiq_t = \beta_0 + \beta_1 IVol_t + \beta_2 SVol_t + \beta_3 Rm,t + \beta_4 Mis_{m,t-1} + \omega_t$$

(21)

Table II presents the results. As can be seen from Panels A and B, the coefficient of misperception of risk is positive and significant at 5% level. When investors perceive the risk to be lower, i.e., when misperception is low, price impact of trade is lower.
The variable $Mis_t$ captures the difference between implied Volatility, which reflects the market’s expectation of realized Volatility, and a model-based forecast of realized volatility. However, the difference between implied volatility and realized volatility also results from the volatility risk premium or the disparity between volatility under objective and risk-neutral probability measures (for example, see Chernov (2002) among many others). However, to the extent this volatility risk premium is constant or dependent on the level of volatility, it does not affect the results here since the regressors include a constant and level of volatility. As shown by Chernov (2002), in a model of stock price distribution that incorporates affine jump diffusion, volatility risk premium turns out to be a linear function of latent spot volatility. The paper finds that in the context of S&P 100 index, controlling for the latent spot volatility, as proxied by the difference between daily high and low prices, corrects the bias in forecasting volatility. I control for this measure of latent spot volatility and find that the results here are robust to such specification. Thus, even though the misperception variable could be partly capturing volatility risk premium, its relationship with illiquidity is not driven by the risk premium.

To get a sense of the economic significance of these results, one can convert the coefficient to standardized units. In such a set-up, the coefficient on $Mis_{m,t-1}$ would be 0.04 in the case of NYSE / Amex, and 0.06 in the case of Nasdaq. Thus, a one-standard deviation move in misperception is associated with a change in illiquidity of about 5% of one standard deviation.
Robustness

I also construct other measures of price impact to test the robustness of these results. Consider the following regression.

\[ R_{e,i,d,t} = \alpha_i + \lambda_{i,t} \text{sign}(R_{e,i,d,t}) \text{Volume}_{i,d,t} + \gamma_{i,t} \text{sign}(R_{e,i,d-1,t}) \text{Volume}_{i,d-1,t} + \delta_{i,t} R_{i,d-1,t} + \nu_{i,d,t} \]  

(22)

where \( R_{e,i,d,t} \) is the return on stock \( i \) on day \( d \) in month \( t \) in excess of the market return, \( \text{Volume}_{i,d,t} \) is the dollar volume for stock \( i \) on day \( d \) in month \( t \), and \( R_{i,d,t} \) is the raw return on stock \( i \) on day \( d \) in month \( t \).

A buy order will result in a price moving upward whereas a sell order will result in price moving downward. In the regression above, dollar volume signed by the excess return is a proxy for order-flow. If the return on a stock is higher than the market return, the volume probably consists of buy orders. Positive information is inferred from the buy orders and the excess return is positive. Similarly, a negative excess return indicates that volume is primarily made up of sell orders. The coefficient of interest here is \( \lambda_{i,t} \). This is the price impact coefficient. It measures the extent to which the order-flow moves the price. Signed lagged volume is included in the regression to control for volume-based return continuation. Lagged return is included to control for autocorrelation in daily returns. The regression is very similar to the one in Pastor and Stambaugh (2003), except that they include only lagged signed volume, and their measure of liquidity is based on return reversal rather than price impact. The approach here is closer in spirit to \( \text{Iliq} \), discussed earlier, but controls for lagged volume and lagged return. An aggregate price impact measure is constructed by averaging \( \lambda_{i,t} \) across stocks and scaling by market capitalization as in the case of \( \text{Iliq} \).
Following Gabaix, Gopikrishnan, Plerou, and Stanley (2003), I use the square-root of dollar volume in regression (22) to get $\lambda_{it}^{\text{sq}}$ which, when averaged, gives an aggregate price impact based on the square-root of dollar volume. Results using these two price impact measures are similar to those provided in Table II.

As shown in Kyle (1985) and Glosten and Milgrom (1985), potential informed trading is an important determinant of liquidity. Time-varying informed trading would be less of a concern at a market-wide level than for individual stocks. Even then, idiosyncratic volatility is included in specification (21) to control for potential inside information and informed trading. As a robustness check, I also include two other measures of market-wide information asymmetry. One measure is month-of-the-quarter dummies. One dummy variable is assigned if the month is the first month of a quarter; for example, January, April, etc. Two other dummy variables are similarly assigned to the second month of each quarter and the last month of each quarter. These dummies are included in regression (21) in place of the constant. Due to the quarterly release of financial information by companies, it is likely that overall information asymmetry varies by month of the quarter. However, after controlling for these dummies, the significance or the magnitude of the coefficient on $Mis_m$ does not change.

Another measure of information asymmetry I use is the total number of companies covered by analysts in a month. I construct this measure using data from Institutional Brokers’ Estimate System (IBES). Arguably, the number of companies analyzed affects information production in the market and hence the potential for informed trading. The results show that the relationship between price-impact and misperception is again robust to the inclusion of this measure.
Evidence from Individual Stocks

To present, the empirical work has examined the relationship between aggregate liquidity and market-wide misperceived volatility. Market-wide movements are important as they indicate a systematic pattern in time-variation in these variables. However, the model is also applicable to individual stocks. Measures of price impact constructed from daily data for individual stocks are too noisy, so I construct measures of price impact using intra-day data from NYSE Trade and Quote (TAQ) database. I adopt a method suggested in Hasbrouck (2006). I cumulate trades for each five-minute interval. Trades are assumed to be buyer motivated or seller motivated based on the algorithm in Lee and Ready (1991). I calculate returns based on quotes every five minutes. I calculate the price impact of trade for each month for each stock by regressing returns over five-minute intervals on net trades (number of buyer motivated trades - number of seller motivated trades) during the same interval. This price impact calculated from high frequency data is called $\lambda_{hf}^{i,t}$. I match this with the implied volatility at the beginning of the month of a standardized option on that stock obtained from Optionmetrics.⁷ A standardized option is an at-the-money option with 30 days until maturity. For each stock, I construct a measure of misperception of volatility similar to the one for the index. The expectation of realized volatility is modeled as a function of two lags of realized volatility and lagged implied volatility. Volatility misperception is calculated as the difference between the implied volatility and expected realized volatility and is called $Mis_{i,t}$.

I estimate the following monthly regression for each individual Stock.

$$
\lambda_{hf}^{i,t} = \beta_{i,0} + \beta_{i,1}Vol_{i,t} + \beta_{i,2}R_{i,t} + \beta_{i,3}Mis_{i,t-1} + \omega_{i,t}
$$

(23)
where $Vol_{i,t}$ is the realized volatility of the return on stock $i$ during month $t$ and $R_{i,t}$ is the return on stock $i$ during month $t$. A summary of the monthly time-series regressions for individual stocks is presented in Table III.

The period covered is from January 1996 to December 2004, or shorter if data are not available for the full period. I require that at least 31 valid monthly observations are available for a stock to be included. Standard errors are calculated from the standard deviation of the coefficients across stocks. These standard errors are similar in spirit to those suggested by Fama and MacBeth (1973). I also calculate the standard errors based on the covariance matrix of the coefficients across stocks. This method takes care of the cross-correlation in coefficients across stocks. The average coefficient of misperception is positive, as expected, and significant for both NYSE / Amex as well as Nasdaq, using either method of calculating standard errors. About 62% to 65% of the coefficients are positive. A binomial test convincingly rejects (p-values of less than 0.001) the null that the coefficient of $Mis_{i,t-1}$ is positive or negative with equal probability.

In summary, the results for individual stocks, like the aggregate results, support the untested predictions of the model.

An interesting relationship to examine is that between illiquidity of individual stocks and market-wide misperception of volatility. I regress the price impact for each individual stock on its own misperception of volatility ($Mis_{i,t}$), market-wide misperception of volatility ($Mis_m$, as defined in (16)), the realized volatility of the return on the stock, and the contemporaneous return on the
stock as follows.

\[
\lambda_{i,t}^{hf} = \kappa_{i,0} + \kappa_{i,1} Vol_{i,t} + \kappa_{i,2} R_{i,t} + \kappa_{i,3} Mis_{i,t-1} + \kappa_{i,4} Mis_{m,t-1} + \zeta_{i,t}
\]  

The results are presented in Table IV. Again the t-stats are calculated using Fama-MacBeth type standard errors as well as standard errors based on the covariance matrix for coefficients across stocks. As seen in the table, the coefficients of own misperception as well as market-wide misperception are positive and significant. A binomial test again convincingly rejects (p-values of less than 0.001) the null hypotheses that the coefficient of own misperception and market-wide misperception are positive or negative with equal probability. Thus the illiquidity of a stock responds positively to both misperception at the individual level and at the market level.

Next, I examine the pattern of coefficient \( \kappa_{i,4} \) across stocks. \( \kappa_{i,4} \) captures the responsiveness of illiquidity of individual stocks to market-wide misperception of volatility. I examine whether the illiquidity of more volatile stocks is more responsive to the market-wide misperception of volatility. I rank stocks based on the volatility of their daily returns during December 1995, the month directly preceding the period under examination (January 1996-December 2004). Table V presents the results for the coefficient \( \kappa_{i,4} \) for different volatility quintiles for NYSE / Amex and Nasdaq. \( \kappa_{i,4} \) is generally increasing for more volatile stocks. The difference between the average coefficients for the highest and the lowest volatility quintiles is positive. It is statistically significant at 5% level using standard errors corrected for cross-sectional correlation across stocks. I find similar results when the samples are split into two or four groups of equal size. Thus, illiquidity of more volatile stocks
is also more responsive to the market-wide misperception of volatility. I also calculate Spearman’s rank correlation coefficient between to the historical volatility and $\kappa_{i,t}$ across stocks. For NYSE / Amex the correlation is 0.13, and for Nasdaq it is 0.17. Both numbers are significant at 1% level. This examination rules out the possibility that the pattern across stocks is driven by outliers, and confirms that illiquidity of the more volatile stocks comoves more with $Mis_m$.

Part of the above effect comes from the fact that more volatile stocks are also more illiquid and have more volatile illiquidity. However, the higher magnitude and higher variability is not the only factor in determining why volatile stocks have higher $\kappa_{i,t}$. The standardized coefficient (scaled by standard deviation of $\lambda_{i,t}$) is also higher for the top volatility quintile as compared to the bottom volatility quintile. The difference is not significant for NYSE / Amex, but is significant at 10% level for Nasdaq. The rank correlation between the standardized coefficient is positive for NYSE / Amex, and positive and significant at 5% level for Nasdaq. I also looked at the pattern of Pearson correlation coefficient $\rho$ between illiquidity of the individual stocks, $\lambda_{i,t}$, and market-wide misperception, $Mis_{m,t-1}$. This correlation is also higher for more volatile stocks. The rank correlation between $\rho$ and historical volatility is positive, though not significant, for NYSE / Amex and positive and significant at 1% level for Nasdaq.

These results are similar to the ones in Acharya and Pedersen (2005) which suggest that the illiquidity of more volatile stocks also has higher sensitivity to aggregate liquidity. Thus when the aggregate liquidity dries up, liquidity for the more volatile stocks dries up even more. Here, I
find that illiquidity of more volatile stocks is more responsive to the market-wide misperception of
volatility. When the investors perceive the market to be volatile, they withdraw liquidity, parti-
cularly from the more volatile stocks. Thus, there is flight to safety within the universe of stocks.

III Conclusions

This paper presents a model that explains the joint behavior of liquidity and volatility. I show that
when investors extrapolate recent price movements to form expectations about future volatility,
liquidity and realized volatility are time-varying and persistent. The assumption that the investors
use recent history of volatility as an input for their volatility forecast is robustly motivated by
evidence from psychology and financial markets, and industry standards for risk management. The
model explains a number of empirical regularities about liquidity, such as co-movement with con-
temporaneous return, co-movement with expected return, and flight to safety. Even though the
fundamentals are homoskedastic, the model endogenously generates empirically observed phenom-
ena of heteroskedasticity and volatility clustering in the price of the risky asset.

The empirical work provides support for the model’s previously untested prediction about volatility
misperceptions and stock market liquidity. I proxy for the misperception of volatility using the
difference between option-implied volatility and expected realized volatility. Consistent with the
model, volatility misperception is negatively correlated with liquidity. This is true at the aggregate
level as well as at the level of individual stocks. Illiquidity of individual stocks responds to own
misperception as well as market-wide misperception of volatility. The latter relationship is stronger
in the case of more volatile stocks. Thus it appears that when the investors believe that markets are more volatile, they pull out of providing liquidity for already volatile stocks in a flight to safety behavior.

Forecasts of volatility are inputs for various financial decisions, including portfolio choice and risk management at the individual as well as institutional level. This paper examines the equilibrium implications of the wide-spread use of extrapolative volatility forecasts, and contributes to the understanding of liquidity and volatility by providing a unified and endogenous explanation for their variation and persistence over time.
Appendix

This appendix contains a collection of derivatives and proofs.

A Expression for $\sigma_p^2$ in (5)

The price of the risky asset is given by (5) as follows:

$$P_t = \tilde{d} - \frac{\gamma(\tilde{\theta}\sigma^2_P + \sigma^2_d)}{r} - \frac{\gamma(\theta_t - \tilde{\theta})\sigma_P^2}{1 + r} + \frac{\gamma(\theta_t\sigma^2_P + \sigma^2_d)}{1 + r} z_t$$

Taking variance of both sides,

$$\sigma^2_p = \frac{\gamma^2\sigma^4_P}{(1 + r)^2} Var(\theta_t - \tilde{\theta}) + \frac{\gamma^2(0.5(\theta_H\sigma^2_p + \sigma^2_d)^2 + 0.5(\theta_L\sigma^2_p + \sigma^2_d)^2)}{(1 + r)^2} \sigma^2_z$$

$$\sigma^2_p = 0.25\gamma^2\sigma^4_P (\theta_H - \theta_L)^2 + 0.5\gamma^2\sigma^2_\theta [\theta_H^2 + \theta_L^2] \sigma^4_p + 2(\theta_H + \theta_L)\sigma^2_p \sigma^2_d + 2\sigma^4_d$$

Solving this quadratic equation in $\sigma^2_p$ we get,

$$\sigma^2_p = \frac{\gamma^2\sigma^4_P}{(1 + r)^2} \left[0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H + \theta_L)\sigma^2_\theta\sigma^2_Z \right] + \frac{\gamma^2\sigma^2_\theta}{(1 + r)^2} (\theta_H + \theta_L)\sigma^2_d + \frac{\gamma^2\sigma^2_d}{(1 + r)^2}$$

Solving this quadratic equation in $\sigma^2_p$ we get,

$$\sigma^2_p = \frac{(1 + r)^2 - \gamma^2\sigma^2_d\sigma^2_\theta(\theta_H + \theta_L) + \gamma^2\sigma^2_d\sigma^2_\theta(\theta_H + \theta_L)^2 + 0.5(\theta_H + \theta_L)^2}{2\gamma^2 \left[0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma^2_\theta\sigma^2_Z \right]}$$

$$\pm \sqrt{(1 + r)^4 - 2\gamma^2\sigma^2_d\sigma^2_\theta(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma^4_\theta\sigma^2_d(\sigma^2_\theta + 1)(\theta_H - \theta_L)^2}$$

$$\frac{2\gamma^2 \left[0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma^2_\theta \sigma^2_Z \right]}{2\gamma^2 \left[0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma^2_\theta \sigma^2_Z \right]}$$

(A.1)

Provided

$$(1 + r)^4 - 2\gamma^2\sigma^2_d\sigma^2_\theta(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma^4_\theta\sigma^2_d(\sigma^2_\theta + 1)(\theta_H - \theta_L)^2 \geq 0$$

This translates to

$$\frac{(1 + r)^2 \geq \gamma^2\sigma^2_d\sigma^2_\theta(\theta_H + \theta_L) + \sqrt{\gamma^4\sigma^4_\theta\sigma^2_d(2\theta^2_H + 2\theta^2_L)\sigma^2_Z + (\theta_H - \theta_L)^2}}{(1 + r)^2}$$

(A.2)
Condition (A.2) simply puts a restriction on the parameters so that $\gamma$, $\sigma_1^2$, $\sigma_2^2$, $\theta_H$ and $\theta_L$ are not too large relative to $1 + r$. If this condition is not satisfied, a stationary equilibrium does not exist.

Both values of $\sigma_P^2$ in (A.1) are positive. However, we want $\sigma_P^2 \to 0$ as $\sigma_d^2 \to 0$ or as $\gamma \to 0$. This makes sense, because with no fundamental volatility, the volatility of the price should also be 0. Also, the volatility of price comes about because of risk premium and illiquidity, both of which are results of risk-aversion. Thus if the investors were risk-neutral, there would be no risk-premium, no illiquidity and hence no volatility of price. Imposing the condition that $\sigma_P^2 \to 0$ as $\sigma_d^2 \to 0$, we get

$$\sigma_P^2 = \frac{(1 + r)^2 - \gamma^2 \sigma_d^2 \sigma_2^2 (\theta_H + \theta_L)}{2\gamma^2 [0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2]} - \frac{\sqrt{(1 + r)^4 - 2\gamma^2 \sigma_d^2 \sigma_2^2 (1 + r)^2 (\theta_H + \theta_L) - \gamma^4 \sigma_d^4 \sigma_2^2 (\sigma_z^2 + 1)(\theta_H - \theta_L)^2}}{2\gamma^2 [0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2]}$$

(B.3)

**B Proof of Proposition 1**

Proposition 1 gives various comparative statics for $\lambda_t$. From (7)

$$\frac{\gamma(\theta_t \sigma_P^2 + \sigma_d^2)}{1 + r}$$

To calculate comparative statics for $\lambda_t$, we need comparative statics for $\sigma_P^2$. (A.3) gives an expression for $\sigma_P^2$. From that we have

$$\frac{d\sigma_P^2}{d\sigma_d^2} = \frac{1}{2\gamma^2 [0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2]} - \frac{-2\gamma^2 \sigma_d^2 (1 + r)^2 (\theta_H + \theta_L) - 2\gamma^4 \sigma_d^4 \sigma_2^2 (\sigma_z^2 + 1)(\theta_H - \theta_L)^2}{2\sqrt{(1 + r)^4 - 2\gamma^2 \sigma_d^2 \sigma_2^2 (1 + r)^2 (\theta_H + \theta_L) - \gamma^4 \sigma_d^4 \sigma_2^2 (\sigma_z^2 + 1)(\theta_H - \theta_L)^2}}$$

35
\[
\frac{d\sigma_P^2}{d\sigma_z^2} \left[ \frac{2\gamma^2 \left[ 0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2 \right]}{\sqrt{(1 + r)^4 - 2\gamma^2\sigma_d^2\sigma_z^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_d^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2}} \right]
\]

\[
= -\gamma^2 \sigma_z^2 (\theta_H + \theta_L) \sqrt{(1 + r)^4 - 2\gamma^2\sigma_d^2\sigma_z^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_d^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2} 
\]

\[
+ \gamma^2 \sigma_d^2(1 + r)^2(\theta_H + \theta_L) + \gamma^4\sigma_d^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2 
\]

\[
= \gamma^2 \sigma_z^2 \left[ \frac{(1 + r)^2(\theta_H + \theta_L) + \gamma^2\sigma_d^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2}{-(\theta_H + \theta_L)\sqrt{(1 + r)^4 - 2\gamma^2\sigma_d^2\sigma_z^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_d^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2}} \right] 
\]

\[
\geq \gamma^2 \sigma_z^2 \left[ \frac{(1 + r)^2(\theta_H + \theta_L)}{-(\theta_H + \theta_L)\sqrt{(1 + r)^4 - 2\gamma^2\sigma_d^2\sigma_z^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_d^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2}} \right] 
\]

\[
= \gamma^2 \sigma_z^2 (\theta_H + \theta_L) \left[ \sqrt{(1 + r)^4 - (1 + r)^4 - 2\gamma^2\sigma_d^2\sigma_z^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_d^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2} \right] 
\]

\[
\geq 0 
\]

Hence

\[
\frac{d\sigma_P^2}{d\sigma_z^2} \geq 0 
\]

This says that the volatility of price is nondecreasing in the fundamental volatility. This result is quite intuitive. More volatile the dividends, more volatile is the price of the risky asset.

Following derivation finds the effect of \( \sigma_z^2 \) on \( \sigma_P^2 \).

\[
\frac{d\sigma_P^2}{d\sigma_z^2} = \frac{1}{(2\gamma^2 \left[ 0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2 \right])^2} 
\]
\[
\frac{d\sigma_p}{d\sigma_2} \left[ \frac{(2\gamma^2 [0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma^2_2])}{\sqrt{(1 + r)^4 - 2\gamma^2\sigma_2^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_2^4\sigma_2^2(\sigma_2^2 + 1)(\theta_H - \theta_L)^2}} \right] \\
= -0.5\gamma^4\sigma_2^4(\theta_H + \theta_L)(\theta_H - \theta_L)^2 \sqrt{(1 + r)^4 - 2\gamma^2\sigma_2^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_2^4\sigma_2^2(\sigma_2^2 + 1)(\theta_H - \theta_L)^2} \\
- \gamma^4\sigma_2^4(\theta_H + \theta_L)(\theta_H^2 + \theta_L^2) \sqrt{(1 + r)^4 - 2\gamma^2\sigma_2^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_2^4\sigma_2^2(\sigma_2^2 + 1)(\theta_H - \theta_L)^2} \\
+ \gamma^4\sigma_2^4(\theta_H + \theta_L)(\theta_H^2 + \theta_L^2) \sqrt{(1 + r)^4 - 2\gamma^2\sigma_2^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_2^4\sigma_2^2(\sigma_2^2 + 1)(\theta_H - \theta_L)^2} \\
+ 0.5\gamma^6\sigma_2^4(1 + r)^2(\theta_H + \theta_L)(\theta_H - \theta_L)^2 + 0.5\gamma^6\sigma_2^4(\theta_H - \theta_L)^4 + 0.25\gamma^6\sigma_2^4(\theta_H - \theta_L)^4 \\
+ \gamma^4\sigma_2^4(1 + r)^2(\theta_H + \theta_L)(\theta_H^2 + \theta_L^2) + 2(\theta_H^2 + \theta_L^2)(1 + r)^4 \\
- 2\gamma^4\sigma_2^4(1 + r)^2(\theta_H + \theta_L)(\theta_H^2 + \theta_L^2) - \gamma^6\sigma_2^4(\theta_H - \theta_L)^2(\theta_H^2 + \theta_L^2) \\
- \gamma^6\sigma_2^4(\theta_H - \theta_L)^2(\theta_H^2 + \theta_L^2) \\
= 0.5\gamma^4(\theta_H + \theta_L)(\theta_H - \theta_L)^2 \\
\left[ (1 + r)^2 - \sqrt{(1 + r)^4 - 2\gamma^2\sigma_2^2(1 + r)^2(\theta_H + \theta_L) - \gamma^4\sigma_2^4(\sigma_2^2 + 1)(\theta_H - \theta_L)^2} \right] \\
+ 0.5\gamma^6\sigma_2^4(\theta_H - \theta_L)^4 + 0.25\gamma^6\sigma_2^4(\theta_H - \theta_L)^4
\]
\[ + \gamma^2 (\theta_H^2 + \theta_L^2) \begin{bmatrix}
(1 + r)^4 - \gamma^2 \sigma_p^2 (1 + r)^2 (\theta_H + \theta_L) & -0.5 \gamma^4 \sigma_d^4 \sigma_z^2 (\sigma_H^2 + 1)(\theta_H - \theta_L)^2 \\
-(1 + r)^2 \sqrt{(1 + r)^4 - 2 \gamma^2 \sigma_p^2 (1 + r)^2 (\theta_H + \theta_L) - \gamma^4 \sigma_d^4 \sigma_z^2 (\sigma_H^2 + 1)(\theta_H - \theta_L)^2} & \end{bmatrix} \geq 0 \]

since

\[ (1 + r)^2 \geq \sqrt{(1 + r)^4 - 2 \gamma^2 \sigma_p^2 (1 + r)^2 (\theta_H + \theta_L) - \gamma^4 \sigma_d^4 \sigma_z^2 (\sigma_H^2 + 1)(\theta_H - \theta_L)^2} \]

and

\[ (1 + r)^4 - \gamma^2 \sigma_p^2 (1 + r)^2 (\theta_H + \theta_L) - 0.5 \gamma^4 \sigma_d^4 \sigma_z^2 (\sigma_H^2 + 1)(\theta_H - \theta_L)^2 \geq (1 + r)^2 \sqrt{(1 + r)^4 - 2 \gamma^2 \sigma_p^2 (1 + r)^2 (\theta_H + \theta_L) - \gamma^4 \sigma_d^4 \sigma_z^2 (\sigma_H^2 + 1)(\theta_H - \theta_L)^2} \]

Thus

\[ \frac{d\sigma_p^2}{d\gamma^2} \geq 0 \]

This means that the volatility of price of the risky asset \((\sigma_p^2)\) is nondecreasing in the volatility of the liquidity demand \((\sigma_z^2)\). It is intuitive that greater volatility of the liquidity demand would result in more volatile prices.

Following derivation finds the effect of \(\gamma\) on \(\sigma_p^2\).

\[ \frac{d\sigma_p^2}{d\gamma^2} = \frac{1}{2 \gamma^2 \left[ 0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2) \sigma_z^2 \right]^2} \]

\[ - \left( \sigma_p^2 (\theta_H + \theta_L) \left[ 2 \gamma^4 \left[ 0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2) \sigma_z^2 \right] \right] \right) \]

\[ - \left( 2 \left[ 0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2) \sigma_z^2 \right] \left[ (1 + r)^2 - \gamma^2 \sigma_p^2 (\theta_H + \theta_L) \right] \right) \]

\[ - \frac{1}{2 \sqrt{(1 + r)^4 - 2 \gamma^2 \sigma_p^2 (1 + r)^2 (\theta_H + \theta_L) - \gamma^4 \sigma_d^4 \sigma_z^2 (\sigma_H^2 + 1)(\theta_H - \theta_L)^2}} \]

38
\[
\begin{align*}
&\left[-2\sigma_r^2\sigma_z^2(1+r)^2(\theta_H + \theta_L) - 2\gamma^2\sigma_r^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2\right] \\
&\quad + \left[2\gamma^2 \left[0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2\right]\right] \\
&\quad + \left(\frac{2 \left[0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2\right]}{\sqrt{(1+r)^4 - 2\gamma^2\sigma_r^2\sigma_z^2(1+r)^2(\theta_H + \theta_L) - \gamma^4\sigma_r^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2}}\right)
\end{align*}
\]

\[
d\frac{\sigma_p^2}{d\gamma^2} = \begin{cases}
\left[2\gamma^4 \left[0.25(\theta_H - \theta_L)^2 + 0.5(\theta_H^2 + \theta_L^2)\sigma_z^2\right]\right] \\
\left[\left[(1+r)^2 - \gamma^2\sigma_r^2\sigma_z^2(\theta_H + \theta_L)\right]\right] \\
\left[(1+r)^4 - 2\gamma^2\sigma_r^2\sigma_z^2(1+r)^2(\theta_H + \theta_L) - \gamma^4\sigma_r^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2\right]
\end{cases}
\]

Since

\[
\left[(1+r)^2 - \gamma^2\sigma_r^2\sigma_z^2(\theta_H + \theta_L)\right] \\
\geq \sqrt{(1+r)^4 - 2\gamma^2\sigma_r^2\sigma_z^2(1+r)^2(\theta_H + \theta_L) - \gamma^4\sigma_r^2\sigma_z^2(\sigma_z^2 + 1)(\theta_H - \theta_L)^2},
\]

\[
\frac{d\sigma_p^2}{d\gamma^2} \geq 0
\]

39
Hence
\[ \frac{d\sigma_P^2}{d\gamma} \geq 0 \]

This comparative static means that more risk-averse the investor greater is the volatility of price. More risk-averse investors demand a greater compensation for accommodating the liquidity demand and thus moving prices around more.

From the comparative statics for \( \sigma_P^2 \) and from the expression for \( \lambda_t \), it is straightforward that
\[ \frac{d\lambda_t}{d\sigma^2} \geq 0 \]
\[ \frac{d\lambda_t}{d\sigma^2_z} \geq 0 \]
\[ \frac{d\lambda_t}{d\sigma^2_z} \geq 0 \]

C Proof of Proposition 4

Proposition 4 states that liquidity and realized volatility of the price of the risky asset are persistent. The first step of proving this is to prove that the investors are more likely to believe that volatility is high at \( t + 1 \) if they believed so at \( t \); i.e.,
\[ \mathcal{P}(\theta_{t+1} = \theta_H|\theta_t = \theta_H) \geq \mathcal{P}(\theta_{t+1} = \theta_H|\theta_t = \theta_L) \]

This can be shown as follows. Let \( \mathcal{P} \) be the probability.
\[ \mathcal{P}(\theta_{t+1} = \theta_H|\theta_t) = \mathcal{P}(\frac{\gamma(\theta_t \sigma_P^2 + \sigma^2_z)}{1 + r}|z_t > M) \]
\[ = \mathcal{P} \left( z_t > \frac{M(1 + r)}{\gamma(\theta_t \sigma_P^2 + \sigma^2_z)} \right) + \mathcal{P} \left( z_t < -\frac{M(1 + r)}{\gamma(\theta_t \sigma_P^2 + \sigma^2_z)} \right) \]
Since $\theta_H \geq \theta_L$,

$$\frac{M(1+r)}{\gamma(\theta_H \sigma_P^2 + \sigma_d^2)} \leq \frac{M(1+r)}{\gamma(\theta_L \sigma_P^2 + \sigma_d^2)} \quad \text{and} \quad \frac{-M(1+r)}{\gamma(\theta_H \sigma_P^2 + \sigma_d^2)} \geq \frac{-M(1+r)}{\gamma(\theta_L \sigma_P^2 + \sigma_d^2)}$$

and

$$\mathcal{P}\left(z_t > \frac{M(1+r)}{\gamma(\theta_H \sigma_P^2 + \sigma_d^2)}\right) \geq \mathcal{P}\left(z_t > \frac{M(1+r)}{\gamma(\theta_L \sigma_P^2 + \sigma_d^2)}\right)$$

and

$$\mathcal{P}\left(z_t < \frac{-M(1+r)}{\gamma(\theta_H \sigma_P^2 + \sigma_d^2)}\right) \geq \mathcal{P}\left(z_t < \frac{-M(1+r)}{\gamma(\theta_L \sigma_P^2 + \sigma_d^2)}\right)$$

$$\mathcal{P}(\theta_{t+1} = \theta_H | \theta_t = \theta_H) \geq \mathcal{P}(\theta_{t+1} = \theta_H | \theta_t = \theta_L) \quad \text{(A.4)}$$

The next step is to prove that $\theta_t$ is positively autocorrelated. This is done below.

$$Cov(\theta_t, \theta_{t+1}) = E[\theta_t \theta_{t+1}] - E[\theta_t]E[\theta_{t+1}]$$

Let

$$\mathcal{P}(\theta_{t+1} = \theta_H | \theta_t = \theta_H) \equiv \mathcal{P}_{t+1}^{HH}$$

$$\mathcal{P}(\theta_{t+1} = \theta_H | \theta_t = \theta_L) \equiv \mathcal{P}_{t+1}^{HL}$$

$$\mathcal{P}(\theta_{t+1} = \theta_L | \theta_t = \theta_H) \equiv \mathcal{P}_{t+1}^{LH}$$

$$\mathcal{P}(\theta_{t+1} = \theta_L | \theta_t = \theta_L) \equiv \mathcal{P}_{t+1}^{LL}$$

$$\mathcal{P}(\theta_t = \theta_H) \equiv \mathcal{P}_t^H$$

$$\mathcal{P}(\theta_t = \theta_L) \equiv \mathcal{P}_t^L$$

Using the above definitions,

$$Cov(\theta_t, \theta_{t+1}) = \mathcal{P}_t^H \theta_H \left[ \mathcal{P}_{t+1}^{HH} \theta_H + \mathcal{P}_{t+1}^{HL} \theta_L \right] + \mathcal{P}_t^L \theta_L \left[ \mathcal{P}_{t+1}^{HL} \theta_H + \mathcal{P}_{t+1}^{LL} \theta_L \right]$$

$$- \left[ \mathcal{P}_t^H \theta_H + \mathcal{P}_t^L \theta_L \right] \left[ \theta_H (\mathcal{P}_{t+1}^{HH} \mathcal{P}_t^H + \mathcal{P}_{t+1}^{HL} \mathcal{P}_t^L) + \theta_L (\mathcal{P}_{t+1}^{LL} \mathcal{P}_t^L + \mathcal{P}_{t+1}^{LH} \mathcal{P}_t^H) \right]$$
Using

\[ P_t^L = 1 - P_t^H, \ P_{t+1}^{LL} = 1 - P_{t+1}^{HL}, \text{ and } P_{t+1}^{LH} = 1 - P_{t+1}^{HH} \]

and rearranging,

\[ \text{Cov}(\theta_t, \theta_{t+1}) = \ P_t^H(1 - P_t^H)(\theta_H - \theta_L)^2 \left[ P_{t+1}^{HH} - P_{t+1}^{HL} \right] \]

From (A.4) we know that

\[ \left[ P_{t+1}^{HH} - P_{t+1}^{HL} \right] \geq 0 \]

Thus,

\[ \text{Cov}(\theta_t, \theta_{t+1}) \geq 0 \]

Illiquidity, \( \lambda_t \), and conditional volatility of the price, \( Vol_t \), are linear functions of \( \theta_t \). Given the positive autocorrelation in \( \theta_t \), it is evident that

\[ \text{Cov}(\lambda_t, \lambda_{t+1}) \geq 0 \text{ and } \text{Cov}(Vol_t, Vol_{t+1}) \geq 0 \]

D Equilibrium in Subsection I.B

Subsection I.B develops an extension of the model in which the risk-free rate is endogenized. Expression (10) gives the price of the risky asset as

\[ P_t = \frac{\bar{d}}{1 + r_t} + \frac{E_t[P_{t+1}]}{1 + r_t} - \gamma(\theta_t\sigma_p^2 + \sigma_d^2) + \gamma(\theta_t\sigma_p^2 + \sigma_d^2)z_t \]

\[ P_t = (\bar{d} + E_t[P_{t+1}] - \gamma(\theta_t\sigma_p^2 + \sigma_d^2) + \gamma(\theta_t\sigma_p^2 + \sigma_d^2)z_t) P_t^s \]
Given the beliefs of the investors,

\[ E^t_i[P_{t+1}] = E^t_i[P_{t+2}] = \cdots = \mathcal{P} \]

\[ \mathcal{P} = \frac{dP^s - \gamma(\theta P^s \sigma_P^2 + \bar{P} \sigma_d^2) - \gamma(\theta P^s z \sigma_P^2 + \bar{P} z \sigma_d^2)}{1 - P^s} \]

Substituting this in the expression for \( P_t \) above gives

\[ P_t = \left( \frac{\bar{d}}{1 - P^s} - \left( \frac{\gamma(\theta P^s \sigma_P^2 + \sigma_d^2 + \theta P^s z \sigma_P^2 + \bar{P} z \sigma_d^2)}{1 - P^s} \right) - \gamma(\theta t \sigma_P^2) + \gamma(\theta t \sigma_P^2 + \sigma_d^2)z_t \right) P^s_t \]

where

\[ \mathcal{P} = \left( \frac{\gamma(\theta P^s \sigma_P^2 + \sigma_d^2 + \theta P^s z \sigma_P^2 + \bar{P} z \sigma_d^2)}{1 - P^s} \right) \]

Expressions of the form \( \bar{\cdot} \) give various unconditional expectations as defined below.

\[ \bar{P}^s = 0.5E_H[P^s_t] + 0.5E_L[P^s_t] \]

\[ \bar{P} = 0.5E_H[P_t] + 0.5E_L[P_t] \]

\[ \bar{\theta} P^s = 0.5\theta_H E_H[P^s_t] + 0.5\theta_L E_L[P^s_t] \]

\[ \bar{\theta} P^s = 0.5\theta_H E_H[P^s_t] + 0.5\theta_L E_L[P^s_t] \]

\[ \bar{\theta} P^s z = 0.5\theta_H E_H[P^s_t z_t] + 0.5\theta_L E_L[P^s_t z_t] \]

\[ \bar{P} z = 0.5E_H[P^s_t z_t] + 0.5E_L[P^s_t z_t] \]

where

\[ E_H[.] = E^t[.|\theta_t = \theta_H] \]

\[ Var_H[.] = Var^t[.|\theta_t = \theta_H] \]

\[ Cov_H[.] = Cov^t[.|\theta_t = \theta_H] \]
\( E_L[.] \), \( \text{Var}_L[.] \) and \( \text{Cov}_L[.] \) are defined accordingly. In a steady-state equilibrium, \( \sigma^2_P \) is determined as follows:

\[
\sigma^2_P = 0.5 \left[ \text{Var}_H[P_t] + (E_H[P_t])^2 \right] + 0.5 \left[ \text{Var}_L[P_t] + (E_L[P_t])^2 \right] - (\bar{P})^2
\]

(A.5)

All expressions involving \( E_H[.] \), \( \text{Var}_H[.] \), \( \text{Cov}_H[.] \), \( E_L[.] \), \( \text{Var}_L[.] \) and \( \text{Cov}_L[.] \) are functions of \( \sigma^2_P \) and the basic parameters of the model (i.e., \( \sigma^2_d \), \( \bar{d} \), \( \gamma \), \( S \), \( W \), \( \theta_H \), \( \theta_L \) and parameters describing the distribution of \( z_t \)). Thus, (A.5) gives the fixed point condition for \( \sigma^2_P \), implicitly defining \( \sigma^2_P \) in terms of basic parameters. This is the condition for stationary equilibrium of the model with an endogenous riskless rate.

### E Derivation of (15)

Expression (15) defines \( \lambda_t \) when the risk-free rate is endogenized. Equation (13) gives the price of the risky asset

\[
P_t = \left( \frac{\bar{d}}{1 - P^s} - \gamma \theta_t \sigma^2_P + \gamma(\theta_t \sigma^2_P + \sigma^2_d) z_t \right) P^s_t = XP_t^s
\]

This makes \( P_t^s = \frac{W}{S+X} \).

\[
\lambda_t = \frac{dP_t}{dz_t} = \gamma(\theta_t \sigma^2_P + \sigma^2_d) P^s_t + \frac{W}{(S+X)^2} \gamma((\theta_t \sigma^2_P + \sigma^2_d)) P^s_t - X \left( \frac{P^s_t}{(S+X)^2} \gamma((\theta_t \sigma^2_P + \sigma^2_d)) \right) = \gamma(\theta_t \sigma^2_P + \sigma^2_d) P^s_t - X \left( \frac{P^s_t}{(S+X)^2} \gamma((\theta_t \sigma^2_P + \sigma^2_d)) \right) = \gamma(\theta_t \sigma^2_P + \sigma^2_d) P^s_t \left( 1 - X \frac{1}{(S+X)} \right)
\]
\[
\begin{align*}
&= \gamma(\theta_t \sigma_P^2 + \sigma_d^2) P_t^s \frac{S}{(S + X)} \\
&= \frac{S}{W} \gamma(\theta_t \sigma_P^2 + \sigma_d^2)(P_t^s)^2
\end{align*}
\]

\section*{F Proof of Proposition 5}

Proposition 5 states that \(Cov(\lambda_t, r_t) \geq 0\). Using the expression for \(\lambda_t\) in (15),

\[
Cov(\lambda_t, 1 + r_t) = Cov\left(\frac{S}{W} \gamma(\theta_t \sigma_P^2 + \sigma_d^2)(P_t^s)^2, \frac{1}{P_t^s}\right)
\]

It can be shown that

\[
\frac{d\lambda_t}{d\theta_t} \geq 0, \quad \frac{d\lambda_t}{dz_t} \geq 0, \quad \frac{d(1 + r_t)}{d\theta_t} \leq 0 \quad \text{and} \quad \frac{d(1 + r_t)}{dz_t} \leq 0
\]

\(\theta_t\) and \(z_t\) are the two shocks that drive both \(\lambda_t\) and \(r_t\), and they are independent from each other. Thus, using a Taylor series expansion of \(\lambda_t\) and \(1 + r_t\) in terms of \(\theta_t\) around \(\bar{\theta}\) and 0, respectively,

\[
\begin{align*}
Cov(\lambda_t, 1 + r_t) &\approx \frac{d\lambda_t}{d\theta_t} \frac{d(1 + r_t)}{d\theta_t} Var(\theta_t) + \frac{d\lambda_t}{dz_t} \frac{d(1 + r_t)}{dz_t} \sigma_z^2 \\
&\leq 0
\end{align*}
\]
References


46


Notes

Here $\lambda_t$ depends on the risk-aversion of the market-makers, similar to the one in Grossman and Miller (1988), but unlike the one in Kyle (1985) or Glosten and Milgrom (1985). In the latter two papers, insider information rather than the risk-aversion of the market-maker is the source of illiquidity. Hence, price impact in those models does not depend on risk-aversion.

Conditions for determining $\sigma^2_P$ in case of endogenous risk-free rate are given in the appendix.

Anecdotal evidence suggests that when liquidity of the risky asset dries up, investors transfer the capital to the safe asset, pushing down the interest rate. Some related evidence is provided in Krishnamurthy (2002) and Longstaff (2004). They find that spreads between illiquid/risky assets and liquid/safe assets widen during flight to safety episodes.

Recently the CBOE has changed the construction of VIX. The new VIX is based on S&P 500 index options. In this paper I use the original VIX, but all the results are similar for tests using the new VIX.

The results that follow are not dependent on this particular specification. For example, similar results are obtained using a GARCH(1,1) or with a specification that doesn’t include VIX as the explanatory variable for $E_t[RV_{t+1}]$. For a survey of volatility forecasting, see Poon and Granger (2003).

When calculating this average, stocks with prices less than $5$ or greater than $1000$ at the end of the previous month are excluded. Also, stocks that appear for the first time or for the last
time in that month are excluded. Finally, I exclude observations in the 0.5% tail on each side to eliminate outliers.

7I thank Yang Lu for providing the key for matching CRSP ids with Optionmetrics ids.

8There are about 1000 stocks in the cross-section. They all have different sample sizes. Some pairs of stocks do not have overlapping samples. I set the missing covariances equal to the average covariance of the remaining sample. The results are robust to setting the missing covariances equal to zero.

9The rank correlation between historical volatility of returns and average price impact, $\lambda_{hf,i,t}$, across stocks is 0.26 for NYSE / Amex, and 0.11 for Nasdaq—both significant at 1% level. The rank correlation between historical volatility of returns and volatility of price impact, $\lambda_{i,t}^{hf}$, is also positive (0.27 for NYSE / Amex and 0.11 for Nasdaq) and significant.
High Perceived Volatility

High Realized Volatility

High Illiquidity

Figure 1. Feedback loop for misperception of volatility, illiquidity and realized volatility. High perceived volatility means investors charge greater compensation to accommodate the liquidity traders, leading to high illiquidity. High illiquidity amplifies the demand shocks to generate high realized volatility of the price of the risky asset. High realized volatility feeds into misperception next period, resulting in high perceived volatility next period.
Figure 2. Time series of aggregate illiquidity, \textit{Iliq} and market-wide misperception of volatility, \textit{Mis}_m. \textit{Iliq} is the price impact of trade for individual stocks from daily data averaged across stocks, calculated separately for NYSE / Amex and Nasdaq. \textit{Mis}_m is calculated as the difference between VIX, the volatility index by CBOE based on implied volatility of S&P 100 index options, and forecasted volatility of the index returns.
Table I. Descriptive Statistics

This table provides descriptive statistics for the misperception of volatility ($Mis_{m,t}$) and aggregate illiquidity ($Illiq_t$). $Mis_{m,t}$ is calculated as the difference between VIX, the volatility index by CBOE based on implied volatility of S&P 100 index options, and forecasted volatility of the index returns. The period covered is from February 1986 to December 2004. $Illiq_t$ is the price impact of trade for individual stocks from daily data averaged across stocks, calculated separately for NYSE / Amex and Nasdaq.

<table>
<thead>
<tr>
<th></th>
<th>$Mis_{m,t}$ (Percentage)</th>
<th>$Illiq_t$ (NYSE/Amex) (Basis points per thousand dollars)</th>
<th>$Illiq_t$ (Nasdaq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.6</td>
<td>6.7</td>
<td>36.7</td>
</tr>
<tr>
<td>Median</td>
<td>5.3</td>
<td>5.9</td>
<td>37.8</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.2</td>
<td>2.5</td>
<td>11.3</td>
</tr>
</tbody>
</table>
Table II. Illiquidity and Misperception of Volatility

This table presents the results of time-series regression of $Illi_q$, the aggregate illiquidity measure for stocks calculated from daily data, on control variables and misperception of volatility. The error term has an AR(2) structure. $Illi_q$ is measured separately for NYSE/Amex and Nasdaq. $IVol_t$ and $SVol_t$ are aggregate idiosyncratic and systematic volatilities, respectively. $R_{m,t}$ is the monthly CRSP value-weighted return. $Mis_{m,t}$ is the market-wide misperception of volatility. It is calculated as the difference between VIX, the volatility index by CBOE based on implied volatility of S&P 100 index options, and forecasted volatility of the index returns. The period covered is from April 1986 to December 2004. * indicates significance at 10%, ** at 5%, and *** at 1% levels using t-statistic corrected for heteroscedasticity.

### Panel A: NYSE / Amex

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>$IVol_t$</th>
<th>$SVol_t$</th>
<th>$R_{m,t}$</th>
<th>$Mis_{m,t-1}$</th>
<th>Adj $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.86***</td>
<td>-0.02</td>
<td>0.09***</td>
<td>-5.10***</td>
<td>0.05***</td>
<td>0.91</td>
</tr>
<tr>
<td>t-stats</td>
<td>5.86</td>
<td>-0.98</td>
<td>5.48</td>
<td>-5.00</td>
<td>4.72</td>
<td></td>
</tr>
</tbody>
</table>

### Panel B: Nasdaq

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>$IVol_t$</th>
<th>$SVol_t$</th>
<th>$R_{m,t}$</th>
<th>$Mis_{m,t-1}$</th>
<th>Adj $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>29.63***</td>
<td>-0.03</td>
<td>0.52***</td>
<td>-14.12***</td>
<td>0.32***</td>
<td>0.88</td>
</tr>
<tr>
<td>t-stats</td>
<td>7.20</td>
<td>-0.33</td>
<td>5.61</td>
<td>-2.65</td>
<td>6.06</td>
<td></td>
</tr>
</tbody>
</table>
Table III. Illiquidity and Misperception of Volatility for Individual Stocks

This table presents a summary of the results of monthly time-series regressions of $\lambda_{hf}^{i,t}$, the price impact for individual stocks calculated from high frequency data, on control variables and own misperception of volatility ($Mis_{i,t}$). The error term has an AR(2) structure. $Vol_{i,t}$ and $R_{i,t}$ are the monthly volatility of daily returns and monthly return, respectively, for the individual stock. $Mis_{i,t}$ is the difference between the implied volatility of the individual stock options and forecasted volatility of the individual stock returns. The period covered is from January 1996 to December 2004. Binomial p-value is the p-value using binomial distribution for the hypothesis that positive and negative coefficients are equally likely. * indicates significance at 10%, ** at 5%, and *** at 1% levels.

Panel A: NYSE / Amex

<table>
<thead>
<tr>
<th>Intercept</th>
<th>$Vol_{i,t}$</th>
<th>$R_{i,t}$</th>
<th>$Mis_{i,t-1}$</th>
<th>Stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Coefficient</td>
<td>0.355</td>
<td>0.315</td>
<td>-0.129</td>
<td>0.192</td>
</tr>
<tr>
<td>t-stats$^\dagger$</td>
<td>35.99***</td>
<td>27.87***</td>
<td>-13.71***</td>
<td>7.99***</td>
</tr>
<tr>
<td>t-stats$^\ddagger$</td>
<td>27.95***</td>
<td>8.92***</td>
<td>-4.75***</td>
<td>4.33***</td>
</tr>
<tr>
<td>% of positive coefficients</td>
<td>62%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binomial p-value</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Nasdaq

<table>
<thead>
<tr>
<th>Intercept</th>
<th>$Vol_{i,t}$</th>
<th>$R_{i,t}$</th>
<th>$Mis_{i,t-1}$</th>
<th>Stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Coefficient</td>
<td>0.31</td>
<td>0.26</td>
<td>-0.07</td>
<td>0.254</td>
</tr>
<tr>
<td>t-stats$^\dagger$</td>
<td>27.29***</td>
<td>16.97***</td>
<td>-4.64***</td>
<td>8.16***</td>
</tr>
<tr>
<td>t-stats$^\ddagger$</td>
<td>15.63***</td>
<td>8.15***</td>
<td>-2.56**</td>
<td>7.69***</td>
</tr>
<tr>
<td>% of positive coefficients</td>
<td>66%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binomial p-value</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^\dagger$ t-stats based on standard errors calculated from the standard deviation of coefficients across stocks.
$^\ddagger$ t-stats based on standard errors calculated from the covariance matrix of coefficients across stocks, taking into account cross-correlation across stocks.
Table IV. Illiquidity of Individual Stocks and Market-wide Misperception of Volatility

This table presents a summary of the results of monthly time-series regressions of $\lambda_{i,t}^{hf}$, the price impact for individual stocks calculated from high frequency data, on control variables, own misperception of volatility ($Mis_{i,t}$) and market-wide misperception of volatility ($Mis_{m,t}$). The error term has an AR(2) structure. $Vol_{i,t}$ and $R_{i,t}$ are the monthly volatility of daily returns and monthly return, respectively, for the individual stock. $Mis_{i,t}$ is the difference between the implied volatility of the individual stock options and forecasted volatility of the individual stock returns. $Mis_{m,t}$ is the difference between VIX, the volatility index by CBOE based on implied volatility of S&P 100 index options, and forecasted volatility of the index returns. The period covered is from January 1996 to December 2004. Binomial p-value is the p-value using binomial distribution for the hypothesis that positive and negative coefficients are equally likely. * indicates significance at 10%, ** at 5%, and *** at 1% levels.

<table>
<thead>
<tr>
<th>Panel A: NYSE / Amex</th>
<th>Intercept</th>
<th>$Vol_{i,t}$</th>
<th>$R_{i,t}$</th>
<th>$Mis_{i,t-1}$</th>
<th>$Mis_{m,t-1}$</th>
<th>Stocks</th>
<th>Average Coefficient</th>
<th>t-stats†</th>
<th>t-stats‡</th>
<th>% of positive coefficients</th>
<th>Binomial p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.250</td>
<td>0.292</td>
<td>-0.129</td>
<td>0.196</td>
<td>0.015</td>
<td>1025</td>
<td></td>
<td>23.72***</td>
<td>7.89***</td>
<td>61%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>t-stats†</td>
<td>16.98***</td>
<td>13.16***</td>
<td>-4.76***</td>
<td>7.82***</td>
<td>919</td>
<td></td>
<td>4.32***</td>
<td>3.17***</td>
<td>62%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>t-stats‡</td>
<td>13.16***</td>
<td>-4.76***</td>
<td>7.82***</td>
<td>11.86***</td>
<td></td>
<td></td>
<td>3.83***</td>
<td>3.01***</td>
<td>80%</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Nasdaq</th>
<th>Intercept</th>
<th>$Vol_{i,t}$</th>
<th>$R_{i,t}$</th>
<th>$Mis_{i,t-1}$</th>
<th>$Mis_{m,t-1}$</th>
<th>Stocks</th>
<th>Average Coefficient</th>
<th>t-stats†</th>
<th>t-stats‡</th>
<th>% of positive coefficients</th>
<th>Binomial p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.171</td>
<td>0.227</td>
<td>-0.072</td>
<td>0.226</td>
<td>0.024</td>
<td>919</td>
<td></td>
<td>16.98***</td>
<td>3.17***</td>
<td>62%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>t-stats†</td>
<td>13.16***</td>
<td>-4.76***</td>
<td>7.82***</td>
<td>11.86***</td>
<td></td>
<td></td>
<td>3.83***</td>
<td>3.01***</td>
<td>80%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>t-stats‡</td>
<td>3.83***</td>
<td>-1.02</td>
<td>5.52***</td>
<td>3.01***</td>
<td></td>
<td></td>
<td>3.83***</td>
<td>3.01***</td>
<td>80%</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

†t-stats based on standard errors calculated from the standard deviation of coefficients across stocks.  
‡t-stats based on standard errors calculated from the covariance matrix of coefficients across stocks, taking into account cross-correlation across stocks.
Table V. Response of Price Impact for Individual Stocks to Market-wide Misperception of Volatility for Different Volatility Quintiles

This table presents coefficients for the response of price impact for individual stocks to market-wide misperception of volatility ($ Mis_{m,t} $). The regressions also include monthly volatility of daily return for the individual stock, monthly return on the individual stock, and misperception of volatility for the individual stock. The error term has an AR(2) structure. $ Mis_{m,t} $ is calculated as the difference between VIX, the volatility index by CBOE based on implied volatility of S&P 100 index options, and forecasted volatility of the index returns. The period covered is from January 1996 to December 2004. Stocks are divided into quintiles based on volatility of daily returns for December 1995. t-stats are calculated based on the covariance matrix of coefficients across stocks, taking into account cross-correlation across stocks. * indicates significance at 10%, ** at 5%, and *** at 1% levels.

<table>
<thead>
<tr>
<th></th>
<th>Historical Volatility Quintiles</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Least Volatile)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Coefficient</td>
<td>0.011</td>
<td>0.013</td>
<td>0.012</td>
<td>0.013</td>
<td>0.022</td>
</tr>
<tr>
<td>t-stats</td>
<td>2.53**</td>
<td>2.78***</td>
<td>2.75***</td>
<td>2.00**</td>
<td>2.85***</td>
</tr>
<tr>
<td>Number of stocks</td>
<td>174</td>
<td>174</td>
<td>174</td>
<td>174</td>
<td>174</td>
</tr>
</tbody>
</table>

Panel B: Nasdaq

<table>
<thead>
<tr>
<th></th>
<th>Historical Volatility Quintiles</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Least Volatile)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Coefficient</td>
<td>0.011</td>
<td>0.021</td>
<td>0.015</td>
<td>0.022</td>
<td>0.039</td>
</tr>
<tr>
<td>t-stats</td>
<td>1.36</td>
<td>2.60***</td>
<td>2.23**</td>
<td>2.46**</td>
<td>2.93***</td>
</tr>
<tr>
<td>Number of stocks</td>
<td>118</td>
<td>119</td>
<td>119</td>
<td>119</td>
<td>119</td>
</tr>
</tbody>
</table>