Why Do Investors Trade?

Eric K. Kelley and Paul C. Tetlock*

November 2012

Preliminary Draft: Please do not cite without permission

Abstract

We propose and estimate a structural model of daily stock market activity to test competing theories of trading volume. The model features informed rational speculators and risk-averse uninformed agents who may trade to hedge endowment shocks or to speculate on perceived information. To identify the model parameters, we exploit enormous empirical variation in trading volume, market liquidity, and return volatility associated with regular and extended-hours markets as well as news arrival. We find that the model matches market activity well when we allow for overconfidence. At plausible values of overconfidence and risk aversion, overconfidence—not hedging—explains nearly all uninformed trading, while rational informed speculation accounts for most overall trading. Without overconfident investors, over 99% of trading volume disappears even when informed rational traders disagree maximally. These findings illustrate that modest overconfidence can help explain stark patterns in stock market activity.

* University of Arizona, ekelley@eller.arizona.edu and Columbia University, paul.tetlock@columbia.edu. This draft incorporates elements of a previous working paper entitled “How (Un)Informed Is Trading?” We thank Lauren Cohen, Kent Daniel, Larry Glosten, Gur Huberman, Charles Jones, Rick Sias, Clara Vega, and seminar participants at Arizona, Columbia, Georgia, The NYU Stern Microstructure Conference, and Tennessee for helpful discussions. We thank Dow Jones for providing news data.
This study tests competing theories of why investors trade. DeBondt and Thaler (1995) contend that the high trading volume observed in financial markets “is perhaps the single most embarrassing fact to the standard finance paradigm.” They speculate that most volume comes from traders who are overconfident in their abilities to pick stocks. Odean (1998) proposes several models in which overconfidence results in trade and argues this activity is costly to society. Others, such as Harris and Raviv (1993) and Kim and Verrecchia (1997) model trading that results from disagreement among agents. Chordia, Roll, and Subrahmanyam (2011) show that turnover in US markets has recently increased fivefold to tens of trillions of dollars annually, but they suggest the activity actually results from informed trading by institutions. Despite this tension between various behavioral and rational explanations, there are disappointingly few tests of which models of market activity are consistent with the data.

We attempt to fill this void by estimating a structural model with agents who exhibit canonical theoretical motives for trading. Rational speculators may collect and trade on valuable information, while risk-averse uninformed agents may trade to hedge endowment shocks or to speculate on perceived information. Agents optimally choose whether to enter the market and how much to trade based on real or perceived signals, endowment shocks, and expectations of other traders’ behavior. A risk-neutral competitive market maker sets the equilibrium stock price in response to agents’ aggregate demand. Trading volume, liquidity, and return variance arise as endogenous outcomes of this market equilibrium. Importantly, the model is flexible enough to allow for disagreement among rational informed agents as well as their overconfident uninformed counterparts. It also nests both rational heterogeneous belief models and behavioral models with overconfident speculation and by endogenizing entry into the market.

\[\text{1 The model builds on Spiegel and Subrahmanyam’s (1992) analysis of rational uninformed trading by allowing for overconfident speculation and by endogenizing entry into the market.}\]
no hedgers. To our knowledge, we are the first to use trading volume and liquidity as the basis for distinguishing rational and behavioral models of stock market activity.

To identify the model parameters, we exploit enormous empirical variation in trading volume, market liquidity, and return volatility associated with the time of day and news arrival in the electronic trading era. Since 2000, revolutions in information and trading technology have enabled a dramatic expansion in the hours of market activity in US stocks. Newswires now arrive continuously throughout the day, and electronic communication networks give any institution or retail trader with a brokerage account the ability to trade outside regular market hours (from 9:30am to 4pm). Despite having round-the-clock information and increased access to the financial markets, most still choose to trade in regular market hours. Less than 5% of total trading in our sample takes place in the pre-market (7am to 9:30am in our study) and the after-market (4pm to 6:30pm), and both are far less liquid than the regular market. In contrast, stock return volatility during extended hours periods is more than half of that during the regular market. In fact, in periods when news arrives, extended hours volatility is on par with that in regular hours. This rich set of stylized facts informs the estimation of our structural model’s parameters.

Our estimates indicate that an unrestricted model allowing for rational informed trading, uninformed hedging, and overconfident speculation fits the data well. When we allow the information environment to vary over time, the model replicates the stark empirical differences in trading activity, liquidity, and volatility across intraday periods. We cannot statistically reject the model’s over-identifying restrictions as the model closely matches the magnitudes of virtually all of the empirical moments. Equally important, the key model parameters are well-identified and exhibit plausible magnitudes.
Our parameter estimates shed light on the nature of trading volume. To explain market activity, uninformed agents, whose perceived signals correctly predict the direction of future returns with a 0.5 probability (i.e., they are pure noise), only need to believe that they observe a signal that is correct 56% of the time. The strength of their belief in their signal’s accuracy is modest in comparison to survey and experimental evidence on overconfidence in the precision of information. Moreover, it would take hundreds of independent trials for uninformed agents to learn that they are overconfident, even if they accurately remember their successes and failures.

Another stark result is that hedging activity is negligible in comparison to overall trading and even to total uninformed trading when we assume plausible parameter values. Specifically, this result obtains whenever the relative risk aversion (RRA) of uninformed hedgers is reasonable (e.g., 2) and there is even a slight positive correlation (e.g., 0.01) among their endowment shocks. In contrast, a model in which uninformed trading arises only from hedging and not overconfidence can only fit the data reasonably well by resorting to RRA values that are too high by at least an order of magnitude.

We conduct further analyses and controlled experiments using the main model that allows for overconfidence. Interestingly, overconfident speculators account for just 12% of trading volume in this model, as compared to over 80% from fully rational informed agents. However, holding all other estimated parameters fixed, we find that predicted volume drops by over 99% if uninformed traders exhibit no overconfidence. Put another way, even though rational agents’ with heterogeneous beliefs account for most trading, such trading cannot by itself explain why volume is high. The main reason is that all traders in our model act to optimize their expected utility, including uninformed traders. This feature implies that market breakdowns with little or no trading occur when adverse selection is sufficiently severe. At plausible risk aversion values, there is not
enough hedging to overcome the adverse selection problem, so overconfident speculation is necessary for the market to remain liquid. In this sense, overconfident traders provide a societal benefit. In fact, stock price discovery in the market with overconfident agents is extremely efficient: prices incorporate 91% of private information in the same period that it is collected by rational agents. In sum, our findings demonstrate that overconfidence can play a critical role in explaining trading volume and market liquidity without distorting market prices, and that it may actually improve the information content of prices.

The remainder of the paper is organized as follows. Sections 1 and 2 describe our model of market activity and identification strategy. Section 3 describes the data. Sections 4 and 5 summarize and analyze our results. Section 6 concludes.

1. Model of Market Activity

We consider a model of a single risky asset (hereafter “stock”), which features rational informed traders along with uninformed traders who have hedging motives and may also be overconfident. The model below borrows key elements from the models in Spiegel and Subrahmanyam (1992), Kyle and Wang (1997), and Scheinkman and Xiong (2003). In our empirical work, we apply this model of a single stock to all stocks under the simplifying assumption that market activity is independent across stocks after we control for systematic factors, such as market returns.

The value of the stock at some (distant) future trading round $T$ is given by

$$F = \bar{F} + \sum_{t=1}^{T} d_t,$$

---

2 Because there is competition among rational risk-neutral market makers in our model, market prices are always *semi-strong* efficient, meaning that they fully incorporate public information (Fama (1970)). The 91% estimate above is one way to quantify the degree to which the *strong* form of the efficient market hypothesis holds.
where \( F \) is the initial value of the stock at time 0 and the \( d_t \) terms are independently normally distributed with means of zero and variances of \( \sigma_{dt} \). In each trading round \( t \), public information reveals the previous value of \( d_{t-1} \), traders choose whether to enter the market and submit orders based on private signals, and a market maker sets the equilibrium price, which is the final value for the period.

Both risk-neutral informed agents and risk-averse uninformed agents are drawn from homogeneous populations. Each maximizes expected utility defined over next period’s profits. Traders’ entry decisions endogenously determine the amounts of informed and uninformed traders in each period. Informed traders are strategic and large, meaning that each of the \( m_I \) participating informed traders takes into account his/her impact on prices. For tractability, we assume that each uninformed trader is small and behaves competitively, taking the market price as given. The mass of uninformed traders who participate is \( m_H \). Allowing for endogenous entry and assuming short-horizons may be a reasonable approximation in a model that is designed to explain trading volume.\(^3\) Although there is effectively only one trading round in the model, time variation can be rich because all traders endogenously choose whether to participate in each period.

In trading round \( t \), informed trader \( i \) can observe a noisy signal \( (s_{it}) \) of \( d_t \):

\[
s_{it} = \gamma d_t + \sqrt{1 - \gamma^2} z_{it},
\]

\[
z_{it} = \rho z_i + \sqrt{1 - \rho^2} \delta_{it},
\]

where \( \delta_{it} \) are independent across traders. The signal noise terms \( z_t \) and \( \delta_t \) are independently normally distributed with means of zero and variances of \( \sigma_{dt} \). The parameter \( \eta \in [0,1] \) is the

---

\(^3\) Because informed traders are risk-neutral and their information is short-lived, informed agents behave identically whether they have single-period or infinitely long horizons. We assume that the uninformed traders are risk averse in order to motivate them to hedge risk. In addition, if they were risk-neutral \( (t \to \infty) \), uninformed traders could demand an infinite quantity of the risky asset.
signal’s informativeness, while the parameter $\rho_z$ governs the extent of disagreement among traders with $\rho_z = 0$ corresponding to maximum disagreement. The pairwise correlation between signals is

$$\rho_I = \sqrt{\eta_z^2 + (1-\eta_z^2)\rho_z^2}. \tag{4}$$

Each informed trader must pay a cost $c_I$ to learn the signal, which includes the opportunity cost of time and attention required to investigate the stock and trade.

We assume that uninformed traders ($j$) have utility functions $u_{jt} = \exp(-\frac{1}{\tau} \pi_{jt})$ with common risk tolerance of $\tau$. The combination of constant absolute risk aversion (CARA) utility and normally distributed signals and dividends implies that trader $j$’s expected utility ($U_{jt}$) depends only on the mean and variance of next period’s profits according to

$$U_{jt} = E(u_{jt} \mid I_{jt}) = E_j(\pi_{jt+1}) - \frac{1}{2\tau} Var_j(\pi_{jt+1}), \tag{5}$$

where $I_{jt}$ is trader $j$’s information at time $t$. Uninformed agents may trade to hedge endowment shocks $h_{jt}$ that follow

$$h_{jt} = \rho_h h_t + (1-\rho_h^2) \psi_{jt}, \tag{6}$$

where $h_{jt}$ are measured in shares and $\psi_{jt}$ are independent across traders. The parameter $\rho_h \in [0,1]$ allows for positive correlation among hedgers’ endowment shocks, which has two important effects. First, it allows aggregate hedging demand to have a non-trivial impact on prices. Second, insofar as the aggregate liquidity shock ($h_t$) affects prices, each hedger has an incentive to speculate on information about the aggregate shock conferred by the realization of their private endowment shock ($h_{jt}$). Similar to informed traders, each uninformed trader must pay a cost ($c_H$) to enter the market, which entails monitoring endowment shocks, processing perceived signals, and trading. Total entry costs are presumably much higher for informed traders who collect valuable information than for uninformed traders, who do not collect informative signals.
We also allow uninformed agents to trade because they are overconfident in their ability to forecast the firm’s dividends. Such traders believe that they observe an informative signal \( s_{jt}^B \) with a structure similar to the \( s_{jt} \) signals such that

\[
s_{jt}^B = \eta_B d_t + \sqrt{1-\eta_B^2} z_{jt}^B,
\]

where the \( B \) sub- or superscript denotes a variable specific to biased agents and \( \eta_B \in [0,1] \) represents perceived signal informativeness. They believe the other (error) component of their signals follows

\[
z_{jt}^B = \rho_{Bu} z_t^B + \sqrt{1-\rho_{Bu}^2} \phi_{jt},
\]

where \( z_t^B \) is independent of \( z_t \) and the \( \phi_{jt} \) are independent across traders. The parameter \( \rho_{Bu} \) governs the perceived correlation among uninformed traders’ signal errors, which we assume to be equal to the correlation among informed traders’ signal errors (i.e., \( \rho_{Bu} = \rho_2 \)). Uninformed traders perceive the pairwise correlation between their signals to be

\[
\rho_B = \sqrt{\eta_B^2 + (1-\eta_B^2)\rho_{Bu}^2}.
\]

In contrast, the true signal process for overconfident agents, which is understood by both informed agents and market makers, is simply

\[
s_{jt} = \rho_a z_t^B + \sqrt{1-\rho_a^2} \phi_{jt}.
\]

The parameter \( \rho_a \) determines the actual correlation among uninformed traders’ signals and does not depend on perceived signal informativeness. Uninformed traders’ signals in Equation (10) are not actually correlated with dividends, even though the traders may believe that \( \eta_B > 0 \). Because overconfidence in our model is based on misplaced faith in the accuracy of a common signal, one
can think of it as trading on common investor sentiment. In this interpretation, overconfident traders follow similar spurious signals, such as investment newsletters, technical trading rules, or flawed valuation methods.

Overconfident agents have correct beliefs about the number of other traders and their strategies and maximize their subjective expected utility accordingly. This formulation follows Scheinkman and Xiong (2003) and Alti and Tetlock (2012) and is similar to that modeled in Kyle and Wang (1997), Odean (1998), and Daniel, Hirshleifer, Subrahmanyam (1998). Aside from incorrectly perceiving their signals, overconfident traders behave rationally. Importantly, this model nests the special case in which uninformed agents are fully rational and recognize they have no special information ($\eta_B = 0$). In this case, uniformed agents’ beliefs about the error term become irrelevant because they all ignore their signals in equilibrium.

As in Kyle (1985), we focus on symmetric linear equilibria in which informed and uninformed traders submit utility-maximizing market orders ($x_{it}$ and $y_{jt}$, respectively) that are linear in their private information. All traders within each group use the same strategies, and the competitive market maker uses a pricing rule that depends linearly with a slope of $\lambda_t$ on the aggregate net order flow ($Q_t$), where

$$Q_t = \sum_{i=1}^{m_i} x_{it} + \int_0^{m_j} y_{jt} dj. \quad (11)$$

Recall that informed trader $i$ observes $I_{it} = \{s_{it}\}$, while uninformed trader $j$ observes $I_{jt} = \{s_{jt}^B, h_{jt}\}$. We therefore denote the aggressiveness of informed trading on $s_{it}$ by $\beta_1$ and the aggressiveness of uninformed trading based on $s_{jt}^B$ and $h_{jt}$ by $\beta_2$ and $\beta_3$.

The market clears at the price in which there is no net demand or supply for the stock, resulting in a price that is linear in traders’ information. We assume that any buy-sell imbalance is

---

4 This use of the term sentiment is consistent with the model by DeLong, Shleifer, Summers, and Waldmann (1990).
accommodated by a rational risk-neutral market maker, which ensures that the market clearing
price is an unbiased estimator of the firm’s value \( E(F|p) = p \). In other words, there are no limits to
arbitrage in the model, so the price is semi-strong efficient. The absence of mispricing allows us to
focus on the model’s predictions for variance, volume, and liquidity.\(^5\)

We solve the model by conjecturing an equilibrium and evaluating whether agents have an
incentive to deviate using backward induction. In Appendix A, we characterize the unique
symmetric equilibrium in which informed traders, uninformed traders, and market makers follow
the linear strategies above and both informed and uninformed traders endogenously choose to
participate in the market. The key endogenous parameters are the equilibrium sensitivity of prices
to order flow \( \lambda_t \), informed and uninformed trader aggressiveness \( \beta_1, \beta_2, \text{ and } \beta_3 \), and the number
of informed \( m_i \) and uninformed traders \( m_{hi} \). The Appendix provides closed form expressions for
these parameters.

2. Identification

A. Model Predictions

We identify the model parameters by their impacts on the model’s predictions of return
variance, expected trading volume, and market liquidity. Adopting the standard definitions of
stock returns \( r_t \) and share volume \( v_t \) used in studies such as Admati and Pfleiderer (1988),

\[
r_t = p_t - p_{t-1}
\]

\[
v_t = \max(\sum \text{buys}_t, \sum \text{sells}_t),
\]

we compute return variance and expected trading volume as

\[
Var(r_t) = Var(\lambda_t Q_t) + Var(d_{t-1} - \lambda_{t-1} Q_{t-1})
\]

\(^5\) Alti and Tetlock (2012) analyze the impact of overconfidence on asset prices and firm investment behavior, but they
do not consider trading volume or market liquidity.
Equations (A48) and (A49) in the Appendix provide detailed closed-form expressions for these two moments in terms of primitive model parameters.

Our definition of illiquidity captures the idea that an illiquid market is one in which stock returns are highly sensitive to trading volume. To avoid the difficulty of measuring signed (i.e., buyer- or seller-initiated) trading volume,, we define illiquidity in the spirit of Amihud (2002) as the regression coefficient of the absolute value of stock returns on unsigned trading volume:

\[
\frac{\text{Cov}(\mid r_i \mid, v_j)}{\text{Var}(v_j)} = \frac{\text{Cov}(\mid r_i \mid, \frac{1}{2}(\sum \mid x_{it} \mid + \sum \mid y_{jt} \mid dj + \mid Q_t \mid))}{\text{Var}(\frac{1}{2}(\sum \mid x_{it} \mid + \sum \mid y_{jt} \mid dj + \mid Q_t \mid))}
\]  

The ratio of Equation (A50) to (A51) in the Appendix provides a closed-form expression for this illiquidity coefficient. In simulations based on a wide range of parameter values, the model’s predicted illiquidity coefficient is very close to the model's predicted value for \( \lambda \), the slope of the market maker's pricing schedule.

To facilitate comparisons of empirical moments and parameter estimates across firms and time, we define scaled versions (denoted by *) of some moments and parameters in terms of each firm’s shares outstanding (\( \theta \)) and price (\( p \)):

\[
c^{\ast}_{ht} = c_{ht} / (p\theta)
\]

\[
c^{\ast}_{htc} = c_{htc} / (p\theta)
\]

\[
\tau^{\ast} = \tau / (p\theta)
\]

\[
\sigma_{dt}^{\ast} = \sigma_{dt} / p
\]

\[
\sigma_{ht}^{\ast} = \sigma_{ht} / \theta
\]

\[
\lambda^{\ast} = \lambda / p
\]
\[ r_i^* = \frac{r_i}{p} \]  
\[ v_i^* = \frac{v_i}{\theta} \]  
\[ \text{Var}(r_i^*) = \text{Var}(\frac{r_i}{p}) \]  
\[ E(v_i^*) = E(\frac{v_i}{\theta}) \]  
\[ \frac{\text{Cov}(r_i^*, v_i^*)}{\text{Var}(v_i^*)} = \frac{\text{Cov}(\frac{r_i}{p}, \frac{v_i}{\theta})}{\text{Var}(\frac{v_i}{\theta})} \]  

With these definitions, one can verify that the equations expressing the parameter estimates in terms of the empirical moments remain identical to the original equations in Appendix A, except that all variables in the new equations have star superscripts. For practical reasons, we measure market capitalization and shares outstanding as of the end of the previous period.\(^6\) Henceforth, we omit star superscripts, except when explicitly comparing raw and scaled values.

### B. Exploitable Variation

Our identification strategy exploits enormous empirical variation in variance, trading volume, and illiquidity moments across two dimensions. First, over the past decade, while virtually any investor can trade stocks at electronic venues both before and after normal market hours, the vast majority of trading still occurs between 9:30 am and 4:00 pm. Second, regardless of the intraday period, all three moments vary with the occurrence of public news. Consequently, we separately estimate variance, volume, and illiquidity moments in three intraday periods—the pre-market from 7:00am to 9:30am; the regular market from 9:30am to 4:00pm; and the after-market from 4:00pm to 6:30pm—and conditional on whether or not public news has arrived.

Some model parameters, such as the amount and precision of acquirable information \((\sigma_{dt})\),

\(^6\) Technically, this timing induces a small approximation error in the parameter estimates, but this is negligible in most cases because the gross returns for each period are usually very close to 1.0.
and $\eta$, respectively), are properties of the information environment and thus may vary. We allow both to depend on the intraday period due to variation in normal business activity. Moreover, intuition suggests that public news about the firm increases the amount of available information. At the same time, interpreting newly arriving information is often difficult, implying that the precision of information may decline. We thus allow these parameters to vary with news arrival as well. Other model parameters, such as risk tolerance and overconfidence ($\tau$ and $\eta_B$, respectively), are properties of the agents in the model and remain fixed.

C. Parameter Restrictions

Table I Panel A describes the parameters and the various restrictions placed on them in two versions of the model. In the overconfidence version, we estimate overconfidence $\eta_B$ and fix risk tolerance $\tau$ to imply a plausible value of risk aversion. In the rational version, we fix overconfidence $\eta_B = 0$ and estimate risk tolerance $\tau$. We are unable to estimate these two parameters simultaneously because they are not separately identifiable. Intuitively, an increase in $\tau$ and an increase in $\eta_B$ both primarily affect the model’s predictions by increasing the aggressiveness of uninformed traders. We argue below that one can distinguish these versions of the model by considering whether the magnitudes of the estimated parameters in each are plausible.

We determine the six remaining parameters ($\rho_z$, $\rho_h$, $\rho_u$, $\sigma_h$, $c_I$, and $c_H$) as follows. First, we set the disagreement parameter that governs the correlation in informed traders’ signal errors to be $\rho_z = 0$, which maximizes the amount of trading that occurs among rational agents. As noted earlier, we assume uninformed traders perceive a similar error structure as the informed, hence $\rho_{Bu} = \rho_z = 0$. As we will show, this $\rho_z = 0$ specification is able to fit the data well. Our (unreported) attempts to
fit the data with $\rho_z >> 0$ have been less successful.

We set the magnitude of endowment shocks ($\sigma_h$) equal to a high value to give the hedging motive a reasonable chance of explaining the data. We also link the magnitude of endowment shocks to $\sigma_d$, which determines return volatility, so that $\sigma_h^* = 0.01 \sigma_d^*$.\(^7\) This restriction reflects the idea that investors are likely to experience larger endowment shocks in volatile stocks whose historical price movements create rebalancing needs. We select the magnitude of the multiple (0.01) so that the endowment shock volatility approaches the size of the portfolio transition trades in Obizhaeva’s (2009) study. She finds that such transitions average 0.03% of firm value, which translates into an endowment shock multiple of 0.01 if return volatility per day is 3%. To ensure that the choice of multiple has a minimal impact on our estimated value of relative risk aversion ($RRA$), we fix the (dollar) volatility of endowment shocks ($p\sigma_h$) at 25% of uninformed investors’ wealth ($w$). This implies that a plus or minus 4 standard deviation shock would either double or eliminate an investor’s wealth. Under these assumptions, we can convert our estimated value of scaled absolute risk tolerance into an $RRA$ coefficient as follows:

$$RRA = \frac{w}{\tau} = \frac{w}{p\sigma_h} \frac{p\sigma_h}{p\theta \tau^*} = \frac{1}{0.25} \frac{0.01\sigma_d^*}{\tau^*} = \frac{0.04\sigma_d^*}{\tau^*}$$

(28)

The estimated $RRA$ coefficient is almost invariant to the value of $\sigma_h$ within a very large range.\(^8\) A higher $\sigma_h$ value increases the estimated $\tau^*$ by approximately the same amount that it increases the investor’s wealth ($w$), which decreases $RRA$ relative to $\tau^*$. In the overconfidence model, we fix $\tau^* = 0.0005$ such that $RRA \approx 2$, rather than estimating $RRA$.

We base the actual correlation among uninformed signals ($\rho_u^2$) on empirical data.

\(^7\) Recall the (starred) versions of $\sigma_d$ and $\sigma_h$ used to estimate the model, Equations (20) and (21), respectively, are both defined in percentage terms and are therefore directly comparable.

\(^8\) When computing $RRA$ in Equation (28), we set the volatility ($\sigma_d$) input equal to the square root of the expected per-period variance of $d$, where the single-period variances used to calculate the expectation are based on our parameter estimates.
Specifically, for two uninformed traders, $j$ and $k$ in our model with signals arising from normal variables with correlation $\rho_u^2$, the probability of trading in the same direction is

$$\Pr(y_j, y_k \geq 0) = 0.5 + \frac{1}{\pi} \arcsin(\rho_u^2)$$

(29)

Dorn, Huberman, and Sengmueller (2008) report that the probability that two retail brokerage customers submit market orders in the same direction (in the same stock) is 0.538. We set Equation (29) equal to this value and solve for $\rho_u$, resulting in an estimate of 0.345.

We set the pairwise correlation $\rho_h^2$ among hedgers’ endowment shocks to be 1.0 (i.e., $\rho_h = 1.0$). When we allow for overconfidence and require that risk aversion is plausible, the model’s main predictions are almost invariant to this correlation, which we verify by re-estimating the model with an extremely low correlation of $\rho_h^2 = 0.01$ (i.e., $\rho_h = 0.1$). When there is no overconfidence and risk aversion is high, the model’s predictions do depend on whether the value of $\rho_h$ is significantly less than 1.0. We estimate and explore the rational $\rho_h = 0.1$ version of the model in the final section of the paper, though we focus on the simpler $\rho_h = 1.0$ model initially.

Lastly, we determine the cost of informed and uninformed entry as follows.

A recent Wall Street Journal story provides a direct estimate of the cost of informed trading in large firms.\(^9\) Several hedge funds paid up to $10,000 each to acquire private information about a December 8th, 2009 health care law that affected four large health care stocks. As in the model, the information was acquired during the regular market, one intraday period in advance of its release during the after-market period. Based on the cumulative market capitalization of the stocks (about $100B), the implied value of $c_i^*$ is less than $10,000 / 100B = 1.0 \times 10^{-7}$. This estimate is an upper bound for three reasons: 1) some hedge funds paid less than $10,000; 2) the meeting may have

---

provided more precise information than investors can typically obtain; and 3) it may have provided relevant information about the entire health care sector, which has a much larger market capitalization than the four firms. We thus use a lower value of $c_I^* = 3.0 \times 10^{-8}$. Our results below are similar using values ranging between $1.0 \times 10^{-8}$ and $1.0 \times 10^{-7}$. As argued earlier, uninformed agents are likely to face much smaller entry costs ($c_H$) than informed agents do ($c_I$). We set uninformed agents’ entry costs to be $c_H = 0.01 c_I$, but our choice of the multiple (0.01) has little impact on the model’s qualitative and quantitative results within the wide range of 0.001 to 0.1.

3. **Data and Empirical Moments**

A. **Data**

Our eligible sample spans 2001 to 2010 and includes NYSE, AMEX, and NASDAQ stocks. We obtain trade-by-trade price and volume data from the NYSE TAQ database. We adjust for non-standard opening and closing times and define the regular trading period as the hours in which the market is open (typically 9:30 am to 4:00 pm ET), and the pre-market and after-market as the 2.5 hours prior to the open and following the close, respectively. We then construct firm-specific return, turnover, and illiquidity observations for each day’s pre-market, regular market, and after-market periods. Each variable is designed to mimic the corresponding theoretical moment introduced in Section 2.

As discussed in the Appendix, we employ standard microstructure techniques to compute accurate trade-based returns. In addition, we adjust all firm returns for market returns by subtracting the contemporaneous intraday return of the SPDR S&P 500 ETF (SPY). When a firm’s

---

10 When we estimate $c_I^*$ jointly with the other model parameters, we obtain somewhat weaker identification, but the $c_I^*$ estimates typically fall within the range of values above.

11 Barclay and Hendershott (2008) argue that extended hours trades and quotes are adequately represented in the TAQ database. See their section 2.1 for a detailed analysis.
return is zero during an extended hours period, we set its market-adjusted return to zero to avoid introducing additional microstructure noise. We compute share turnover as the market value of share volume scaled by market cap. Finally, we compute illiquidity as the regression coefficient of absolute one-minute VWAP returns on contemporaneous turnover.12

We measure firm-specific news using the Dow Jones archive to distinguish between periods with and without news arrival. These data include all DJ newswire and Wall Street Journal stories from 2001 to 2010. For each story, DJ provides stock codes indicating which firms are meaningfully mentioned and a timestamp indicating when the story became publicly accessible. We ensure that we focus on firm-specific news by excluding stories that mention more than two listed U.S. stocks. The variable Newsit equals one when news stories mention firm i during intraday period t and zero otherwise. For small and mid cap (large cap) firms, we require at least one (two) story mention(s) to constitute news. This convention does not count isolated stories for large cap stocks as news because large firms frequently receive news coverage even when no new public information exists.

Motivated by the modern day 24-hour news cycle, we consider whether news has arrived in the past 24 hours. For example, sequences of related news stories about a firm may unfold during a 24-hour period, bringing the news and the stock to the attention of more traders and altering the information gathering process. In our estimation below, we keep only intraday periods in which no news has occurred since the same intraday period in the prior trading day. Imposing a brief period of non-news prior to measuring market activity mitigates the confounding impacts of recently released news stories and better establishes the release of new public information. Thus, we estimate moments conditional on news using observations where Newsit = 1 and no news occurred.

12 In the illiquidity regressions, we drop absolute one-minute returns greater than five percent and only estimate the model for regular (extended hours) periods with at least 50 (20) VWAP returns. We restrict the intercepts for each extended hours regression to equal the average regular period intercept for the same size-group and quarter.
since the same intraday period from the prior trading day. Likewise, we estimate moments conditional on no-news using only observations where $News_{t} = 0$ and no news has occurred since the same intraday period from the prior trading day.

We employ additional sample filters based on size, news coverage, and extended hours trading activity from the prior calendar year. First, we retain only stocks having market capitalizations above $100$ million and share prices greater than $1$ at the end of the prior year. Second, we require a firm to have a news story in a minimum of four pre-market periods and four after-market periods. Third, firms must also have trading in at least 20 pre-market periods and 20 after-market periods. Finally, we divide the firms into three size subsamples based on market capitalization from the prior year-end. We define “large cap,” “mid cap,” and “small cap” stocks as those with market capitalizations within the intervals $[$10B, $\infty)$, $[$1B,$10B)$, and $[$100M,$1B)$, respectively. These size groups contain an average of 95, 245, and 237 firms per year, respectively.

B. Empirical Moments

In each intraday period (pre-market, regular hours, and after-market) and quarter, we pool firm-specific observations across similarly-sized firms. We then estimate pooled variance, volume, and illiquidity moments conditional on the occurrence of either “news” or “no-news”. In all computations, we weight intraday observations such that each firm is given equal representation within a size group and quarter. For our full sample analysis, we average each moment estimate across size groups within each quarter and then average across quarters to obtain a set of eighteen conditional moments. This procedure mitigates measurement error resulting from firm-level estimates and assigns each size group and quarter equal weight. Moreover, we use the
joint quarterly time series of moment estimates to estimate the weighting matrix in the efficient
generalized method of moments procedure described below.

Figure 1 plots news probabilities for each quarter in the full sample. These values indicate the probability of news arrival conditional on no news releases since the same intraday period of the past trading day. Four general patterns are noteworthy. First, while the regular period has the highest probability of news, the majority of news stories occur in the extended hours periods—the probability of news in a regular period is about 0.10, while the probabilities of news in the pre-market and after-market periods are 0.07 and 0.05, respectively. Furthermore, measured as an hourly arrival rate, the probability of news is actually highest during the pre-market period. Second, the probability of news arrival in every intraday period increases substantially around 2003 and plateaus through 2010. Third, there is some seasonality for regular market news in which more stories occur during the first quarter of the year. This is possibly related to annual reports of the disproportionate number of firms with December fiscal year ends. Fourth (not shown), in all periods, news occurs more frequently for large cap stocks than for mid or small caps.

[Insert Figure 1 here.]

The first column in Table II presents full-sample estimates of conditional variance, volume, and illiquidity moments, and Figure 2 highlights how each varies with news. We begin our discussion with the variance. To facilitate comparison across periods, we scale up extended hours variances by 6.5/2.5 (the ratio of the lengths of the regular period to an extended hours period) and in all cases take the square root to report volatilities. In each variance calculation, we account for spurious reversal due to transitory noise in the period \( t \) price by computing \( Var(r_t)^* = Var(r_t) + Cov(r_t, r_{t-1}) + Cov(r_t, r_{t+1}). \) Interestingly, this adjustment affects estimates by less than 15% in each intraday period, implying that microstructure noise is not too severe.
Our estimates reveal that news is consistently associated with higher variance—variance conditional on news always exceeds that conditional on no-news, and the difference is economically staggering. Figure 2 shows that in the regular period, volatility on news days is about 130% of that on no-news days. In the pre-market and after-market periods, this ratio is even higher at 302% and 370%, respectively. Qualitatively similar patterns emerge within each size group (not shown). Second, volatility during extended trading hours is of a similar magnitude as that during regular trading, especially when comparing periods with news. This finding is surprising in light of earlier evidence from French and Roll (1986) that volatility when the regular market is open far exceeds that when the regular market is closed. Our results may differ from theirs because at least some trading now occurs in extended hours markets even though traditional markets are closed and because new technologies have changed the nature of trading and news dissemination since their time period.

Table II also presents the conditional volume moments for the pre-, regular, and after-market periods. Extended hours moments are again scaled up by a factor of 6.5/2.5, and numbers in the table are turnover expressed in percent. Similar to the conditional variance results, conditioning on news matters. As illustrated in Figure 2, turnover with news in the regular, pre-, and after-market periods, respectively, is 145%, 723%, and 688% of that without news. However, unlike the variance patterns, almost all trading takes place during regular market hours irrespective of the occurrence of news.

Finally, Table II contains conditional illiquidity moments, which are average coefficients from regressing absolute one-minute VWAP returns on turnover. Not surprisingly, extended hours
markets are far less liquid than the regular market. Illiquidity in the pre-market and after-market, respectively, is about 30 times and 15 times that of the regular market. The effect of news, however, varies across periods. The regular market and pre-market are more liquid in the presence of news, while the opposite is true for the after-market. Together with those in variance and volume, these stark patterns are key to identifying our model.

4. Structural Estimates of the Model

We use the generalized method of moments (GMM) to estimate the 13 parameters that best match the 18 empirical moments. Formally, GMM minimizes the distance between the model’s predicted moments and the empirical moments. We use an efficient GMM procedure that weights this distance using a matrix equal to the inverse of the covariance matrix of the empirical moments, so that moments that are measured with greater precision receive proportionally greater weight. We measure the covariance matrix using quarterly variation in the moments while taking persistence into account. We describe the details of this procedure in Appendix C.

We estimate both versions of the model described in Panel A in Table I. Recall the overconfidence and rational hedging models differ only in whether risk tolerance ($\tau$) or overconfidence ($\eta_B$) is held fixed. Table II shows how the models fit each empirical moment, and Figure 3 reports the $t$-statistics of prediction errors. These $t$-statistics indicate that the overconfidence model matches all 18 of the empirical moments within two standard errors and all but one moment within one standard error. One cannot statistically reject the overconfidence model based on its five overidentifying restrictions—the $\chi^2(5)$ statistics of 7.2 has a $p$-value of 0.208. The overconfidence model also matches the moments well in both subperiods. This is notable because the recent increase in high-frequency trading is likely to affect our estimates
mainly in the second subperiod. Moreover, the parameters’ small standard errors (shown in parentheses in Table III below) indicate that all 13 of the parameters are reasonably well-identified in the full sample and in both subsamples.\textsuperscript{13} Thus a lack of statistical power cannot explain why we do not reject the hypothesis that the overconfidence model fits the data.

[Insert Figure 3 here.]

Although the rational model fits the main qualitative features of the data, it fails to match the magnitudes of some moments, particularly in the extended hours periods. The $t$-statistics of the rational model’s prediction errors are less than 1.0 for 14 out of 18 moments, but they exceed 2.0 for three of the other four. As a result, one can statistically reject the rational model’s five overidentifying restrictions at the 1% level—the $\chi^2(5)$ statistic of 29.9 has a $p$-value < 0.001. This rejection is driven by the rational model’s inability to match after-market activity, which we discuss further at the end of this section.

For now, we stress that the economic magnitudes of both models’ prediction errors are tiny in comparison to the huge empirical variation in moments across periods—e.g., turnover and illiquidity often vary by orders of magnitude. Most importantly for this study, both models match the enormous trading volume during the regular market period as well as the light trading in extended hours. Both models are also able to replicate the large increases in volatility and volume that accompany the occurrence of news, especially when news arrives in extended hours. In this sense, both models can explain the main features of market activity.

[Insert Table III here.]

We scrutinize this statement further by analyzing the parameter estimates reported in Table III. Most importantly, we consider the economic plausibility of the overconfidence and risk

\textsuperscript{13} The parameters’ standard errors come from the GMM delta method formula based on the covariance matrix of the moments and the sensitivity of each moment to each parameter—e.g., see Cochrane (2001).
aversion parameters that distinguish the two models. If we can reject the hypothesis that a model’s parameter estimates are plausible, we can reject the model as an explanation of market activity even if it fits the data under implausible parameter values. Beginning with the overconfidence model, the full-sample estimates of overconfidence indicate that uninformed agents believe their signals have precision of $\eta_B = 0.188$. The estimates of $\eta_B$ are 0.164 and 0.292 in the two subperiods. Based on the standard errors, one can reject the hypothesis of no overconfidence ($\eta_B = 0$).

Importantly, in all six intraday periods, overconfident agents never exhibit beliefs in their own precision never exceeds the precision of rational agents by a statistically significant margin. There are, however, periods such as the extended hours periods in which rational agents’ precision significantly exceeds the perceived precision of overconfident agents.

We evaluate whether uninformed agents’ overconfidence is high by examining the accuracy of agents’ directional forecasts of dividends based on their signals. Because uninformed agents’ signals are not informative, their directional forecasts of dividends are correct exactly half of the time (probability 0.5). Our parameter estimates suggest uninformed agents believe that they are right 56.0% of the time.\footnote{We compute the overconfident agents’ perceived probability of being correct using Equation (29), except that we replace $y_j y_{\bar{k}}$ with $s_j^B d$ and $\rho_u \sigma^2$ with $\eta_B$.} Even in the second half of the sample when overconfidence is highest, uninformed agents only believe they are correct 59.4% of the time.

Such overconfidence is modest in relation to experimental estimates. In Lichtenstein and Fischhoff’s (1977) experiment, subjects taught how to read stock charts believe they can predict a stock’s directional price movement 65.4% of the time, whereas they are correct slightly less than half of the time. Our smaller overconfidence estimates are consistent with the notion that investors’ behavioral biases may diminish when material sums of money are at stake. In addition, investors with short horizons like those in our model presumably obtain rapid and repeated
feedback on their forecasting performance, which could mitigate their overconfidence. Nevertheless, it would take hundreds of independent trials for uninformed agents to learn that they are overconfident, even if they accurately remember their successes and failures.\footnote{The investor would require on average 274 independent trials to obtain a \( t \)-statistic of 2.0, rejecting the null hypothesis that \( p = 0.56 \).}

We now turn to the rational model. The reasonable estimates of overconfidence described above contrast dramatically with the implausibly high estimates of risk aversion (\( RRA = 132 \)) in the rational model. The subperiod \( RRA \) estimates of 141 and 112 are similar in magnitude. The standard error of the full sample estimate is sufficiently small so that we can reject the hypothesis that \( RRA \) is less than 75.

For comparison purposes, we consider the implied risk aversion of an investor with constant \( RRA \) who holds a portfolio fully invested in the US stock market. Even with the generous assumption of an equity premium of 8\% and a volatility (in log returns) of 20\% per year, this investor must have a risk aversion of \( 2 = 0.08 / (0.20)^2 \) in order to hold the market. Although traders who hedge their endowment shocks may be somewhat more risk averse than the average investor, it is difficult to imagine that they exhibit risk aversion coefficients that are over 50 times higher than average.

One should also interpret the risk aversion estimates in light of the fact that the risk faced by agents in the model affects only the returns of a single stock—i.e., it is idiosyncratic, not systematic. If uninformed investors hold well-diversified portfolios, small idiosyncratic risks should be irrelevant for their overall portfolio risk, implying they should exhibit no risk aversion at all. More generally, under the reasonable assumption that investors exhibit lower effective risk aversion in response to idiosyncratic risks, one would expect a structural estimate of \( RRA \) based on idiosyncratic risks to be lower than an estimate based on systematic risks. In this respect, the
extremely high estimated RRA in Table III is even more surprising.

Moreover, we find that the rational model cannot account for any material amount of trading activity when risk aversion is plausible. Starting from the parameter estimates of the overconfidence model (e.g., RRA = 2), lowering $\eta_B$ to 0 reduces predicted trading activity by over 99% and leads to a complete market breakdown (i.e., no trading) in extended hours periods without news. Alternatively, starting from the rational estimation parameters, decreasing RRA to 50 causes predicted volume to fall by 94% and decreasing RRA to 5 or lower causes the market to break down in all periods with or without news.

With a few exceptions, the values of the other parameter estimates are plausible and stable across the subsamples and the two models. In both models, the estimated values of the amount of acquirable information ($\sigma_d$) closely mimic the patterns in empirical volatility noted in Section 3. The models produce this outcome because market prices incorporate most private information in the period in which it is acquired, as we show in the analysis section that follows. Both models match the evidence that news increases return volatility because the parameter estimates indicate that news increases the amount of acquirable information, especially in extended hours periods.

In both models, the estimates of information precision ($\eta$) are much lower in the regular market than in the two extended hours markets, which is critical in order to match the stark volume differences across periods. Trading volume decreases with $\eta$ in both models, but for different reasons. In the overconfidence model, increases in information precision deter overconfident speculators because they believe that prices already incorporate most of the relevant information. In the rational hedging model, increases in precision exacerbate adverse selection in pricing, which discourages hedging activity.

One notable difference across models, which partly explains the statistical rejection of the
rational version, is that information precision is higher (lower) when news occurs in the rational (overconfidence) model. This is driven by our parameterization in which increases in acquirable information are associated with larger endowment shocks ($\sigma_h = 0.01\sigma_d$), as explained in the identification subsection. In the after-market, the increase in endowment shocks is particularly large because news is associated with high volatility. These large endowment shocks cause predicted after-market liquidity to be higher than the empirical value of after-market liquidity.

5. Analysis of the Model

The model also allows one to analyze the relative contributions of informed and uninformed traders to market activity. We decompose trading volume as in Equation (15) by separately considering the quantities of buy and sell orders that transact between traders and the net order flow ($Q$) where market makers take the other side of the trade. We assign half the volume from a trade to each participating counterparty.

We denote the volume arising from informed traders, uninformed traders, and market makers by $v_I$, $v_H$, and $v_M$. Expected volume attributable to the market maker is half of the volume arising from expected net order flow aggregated across both trader groups, which is

$$E(v_M) = \frac{1}{4} E(|Q|).$$

One can easily compute closed-form solutions for this volume expression and all those below by adjusting the terms in Equation (A49) appropriately. The trading volume attributable to each group of traders is half of the sum of their buy orders and sell orders plus half of their proportion of the net trading with market makers. We apportion the volume arising from net order flow to each trader based on the fraction of variance in net order flow arising from the trader, as measured in Equation (A13). Because the three components of volume equal total volume, we can express each
component as a fraction of the total using

\[
\frac{E(v_i)}{E(v)} + \frac{E(v_H)}{E(v)} + \frac{E(v_M)}{E(v)} = 1.
\] (31)

The volume decompositions reported below use this scaling.

Table IV reports volume analyses and other analyses of market activity based on the parameter estimates in different versions of the model. The three versions include the overconfidence and rational models with the parameter estimates shown in Table III and with the correlation among hedgers’ endowment shocks (\(\rho_h^2\)) assumed to be 1.0. The third version is a newly estimated rational model (i.e., \(\eta_B = 0\)) in which the pairwise correlation among endowment shocks is assumed to be the tiny value of \(\rho_h^2 = 0.01\). The motivation for this model is that hedgers may be more inclined to hedge if their individual endowment shocks are less correlated with the aggregate endowment shock, which adversely affects the price at which they can trade. Indeed, we find that hedgers enter in much larger amounts in the \(\rho_h^2 = 0.01\) model, which leads to a lower estimate of risk aversion but much higher estimates of the magnitude of endowment shocks—as discussed further below.

The volume decompositions indicate that informed trading accounts for the vast majority of trading in all three models. In the behavioral model, overconfidence accounts for just 12% of trading. One plausible explanation for this trading is that a small fraction of trading by mutual fund, pension fund, and hedge fund managers represents overconfident speculation (with a low belief in precision of \(\eta_B = 0.188\)). In addition, much retail trading could arise from overconfident speculation.

In the rational models, hedging accounts for just 4.7% of trading when \(\rho_h^2 = 1.0\), but it accounts for 32.5% of volume when \(\rho_h^2 = 0.01\), as shown in Table IV. In a steady state
equilibrium, the rate at which new endowment shocks arrive must equal the rate at which they are hedged, where the latter is given by the volume attributable to hedgers (e.g., 4.7% or 32.5% of total volume). Thus, the expected magnitude of total endowment shocks must be comparable to volume from hedgers. Based on the average annual turnover of 403% in our sample, 4.7% (32.5%) of volume represents expected annual endowment shocks of 19.0% (131%) of stock market capitalization, which we report in the last row of Table IV.

We now consider whether these hedging turnover magnitudes are plausible in relation to agents’ needs to hedge firm-specific risk. In the model, each firm is held by investors whose uncertainty in wealth comes solely from stock holdings in that firm and an endowment shock that is perfectly correlated with the firm’s stock. Shocks to individuals’ income from non-stock sources could serve as empirical counterparts to endowment shocks. Heaton and Lucas (2000) argue convincingly that proprietary business income is a first-order consideration for the typical stockholder. At the individual level, they report the standard deviation of growth in this income is 65%. Although there is little evidence at the firm level, the correlation between proprietary business income and stock returns is 0.14 at the aggregate level. Heaton and Lucas (2000) also document that the value of proprietary businesses ($1.63T) is almost as large as stockholdings ($2.22T) for entrepreneurs. Using these inputs, one can roughly estimate the volatility of endowment shocks to be $0.14 \times 0.65 \times 1.63 \times 2.22 = 6.7\%$ of such investors’ stockholdings. Even if such undiversified stockholders held 100% of each firm’s stock, the endowment shock volatility and thus steady-state hedging would be just 6.7% of firm value. This aggressive 6.7% estimate is far smaller than 19.0% hedging in the $\rho^2 = 1$ model and is dwarfed by the 131% hedging in the $\rho^2 = 0.01$ model. Thus while the rational model with $\rho^2 = 0.01$ does produce a more plausible risk

---

16 Heaton and Lucas (1996) calibrate a model in which agents trade based on uninsurable labor income shocks that are uncorrelated with stock returns—shocks that we do not model. Such trading amounts to just 15% turnover annually even when there are no trading costs, which is small compared to the 400% empirical turnover.
aversion (RRA) estimate than the other rational model, this comes at the cost of predicting implausibly large endowment shocks.\footnote{One can alternatively interpret the endowment shock as a portfolio imbalance. Portfolio rebalancing between investors with differing risk aversion is also unlikely to account for annual hedging approaching 100\% of firm value.}

Following an analogous procedure to the volume decomposition, one can decompose return variance into the two components shown in Equation (13), the first based on trading on private information \( Var_T(r) \) and the second based on public information \( Var_P(r) \)

\[
\frac{Var_T(r)}{Var(r)} + \frac{Var_P(r)}{Var(r)} = 1. \tag{32}
\]

The first term represents price discovery arising from informed trading on newly acquired private information, while the second term measures the revelation of tractable information that the previous period’s price did not fully reveal. This second term can be viewed as public information that is revealed through various sources, including newswires, social media, television, and radio.

Table IV reveals that market prices are extremely efficient in the overconfidence model, even relative to the two rational models. A remarkably high percentage (90.8\%) of potentially acquirable \textit{private} information is revealed in prices in the same period in which it could be acquired. In this sense, the strong-form of the efficient market hypothesis nearly holds in the behavioral model. Intuitively, overconfidence increases market liquidity because the rational market maker knows that she sometimes transacts with overconfident agents. The increased liquidity motivates rational informed agents to trade more aggressively on their information, which is then incorporated in prices. Uninformed overconfident agents trade more aggressively than uninformed hedgers all else equal because they believe that they are taking advantage of the market maker, who fails to realize that they possess information. This is why liquidity and price informativeness is lower in the hedging models.
Finally, we analyze trader profits and welfare. Informed traders do not receive any expected surplus net of costs because they have homogeneous costs and there is free entry—i.e., perfect competition. However, informed traders still earn positive gross profits excluding entry costs. Because they earn zero net profits, their gross trading profits are equal to their entry costs

\[ \pi(\text{total}) = m_p c_p, \]  

where the above equation aggregates across all informed traders entering the market. We compute the expected daily entry cost by multiplying the news and non-news per-period costs by the probability of either news or non-news occurring and adding across periods.

The total trading profits of the informed are equal to the total trading losses of uninformed traders. Depending on whether they are overconfident, the uninformed traders may incur average losses that exceed their expected losses. We evaluate the welfare of the uninformed based on their average realized utilities, not their expected utilities. Because uninformed agents are homogeneous and there is free entry, the expected utility of the uninformed investors who enter is equal to the expected utility of those who do not participate. The average realized utility of the uninformed is equal to their expected utility if and only if they are not overconfident. Thus, we can compute expected utility under beliefs in which \( \eta_B = 0 \) to evaluate average realized utility. The difference between realized and expected utility depends on the magnitude of unanticipated trading losses and unanticipated volatility in the investor’s stock portfolio. We focus on trading losses because they are less dependent on assumptions about risk aversion.

In all three models, the cost of informed trading is between 30 and 40 bps of firm value per year. This is about half of the 67 bps magnitude of the cost of active management estimated by French (2008). Multiplying our 35 bps percentage estimate by the US equity market capitalization
of $15T suggests that aggregate trading gains for informed traders and losses for uninformed traders are roughly $53B per year. We do not take a stand on whether $53B per year is a fair price to pay for price discovery and liquid capital markets. There may be other costs related to imperfect competition in asset management that lie beyond the scope of our model.

6. Conclusions

We propose and estimate a model in which market activity arises endogenously from the interaction between utility-maximizing economic agents, including uninformed traders who act as either hedgers, overconfident speculators, or both. We find that a model with overconfident speculators fits the data reasonably well even when the magnitude of overconfidence is modest. In this ostensibly behavioral model, most trading volume arises from rational informed agents and market prices are remarkably efficient. Without overconfidence, nearly all trading volume would dissipate and prices would incorporate a negligible amount of information.

Although we are unable to identify a model without overconfidence that fits the data with reasonable parameter values, our results point to several promising directions for future research. In our estimations, models of uninformed trading based on rational hedging require either implausibly high risk aversion or implausibly large endowment shocks. Specifically, the models in which endowment shocks have a low correlation across traders require very large endowment shocks. In unreported tests, we pursue such models further and find that raising the cost of entry for uninformed agents can significantly mitigate the size of endowment shocks necessary to explain trading volume. This approach still suffers from the fundamental criticism that uninformed agents must hold extremely large undiversified holdings in the firm. This is not only inefficient from a risk-sharing standpoint, but it is also unrealistic in light of the growth of mutual funds and
Abandoning the rational model altogether may provide a more satisfactory explanation of volume in the absence of behavioral biases. One could reinterpret the high risk aversion estimates here as hinting that uninformed trading may be driven by agents who face tight constraints on their behavior. Institutional money managers who manage others’ money could act as though their risk aversion is extremely high in situations where they experience large and unanticipated client withdrawals. Modeling this agency problem in conjunction with client endowment shocks could provide valuable new insights into market activity.

Another promising direction for future research would be integrating behavioral models designed to explain trading volume and those focused on mispricing. Our study suggests that a modest amount of overconfidence can explain trading volume, while other studies such as Alti and Tetlock (2012) argue that overconfidence can help explain mispricing. To distinguish among behavioral and rational models, we use empirical moments such as liquidity and volume, while Alti and Tetlock (2012) use empirical patterns in stock returns. Models with heterogeneous agents—rational or behavioral—often make sharp predictions about both volume and return predictability. Using both types of empirical moments to identify the parameters in such models would enable researchers to construct more powerful tests of rational and behavioral theories.
Appendix A – Solving for the Model’s Equilibrium Moment Predictions

Here we solve for the symmetric linear equilibrium and the endogenous outcomes in the model introduced in Section 2. We suppress variables’ time ($t$) subscripts in all equations, except those with ambiguous timing.

We impose market clearing to examine the market maker’s equilibrium pricing function. The aggregate order flow $Q$ is

$$Q = \sum_{i=1}^{m_i} x_i + \int_0^{m_j} y_j dj$$

$$= \beta_1 S + \beta_2 S_B + \beta_3 H,$$  \hspace{1cm} (A1)

where the aggregate signal and endowment shock realizations $S$, $S_B$, and $H$ are

$$S = \sum s_i = m_i [\eta d + \sqrt{1 - \eta^2} \rho_z z] + \sqrt{1 - \eta^2} \sqrt{1 - \rho_z^2} \sum \delta_i$$ \hspace{1cm} (A2)

$$S_B = \int s_j dj = m_H \rho_a z^B$$ \hspace{1cm} (A3)

$$S_{BB} = \int s_j^B dj = m_H \left[ \eta H d + \sqrt{1 - \eta^2} \rho_{BB} z^B \right]$$ \hspace{1cm} (A4)

$$H = \int h_j dj = m_H \rho_a h.$$ \hspace{1cm} (A5)

The linear pricing function with market depth $\lambda_t$ is

$$p_t - F_t = \lambda Q_t$$

$$= \lambda [\beta_1 S + \beta_2 S_B + \beta_3 H],$$ \hspace{1cm} (A6)

where the market maker’s expectation of fundamental value is

$$F_t = E(F | Q_{t-1}, \ldots, d_{t-1}, d_{t-2}, \ldots)$$

$$= F + \sum_{s=1}^{t-1} d_s.$$ \hspace{1cm} (A7)

The zero profit condition for the market maker is:

---

19 When aggregating across uninformed traders’ signals and endowment shocks, we use the approximation that the trader-specific signal and endowment shock components are negligible, which is consistent with our assumption that each uninformed trader is small.
\[ 0 = E[Q_i(d_i - (p_i - F_i))] \] (A8)

which requires that the slope of the market maker’s linear price schedule is

\[
\lambda_i = \frac{\text{Cov}(d_i, Q_i)}{\text{Var}(Q_i)} = \frac{\beta_i m_i \eta \sigma^2_d}{\beta_i^2 \sigma_{ss} + \beta_S^2 \sigma_{SB} + \beta_H^2 \sigma_{HH}}
\] (A9)

where

\[
\sigma_{ss} = \text{Var}(S) = m_i[1 + (m_i - 1)\rho^2_i] \sigma^2_d
\] (A10)

\[
\sigma_{SB} = \text{Var}(S_B) = m_h^2 \rho_h^2 \sigma^2_d
\] (A11)

\[
\sigma_{HH} = \text{Var}(H) = m_H^2 \rho_H^2 \sigma^2_h
\] (A12)

To simplify expressions throughout, we use the equilibrium liquidity equation

\[
\lambda^2 \text{Var}Q = \lambda^2[\beta^2 \sigma_{ss} + \beta_S^2 \sigma_{SB} + \beta_H^2 \sigma_{HH}] = \lambda \beta m_i \eta \sigma^2_d.
\] (A13)

We characterize traders’ beliefs about future dividends. For convenience, we denote the market price that would prevail in the absence of trader \( i \)’s demand as

\[
p_{-i} = p - \lambda x_i.
\] (A14)

We first summarize some properties of informed and uninformed traders’ beliefs

\[
\sigma_{iS} = \text{Cov}(s_i, S - s_i) = (m_i - 1)[\eta^2 + (1 - \eta^2) \rho^2_i] \sigma^2_d
\] (A15)

\[
\sigma_{jS} = \text{Cov}(s_j, S) = m_i \eta \sigma^2_d
\] (A16)

\[
\sigma_{dS} = \text{Cov}(d, S) = m_i \eta \sigma^2_d
\] (A17)

\[
\sigma_{jSB} = \text{Cov}(s_j, S_{BB}) = m_H \rho_h^2 \sigma^2_d
\] (A18)

\[
\sigma_{jHH} = \text{Cov}(h_j, H) = m_H \rho_h^2 \sigma^2_h
\] (A19)

We also define uninformed traders’ perceived variance as
\[ \gamma_H = \text{Var}_h(d - \lambda Q | s_j^H, h_j) \]
\[ = (1-\eta_B^2)\sigma_d^2(1-f_1m_j\eta - f_2m_H\eta_B)^2 + f_1^2(1-\eta^2)\left[ m_j^2\rho_z^2 + (1-\rho_z^2)m_j \right] \sigma_d^2 \]
\[ + f_2^2m_H^2(1-\eta_B^2)\rho_{hb}^2 \left[ 1-(1-\eta_B^2)\rho_{hb}^2 \right] \sigma_d^2 + f_3^2m_H^2\rho_h^2(1-\rho_h^2)\sigma_h^2, \tag{A20} \]

with \( f_k = \lambda \beta_k \) for \( k = 1, 2, 3 \).

Rational informed traders choose demand to maximize their expected utility (i.e., profit):
\[ \max_{x_i} U_i = E[d - (p_i - F) | s_i] x_i - \lambda x_i^2 \tag{A21} \]
\[ x_i = \frac{E[d] - (p - F)}{2\lambda} = \frac{(\eta - \lambda \beta_i \sigma_{as} \sigma_d^2) s_i}{2\lambda} \tag{A22} \]

which implies
\[ f_i = \lambda \beta_i = \eta(2 + \sigma_{as} \sigma_d^2)^{-1}. \tag{A23} \]

Uninformed traders also maximize their perceived expected utility from hedging and overconfident speculation, which is:
\[ \max_{y_j} E_B[h_jd + y_j(d - p) | h_j, s_j^H] - \frac{1}{2} \text{Var}_B[h_jd + y_j(d - p) | h_j, s_j^H] \tag{A24} \]

Maximizing this quadratic equation in \( y_j \) gives:
\[ y_j = \frac{\eta_B - \lambda \beta_i \sigma_{sj} \sigma_d^2 - \lambda \beta_2 \sigma_{sjh} \sigma_d^2 - \frac{1}{\tau} t \gamma_H}{\lambda \beta_1 \sigma_{sj} \sigma_d^2 + \frac{1}{\tau} t \gamma_H} \]
\[ - \frac{\lambda \beta_2 \sigma_{hjit} \sigma_h^2 + \frac{1}{\tau} t \gamma_H}{\lambda \beta_1 \sigma_{sj} \sigma_d^2 + \frac{1}{\tau} t \gamma_H} \tag{A25} \]

which implies
\[ \beta_2 = \frac{\eta_B - f_1 \sigma_{sj} \sigma_d^2 - f_2 \eta_2 \rho_{hb}^2}{\frac{1}{\tau} t \gamma_H} \tag{A26} \]
\[ \beta_3 = - \frac{f_1 \sigma_{hjit} \sigma_h^2 + \frac{1}{\tau} t \gamma_H}{\frac{1}{\tau} t \gamma_H} \tag{A27} \]

This allows one to compute \( f_3 \) in terms of \( f_2 \)
Now we have four equations and four unknowns that one can use to solve for market liquidity and informed trader aggressiveness. We substitute (A23) and (A28) into the equilibrium liquidity condition (A13) and write

\[ g_2^2 \rho_u^2 \sigma_d^2 \left[ \eta_B - f_1 \sigma_{sys} \sigma_d^2 - g_2 (\rho_B^2 - \rho_h^2) \right]^2 + g_2^2 \rho_h^2 \sigma_h^2 \frac{\sigma_d^4}{\tau^2} (1 - \eta_B^2)^2 (1 - f_1 m_1 \eta - g_2 \eta_B)^2 = f_1^2 m_1 \sigma_d^2 \left[ \eta_B - f_1 \sigma_{sys} \sigma_d^2 - g_2 (\rho_B^2 - \rho_h^2) \right]^2 \]  

(A29)

with \( g_k = m_{1g} \delta_k \) for \( k = 2, 3 \). This equation is quartic in \( g_2 \) with coefficients

\[ k_0 = -f_1^2 m_1 \sigma_d^2 (\eta_B - f_1 \sigma_{sys} \sigma_d^2)^2 \]  

(A30)

\[ k_1 = 2 f_1^2 m_1 \sigma_d^2 (\eta_B - f_1 \sigma_{sys} \sigma_d^2)(\rho_B^2 - \rho_h^2) \]  

(A31)

\[ k_2 = -f_1^2 m_1 \sigma_d^2 (\rho_B^2 - \rho_h^2)^2 + \rho_d^2 \sigma_d^2 (\eta_B - f_1 \sigma_{sys} \sigma_d^2)^2 + \rho_h^2 \sigma_h^2 \frac{\sigma_d^4}{\tau^2} (1 - \eta_B^2)^2 (1 - f_1 m_1 \eta)^2 \]  

(A32)

\[ k_3 = -2 \rho_u^2 \sigma_d^2 (\eta_B - f_1 \sigma_{sys} \sigma_d^2)(\rho_B^2 - \rho_h^2) - 2 \rho_h^2 \sigma_h^2 \frac{\sigma_d^4}{\tau^2} (1 - \eta_B^2)^2 (1 - f_1 m_1 \eta) \eta_B \]  

(A33)

\[ k_4 = \rho_u^2 \sigma_d^2 (\rho_B^2 - \rho_h^2)^2 + \rho_h^2 \sigma_h^2 \frac{\sigma_d^4}{\tau^2} (1 - \eta_B^2)^2 \eta_B^2 \]  

(A34)

At empirically plausible parameter values, we find that there is only one positive root for \( g_2 \) satisfying traders’ first- and second-order conditions. The resulting \( g_2 \) value allows us to solve for \( f_2 \) and thus \( f_3 \).

We solve for the equilibrium strategies by substituting the \( f_1, f_2, \) and \( f_3 \) functions into \( \beta_2 \) and \( \beta_3 \) to obtain

\[ \beta_2 = \frac{\tau (\eta_B - f_1 \sigma_{sys} \sigma_d^2 - g_2 \rho_B^2)}{\gamma H} \]  

(A35)
The solution for $\lambda$ is thus

$$\lambda = \frac{f_3}{\beta_3} = \frac{g_3}{m_H \gamma_H \frac{\tau g_3 \rho_h^2 + (1-\eta_B^2)(1-f_1 m_1 \eta - g_2 \eta_B) \sigma_d^2}{\sigma_u \sigma_d^2}}$$  \hspace{1cm} (A37)

This implies that equilibrium aggressiveness of informed traders is

$$\beta_i = \lambda^{-1} \eta (2 + \sigma_u \sigma_d^2)^{-1} \frac{m_H \eta \gamma_H \frac{\tau g_3 \rho_h^2 + (1-\eta_B^2)(1-f_1 m_1 \eta - g_2 \eta_B) \sigma_d^2}{\sigma_u \sigma_d^2}}{g_3 (2 + \sigma_u \sigma_d^2)}.$$  \hspace{1cm} (A38)

The non-existence of equilibrium can occur in certain parameter ranges. One such range corresponds to the no-trade equilibrium described in Spiegel and Subrahmanyam (1992) and arises when there is exactly no overconfidence ($\eta_B = 0$). The correlation among uninformed traders' orders can become so high that they would prefer to speculate, rather than hedge, based on their information about the aggregate liquidity shock. Specifically, there is no equilibrium with positive trading if $\eta_B = 0$ and the following relationship holds

$$\left(f^{-1} - m_1 \eta\right) \tilde{\sigma}_d < \frac{\tau \tilde{\sigma}_h}{\sqrt{m_1}} \frac{\tilde{\sigma}_u \tilde{\sigma}_d}{\tilde{\sigma}_h}.$$

Liquidity is less likely to dry up if informed traders’ information precision is lower, if there is higher correlation among their signal errors, or if uninformed traders are more risk-averse, or if they experience larger endowment shocks.

If the numbers of traders are exogenously given, there is a second range of parameter values that leads to no trading. Specifically, it is possible that overconfident agents (i.e., $\eta_B > 0$) become unwilling to trade because they believe that other overconfident agents have already
traded so much that their information is fully incorporated in prices. Formally, this occurs when

\[ 1 - f_i m_I \eta - g_i \eta_B \leq 0. \quad (A40) \]

One can show that this can only happen if there is more than one informed trader \((m_I > 1)\).

However, if the number of traders is endogenously determined, informed traders will rationally stop entering the market before this type of market breakdown ever occurs.\(^{20}\) Lastly, we note that there must be at least one informed trader must participate for liquidity \((\lambda^{-I})\) to be finite. We now turn to endogenizing information acquisition with the above parameter ranges in mind.

The two trader types simultaneously choose whether to enter the market, which endogenously determines \(m_I\) and \(m_H\) in equilibrium. We substitute agents’ optimal order choices into their utility functions to obtain their equilibrium utilities. The entry condition for informed traders is that their trading profits net of costs \((c_I)\) must be zero, which simplifies to

\[
\max U_i = E[d - (p_i - F) \mid s_i] x_i - \lambda E[x_i^2] - c_i = 0
\]

\[
= \lambda E[x_i^2] - c_i
\]

\[
= \lambda^{-1} f_i^2 \sigma_i^2 - c_i
\]

\[
= m_H (\beta_B / g_3) f_i^2 \sigma_i^2 - c_i
\]

\[
m_H = c_I g_4(m_I),
\]

where

\[
g_4 = \frac{g_3}{f_i^2 \sigma_i^2 \beta_B}. \quad (A42)
\]

Then we set the equilibrium utility of uninformed traders net of portfolio monitoring, participation costs, and information acquisition costs \((c_H)\) equal to the utility from not participating, which results in the following equation

\[^{20}\text{Even with an exogenously given number of traders, one could avoid this no trading equilibrium by assuming that each overconfident agent incorrectly believes that his or her signal is uncorrelated with the trading of other overconfident agents.}\]
\[
E_h U_h = \frac{1}{2} \left( \frac{1}{\tau} \gamma_h \right) (y_j)^2 + E[h_j d] - \frac{1}{2 \tau} Var[h_j d] - c_h = E[h_j d] - \frac{1}{2 \tau} Var[h_j d] \tag{A43}
\]

where

\[
g_s(m_t) = \frac{1}{2} \left( \frac{1}{\tau} \gamma_h \right) \left[ \beta_2^2 \sigma_d^2 + \beta_3^2 \sigma_h^2 \right]. \tag{A44}
\]

Combining the two entry conditions, we obtain

\[
m_H = c_i g_4(g_5^{-1}(c_H)) \tag{A45}
\]

Both informed and uninformed speculators do not enter the market in unlimited amounts because they recognize that other agents like them have access to similar information. Thus, their benefit to speculation declines naturally as more traders like them enter and act on related signals.\(^{21}\)

The equilibrium outcomes with endogenous entry decisions constitute a subset of the possible trading game equilibrium outcomes in which the equilibrium utilities are consistent with the number of traders entering the market. Specifically, each trader type must earn a non-negative equilibrium utility if and only if at least one trader of that type enters the market. Formally, an equilibrium outcome must satisfy

\[
U_i(m_i, \tilde{m}_i, \tilde{m}_{\tilde{hi}}) = c_i \text{ and } \frac{\partial U_i(m_i, \tilde{m}_i, \tilde{m}_{\tilde{hi}})}{\partial m_i} \leq 0 \text{ if and only if } m_i > 0 \tag{A46}
\]

\[
U_{\tilde{hi}}(m_i, \tilde{m}_i, \tilde{m}_{\tilde{hi}}) = c_{\tilde{hi}} \text{ and } \frac{\partial U_{\tilde{hi}}(m_i, \tilde{m}_i, \tilde{m}_{\tilde{hi}})}{\partial m_{\tilde{hi}}} \leq 0 \text{ if and only if } m_{\tilde{hi}} > 0 \tag{A47}
\]

The derivative restriction ensures that no trader has an incentive to deviate from equilibrium by unilaterally entering the market. We also require equilibrium illiquidity \(\lambda(m_{\tilde{hi}}, m_{\tilde{hi}})\) to be non-negative to satisfy each trader’s second-order condition. If these restrictions are met, the entry

\(^{21}\) An alternative approach to limiting entry by the uninformed would be modeling heterogeneous costs of entering the market, but this would require multiple ad hoc assumptions.
cost parameters \((c_{It}, c_{Ht})\) implement the trading game equilibrium with \((m_{It}, m_{Ht})\) traders of each type.

We now analyze the variance, volume, and liquidity implications of the model using the definitions in Equations (14), (15), and (16). First, return variance is

\[
Var(r) = Var(\lambda Q) + Var(d_{t-1} - \lambda_{t-1}Q_{t-1}) \\
= (2 + (m_{It} - 1)\rho_d^2)^{-1}m_{It}\eta_d^2\sigma_{dr}^2 \\
+ (2 + (m_{Ht} - 1)\rho_d^2)^{-1}(2 + (m_{It} - 1)\rho_d^2 - m_{It}\eta_d^2)\sigma_{dr}^2
\]

(A48)

We now compute volume

\[
E(v) = \frac{1}{2}(m_{It}E[|x_i|] + m_{Ht}E[|y_j|]) + \frac{1}{2}E[|Q|] \\
= \frac{1}{\sqrt{2\pi}}\sqrt{Var(\beta_i s_i)} + m_{It}\sqrt{Var(\beta_i s_i)} + \frac{1}{\sqrt{2\pi}}\sqrt{Var(Q)} \\
= \frac{1}{\sqrt{2\pi}}(m_{It}\beta_i\sigma_d + m_{Ht}\beta_i\sigma_d^2 + \beta_i^2\sigma_n^2) \\
+ \frac{1}{\sqrt{2\pi}}\sqrt{\beta_i^2\sigma_S^2 + \beta_d^2\sigma_{SSB} + \beta_d^2\sigma_{HH}}
\]

(A49)

Illiquidity is equal to the ratio of the covariance between absolute returns and volume to the variance of volume. The numerator and denominator are given by:

\[
Cov(\text{r} | v) = Cov(\text{r} | \frac{1}{2}(\sum |x_i| + \int |y_j| df + |Q|) \\
\frac{1}{2}\left[ m_{It}\sqrt{Var(r)}\beta_i\sigma_d g\left[ \frac{\lambda \beta_i (1 + \sigma_d^2)\sigma_n}{\sqrt{Var(r)}} \right] \\
+ m_{Ht}\sqrt{Var(r)}\beta_d^2\sigma_d + \beta_d^2\sigma_n^2 g\left[ \frac{\lambda \beta_i m_{It}\sigma_d^2 + \beta_d^2\sigma_n^2}{\sqrt{Var(r)}} \right] \\
+ \sqrt{Var(r)}\sqrt{Var(Q)} g\left[ \frac{\lambda \beta_i Var(Q)}{\sqrt{Var(r)}} \right] \right]
\]

(A50)
\[
Var(\nu) = Var\left(\frac{1}{2} \sum |x_i| + \int |y_j| dj + |Q| \right)
\]
\[
= \frac{1}{4} \left[ (1 - 2/\pi) Var Q + 2m_i \beta_1 \sigma_d \sqrt{Var Q g} \left[ \frac{\beta_i (1 + \sigma_{\nu} \sigma_{\nu}^2) \sigma_{\nu}}{\sqrt{Var Q}} \right] \\
+ 2m_i \sqrt{\beta_i^2 \sigma_d^2 + \beta_i^2 \sigma_h^2} \sqrt{Var Q g} \left[ \frac{\beta_i^2 m_i \sigma_d^2 + \beta_i^2 \sigma_{\nu}}{\sqrt{\beta_i^2 \sigma_d^2 + \beta_i^2 \sigma_h^2}} \right] \\
+ m_i (1 - 2/\pi) \beta_i^2 \sigma_d^2 + m_i (m_i - 1) \beta_i^2 \sigma_h^2 g(\rho_i) \\
+ m_i^2 (\beta_i^2 \sigma_d^2 + \beta_i^2 \sigma_h^2) g \left[ \frac{\beta_i^2 m_i \sigma_d^2 + \beta_i^2 \sigma_{\nu}}{m_i (\beta_i^2 \sigma_d^2 + \beta_i^2 \sigma_h^2)} \right] \right]
\]  

(A51)

where the covariance between the absolute value of two correlated standard normal random variables with correlation \(\rho\) is given by the function:

\[
g(\rho) = \frac{2}{\pi} \left[ \sqrt{1 - \rho^2} - 1 \right] + \frac{\rho}{\pi} \left[ \arctan \left( \frac{\rho}{\sqrt{1 - \rho^2}} \right) - \arctan \left( \frac{\sqrt{1 - \rho^2}}{\rho} \right) + \frac{\pi}{2} \right]
\]

(A52)

The illiquidity moment equation is thus the ratio of Equation (A50) to Equation (A51). These are the key testable implications of the model.

Appendix B – Measuring Returns Using TAQ Data

During regular trading, we only keep trades and quotes meeting standard filters used in the microstructure literature. We drop trades with non-positive price or size and those with correction codes not equal to zero or condition code of M, Q, T, or U. For the pre-market and after-market periods, however, the filters for trades necessarily differ. Importantly, we do not exclude trades with a condition code of T, which explicitly identifies extended hours trades. For extended hours periods, we exclude those that occur at prices probably determined within the trading day (e.g., crosses and block trades), appear out of sequence, or contain non-standard delivery options. This filter eliminates any trades from NYSE, AMEX, or CBOE and trades with “cond” codes B, G, K, M, L, N, O, P, W, U, Z, 4, 5, 6, 8, or 9. We drop trades of at least 10,000 shares or $200,000 regardless of their “cond” codes as these are likely pre-negotiated blocks. Finally, we drop all
trades and quotes in the final minute of the pre-market and in the first minute of the after-market period to mitigate effects of bid-ask bounce.

Within each of the pre-market, regular, and after-market periods, we construct a beginning and ending trade price as the volume-weighted average price (VWAP) based on the first and last minute of trades in the dataset and then compute trade-based returns. Using a VWAP instead of a single trade price further mitigates the effects of bid-ask bounce in returns. When there is only one trade observation in an intraday period, its return is computed from the last price from the most recent intraday period. When there is no trading in a period, the return is zero.

Appendix C – Data and Estimation Procedures

We estimate the covariance matrix of the empirical moments using a model that allows each moment to be persistent and to depend on persistent systematic factors. For simplicity, we model persistence as an AR(1) process at the quarterly frequency. We define three factors \( f_t \) as the sum across moments of each type (i.e., variance, volume, and illiquidity). We estimate an AR(1) model for each factor:

\[
f_t = \alpha_f + \rho_f f_{t-1} + \epsilon_f,
\]

where \( \alpha_f \) and \( \rho_f \) are estimated via OLS.

For each moment \( m_t \), we estimate an AR(1) model in which the moment can also depend on the systematic factor:

\[
m_t = \alpha_m + \rho_m m_{t-1} + \beta_f f_t + \epsilon_m,
\]

where \( \alpha_m, \rho_m, \) and \( \beta_f \) are estimated via OLS. Using this equation, one can define an “abnormal moment” that is orthogonal to the factor:

---

22 Because dividend and stock split adjustments occur between the final trade on the trading day prior to the ex date and the first trade on the ex date, they are not considered in our pre-market, regular market, and after market returns.
\[ \tilde{m}_i = m_i - \beta_j f_i. \]  
(A55)

Computing expectations of the last three equations leads to the following expressions for the means of the factors, the abnormal moments, and the raw moments:

\[ E[f_i] = \frac{\alpha_f}{1 - \rho_f} \]  
(A56)

\[ E[\tilde{m}_i] = \frac{\alpha_m}{1 - \rho_m} \]  
(A57)

\[ E[m_i] = \frac{\alpha_m + \beta_j E[f_i]}{1 - \rho_m} = \frac{\alpha_m}{1 - \rho_m} + \frac{\alpha_m + \beta_j \left( \frac{\alpha_f}{1 - \rho_f} \right)}{1 - \rho_m}. \]  
(A58)

We compute standard errors by applying the delta method to the abnormal and raw moment mean equations. First, we obtain the covariance matrix of \( E[\tilde{m}_i] \) using the delta method. Second, we separately apply the delta method to the second term in each \( E[ m_i ] \) equation to add in the uncertainty in each raw moment mean that is caused by uncertainty in the factor means. We treat the raw and abnormal moments as conceptually distinct because our model analyzes firm-specific information, rather than systematic information. Thus, the model’s moment predictions do not depend on the realizations of systematic factors. Similarly, our empirical method above eliminates the factors’ influence on the off-diagonal terms in the moment covariance matrix by regressing each moment on the appropriate factor (e.g., pre-market non-news volume is regressed on the volume factor) and using the covariances in the resulting residuals \( e_{mt} \) in our covariance matrix calculation. The covariance matrix of residual moments has off-diagonal correlations that are usually quite reasonable—e.g., 95% of the elements have correlations with absolute values of less than 0.7.
Bibliography


Alti, Aydogan, and Paul C. Tetlock, 2012, Biased beliefs, asset prices, and investment: A structural approach, University of Texas and Columbia University working paper.


Lichtenstein, Sarah, and Baruch Fischhoff, 1977, Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance* 20, 159-183.

Obizhaeva, Anna, 2009, Portfolio Transitions and Stock Price Dynamics, University of Maryland working paper.


Table I
Comparative Statics and Parameter Identification

This table summarizes the model presented in Section 2. Panel A describes exogenous parameters and restrictions imposed in the estimation of the Overconfidence (OC) and the Rational (RAT) versions of the model. Panels B and C provide comparative statics for variance, volume, and illiquidity moments in the Overconfidence and Rational versions, respectively.

Panel A: Parameter Descriptions and Restrictions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>OC Restriction</th>
<th>RAT Restriction</th>
<th>Periodic Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\eta_B)</td>
<td>Overconfidence (uninformed perceived precision)</td>
<td>None</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>RRA</td>
<td>Relative risk aversion of uninformed traders</td>
<td>2</td>
<td>None</td>
<td>No</td>
</tr>
<tr>
<td>(\eta)</td>
<td>Informed signal precision</td>
<td>None</td>
<td>None</td>
<td>Yes</td>
</tr>
<tr>
<td>(\sigma_d)</td>
<td>Volatility of acquirable information</td>
<td>None</td>
<td>None</td>
<td>Yes</td>
</tr>
<tr>
<td>(\sigma_h)</td>
<td>Volatility of endowment shocks</td>
<td>0.01(\sigma_d)</td>
<td>0.01(\sigma_d)</td>
<td>No</td>
</tr>
<tr>
<td>(\rho_z = \rho_{Bu})</td>
<td>Correlation among signal errors</td>
<td>0</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>(\rho_u)</td>
<td>Actual correlation of uninformed signals</td>
<td>0.345</td>
<td>0.345</td>
<td>No</td>
</tr>
<tr>
<td>(\rho_h)</td>
<td>Correlation among endowment shocks</td>
<td>1</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>(c_I)</td>
<td>Cost of informed entry</td>
<td>3*10^{-8}</td>
<td>3*10^{-8}</td>
<td>No</td>
</tr>
<tr>
<td>(c_H)</td>
<td>Cost of uninformed entry</td>
<td>0.01(c_I)</td>
<td>0.01(c_I)</td>
<td>No</td>
</tr>
</tbody>
</table>

Panel B: Overconfidence Model (estimated \(\eta_B > 0\) and fixed plausible RRA)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variance(t)</th>
<th>Variance(t+1)</th>
<th>Volume(t)</th>
<th>Illiquidity(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\eta_B)</td>
<td>Inverted U-shape</td>
<td>U-shaped</td>
<td>Inverted U-shape</td>
<td>U-shaped</td>
</tr>
<tr>
<td>RRA</td>
<td>Almost none</td>
<td>N/A</td>
<td>Almost none</td>
<td>Almost none</td>
</tr>
<tr>
<td>(c_I)</td>
<td>None</td>
<td>N/A</td>
<td>Increase</td>
<td>Decrease</td>
</tr>
<tr>
<td>(\eta)</td>
<td>U-shaped</td>
<td>Inverted U-shape</td>
<td>Inverted U-shape</td>
<td>U-shaped</td>
</tr>
<tr>
<td>(\sigma_d)</td>
<td>Increase</td>
<td>Increase</td>
<td>Decrease</td>
<td>Increase</td>
</tr>
</tbody>
</table>

Panel C: Hedging Model (fixed \(\eta_B = 0\) and estimated high RRA)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variance(t)</th>
<th>Variance(t+1)</th>
<th>Volume(t)</th>
<th>Illiquidity(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\eta_B)</td>
<td>Decrease</td>
<td>Increase</td>
<td>Decrease</td>
<td>Increase</td>
</tr>
<tr>
<td>(c_I)</td>
<td>None</td>
<td>N/A</td>
<td>Increase</td>
<td>Decrease</td>
</tr>
<tr>
<td>(\eta)</td>
<td>U-shaped</td>
<td>Inverted U-shape</td>
<td>Decrease</td>
<td>Increase</td>
</tr>
<tr>
<td>(\sigma_d)</td>
<td>Increase</td>
<td>Increase</td>
<td>Increase</td>
<td>Inverted U-shape</td>
</tr>
</tbody>
</table>
Table II
Empirical and Model Predicted Moments

This table presents empirical moment estimates for variance, volume, and illiquidity moments for the Regular Market (9:30 AM to 4:00 PM), the Pre-Market (7:00 AM to 9:30 AM), and the After Market (4:00 PM to 6:30 PM) conditional on news or no news. Standard errors computed as in Appendix B are in parentheses. For variance and volume moments in either the Pre-Market or After-Market, numbers are scaled up by a factor of 6.5/2.5 for comparison with the Regular Market period. Variance moments are expressed as volatility and volume moments are expressed as turnover. Predicted moments from GMM estimation of both the Overconfidence (OC) and Rational (RAT) versions of the model are also included. For each model, the $\chi^2$ statistic and $p$-value from a test of overidentifying restrictions are provided. Panel A presents full sample results (2001-2010), while Panel B presents results for subperiods separately.
Table II: continued

Panel A: Full Sample (2001-2010)

<table>
<thead>
<tr>
<th>Regular Market Moments</th>
<th>Empirical</th>
<th>OC Prediction</th>
<th>RAT Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>News Volatility (%)</td>
<td>3.999</td>
<td>3.666</td>
<td>4.023</td>
</tr>
<tr>
<td>No-News Volatility (%)</td>
<td>3.040</td>
<td>2.867</td>
<td>3.252</td>
</tr>
<tr>
<td>News Turnover (%)</td>
<td>2.113</td>
<td>2.148</td>
<td>2.095</td>
</tr>
<tr>
<td>No-News Turnover (%)</td>
<td>1.455</td>
<td>1.456</td>
<td>1.490</td>
</tr>
<tr>
<td>News Illiquidity</td>
<td>7.747</td>
<td>8.596</td>
<td>7.319</td>
</tr>
<tr>
<td>No-News Illiquidity</td>
<td>7.920</td>
<td>7.962</td>
<td>8.214</td>
</tr>
</tbody>
</table>

Pre-Market Moments

| News Volatility (%)                        | 3.572     | 3.655         | 3.554          |
| No-News Volatility (%)                     | 1.182     | 1.240         | 1.225          |
| News Turnover (%)                          | 0.061     | 0.055         | 0.053          |
| No-News Turnover (%)                       | 0.008     | 0.008         | 0.007          |
| News Illiquidity                           | 227.125   | 129.466       | 113.261        |
| No-News Illiquidity                        | 259.393   | 227.084       | 244.582        |

After-Market Moments

| News Volatility (%)                        | 6.044     | 6.060         | 5.946          |
| No-News Volatility (%)                     | 1.635     | 1.832         | 2.120          |
| News Turnover (%)                          | 0.148     | 0.143         | 0.110          |
| No-News Turnover (%)                       | 0.021     | 0.020         | 0.019          |
| News Illiquidity                           | 112.389   | 111.994       | 57.276         |
| No-News Illiquidity                        | 90.207    | 83.189        | 76.366         |

\[
\chi^2(5) = 7.172, \quad p-value = 0.208
\]

\[
\chi^2(5) = 29.858, \quad p-value = 0.000
\]
Table II: continued

<table>
<thead>
<tr>
<th>Regular Market Moments</th>
<th>2001-2005</th>
<th>2006-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>News Volatility (%)</td>
<td>Empirical</td>
<td>OC</td>
</tr>
<tr>
<td></td>
<td>(0.590)</td>
<td>(0.780)</td>
</tr>
<tr>
<td>No-News Volatility (%)</td>
<td>3.379</td>
<td>3.266</td>
</tr>
<tr>
<td></td>
<td>(0.460)</td>
<td>(0.585)</td>
</tr>
<tr>
<td>News Turnover (%)</td>
<td>2.057</td>
<td>2.082</td>
</tr>
<tr>
<td></td>
<td>(0.257)</td>
<td>(0.257)</td>
</tr>
<tr>
<td>No-News Turnover (%)</td>
<td>1.314</td>
<td>1.301</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.182)</td>
</tr>
<tr>
<td></td>
<td>(2.105)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.148)</td>
<td></td>
</tr>
</tbody>
</table>

| Pre-Market Moments     |           |           |
|                        | (0.311)   | (0.354)   |           |           |           |           |
| No-News Volatility (%) | 1.356     | 1.388     | 1.367     | 0.976     | 1.053     | 1.053     |
|                        | (0.221)   | (0.307)   |           |           |           |           |
| News Turnover (%)      | 0.060     | 0.055     | 0.054     | 0.062     | 0.057     | 0.053     |
|                        | (0.011)   | (0.011)   |           |           |           |           |
| No-News Turnover (%)   | 0.010     | 0.010     | 0.010     | 0.006     | 0.006     | 0.005     |
|                        | (0.002)   | (0.002)   |           |           |           |           |
| News Illiquidity       | 208.571   | 124.006   | 111.932   | 245.680   | 166.458   | 118.355   |
|                        | (70.371)  | (70.371)  |           |           |           |           |
| No-News Illiquidity    | 212.884   | 185.784   | 217.131   | 305.903   | 299.306   | 292.157   |
|                        | (90.850)  | (90.850)  |           |           |           |           |

| After-Market Moments   |           |           |
| News Volatility (%)    | 7.186     | 7.172     | 7.141     | 4.629     | 4.463     | 4.599     |
|                        | (0.661)   | (1.026)   |           |           |           |           |
| No-News Volatility (%) | 1.439     | 1.805     | 2.008     | 1.811     | 1.844     | 2.256     |
|                        | (0.398)   | (0.316)   |           |           |           |           |
| News Turnover (%)      | 0.173     | 0.168     | 0.144     | 0.122     | 0.123     | 0.079     |
|                        | (0.025)   | (0.025)   |           |           |           |           |
| No-News Turnover (%)   | 0.025     | 0.023     | 0.023     | 0.018     | 0.017     | 0.014     |
|                        | (0.004)   | (0.004)   |           |           |           |           |
| News Illiquidity       | 110.335   | 109.361   | 46.054    | 114.444   | 117.229   | 78.710    |
|                        | (21.303)  | (21.303)  |           |           |           |           |
| No-News Illiquidity    | 80.787    | 64.849    | 63.236    | 99.627    | 98.634    | 88.845    |
|                        | (21.489)  | (21.489)  |           |           |           |           |

| $\chi^2(5)$ | 3.926 | 16.823 | 3.427 | 12.363 |
| $p$-value    | 0.560 | 0.005 | 0.635 | 0.030 |
Table III
Parameter Estimates

This table presents GMM parameter estimates for two versions of the model. The Overconfidence (OC) version restricts the Relative Risk Aversion (RRA) parameter, while the Rational (RAT) version restricts the parameter $\eta_B$. These restricted parameter values are presented in brackets. Other restrictions are listed in Table I Panel A. Parameters vary across intraday periods as listed, and “News” and “No News” indicate parameters from periods with and without public news, respectively. Standard errors are reported in parentheses. Panel A presents full sample results (2001-2010), which Panel B presents results for subperiods separately.

<table>
<thead>
<tr>
<th>Panel A: Full Period</th>
<th>OC</th>
<th>RAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters for all periods</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_B$</td>
<td>0.188</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>RRA</strong></td>
<td>[2.000]</td>
<td>131.529</td>
</tr>
<tr>
<td></td>
<td>(28.470)</td>
<td></td>
</tr>
<tr>
<td><strong>Parameters for Regular Market</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$(News)</td>
<td>0.153</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>$\eta$(No News)</td>
<td>0.191</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>$\sigma_d$(News)</td>
<td>0.039</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\sigma_d$(No News)</td>
<td>0.030</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td><strong>Parameters for Pre-Market</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$(News)</td>
<td>0.507</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>$\eta$(No News)</td>
<td>0.942</td>
<td>0.344</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>$\sigma_d$(News)</td>
<td>0.040</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\sigma_d$(No News)</td>
<td>0.014</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td><strong>Parameters for After-Market</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$(News)</td>
<td>0.325</td>
<td>0.741</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>$\eta$(No News)</td>
<td>0.793</td>
<td>0.259</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>$\sigma_d$(News)</td>
<td>0.064</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\sigma_d$(No News)</td>
<td>0.016</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Parameters for all periods</th>
<th>Panel B: Subperiods</th>
<th>2001-2005</th>
<th>2006-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OC</td>
<td>RAT</td>
<td>OC</td>
</tr>
<tr>
<td>( \eta_B )</td>
<td>0.164</td>
<td>[0.000]</td>
<td>0.292</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.095)</td>
<td></td>
</tr>
<tr>
<td>( RRA )</td>
<td>[2.178]</td>
<td>141.291</td>
<td>[1.733]</td>
</tr>
<tr>
<td></td>
<td>(41.370)</td>
<td>(34.250)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters for Regular Market</th>
<th>2001-2005</th>
<th>2006-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta(\text{News}) )</td>
<td>0.173</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>( \eta(\text{No News}) )</td>
<td>0.225</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>( \sigma_d(\text{News}) )</td>
<td>0.046</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>( \sigma_d(\text{No News}) )</td>
<td>0.034</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters for Pre-Market</th>
<th>2001-2005</th>
<th>2006-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta(\text{News}) )</td>
<td>0.571</td>
<td>0.317</td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>( \eta(\text{No News}) )</td>
<td>0.952</td>
<td>0.815</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.142)</td>
</tr>
<tr>
<td>( \sigma_d(\text{News}) )</td>
<td>0.041</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>( \sigma_d(\text{No News}) )</td>
<td>0.015</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters for After-Market</th>
<th>2001-2005</th>
<th>2006-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta(\text{News}) )</td>
<td>0.343</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>( \eta(\text{No News}) )</td>
<td>0.841</td>
<td>0.605</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.155)</td>
</tr>
<tr>
<td>( \sigma_d(\text{News}) )</td>
<td>0.075</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>( \sigma_d(\text{No News}) )</td>
<td>0.015</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>
Table IV
Analysis of Overconfidence and Rational Models

This table summarizes estimates from the Overconfidence (OC) and Rational (RAT) versions of the model as estimated in Tables II and III. The final column summarizes a second version of the Rational model in which the $\rho_h$ parameter is restricted to 0.1. For the $RRA$ and $\eta_B$ parameters, standard errors are provided in parentheses. The Volume and Variance Decompositions are based on Equations (31) and (32), respectively, with parameter estimates for each relevant model. The Annual Cost of Informed Trading is the sum across all intraday periods of the parameter $c_I$ times the expected endogenous parameter $m_I$ (the number of informed traders) times 252 trading days. The annual amount of hedging is the fraction of trading volume explained by hedging multiplied by expected annual turnover.

<table>
<thead>
<tr>
<th>Parameter Values</th>
<th>OC</th>
<th>RAT ($\rho_h^2 = 1.0$)</th>
<th>RAT ($\rho_h^2 = 0.01$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Risk Aversion (RRA)</td>
<td>2.0</td>
<td>131.5</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(28.5)</td>
<td>(2.8)</td>
</tr>
<tr>
<td>Overconfidence ($\eta_B$)</td>
<td>0.188</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Volume Decomposition**

- Informed Trading: 80.5%  89.5%  63.3%
- Uninformed - Overconfident: 12.0%  0.0%  0.0%
- Uninformed - Hedging: < 0.1%  4.7%  32.5%
- Market Making: 7.5%  5.8%  4.2%

**Variance Decomposition**

- Private Information: 90.8%  72.8%  68.7%
- Public Information: 9.2%  27.2%  31.3%

**Aggregated Statistics (% mkt cap)**

- Annual Cost of Informed Trading: 0.31%  0.40%  0.30%
- Annual Amount of Hedging: < 0.1%  19.0%  131.0%
Figure 1: Probability of news arrival.
This figure plots quarterly probabilities of news arrival for each of three intraday periods (the Regular Market, the Pre-Market, and the After Market) during 2001-2010. The indicator variable News is 1 if there is at least one story (two stories) in the Dow Jones Newswires mentioning a particular Small or Mid Cap (Large Cap) firm and 0 otherwise. To precisely measure news arrival, only periods for which there are no news stories since the same intraday period on the prior trading day are considered. All probabilities are calculated by pooling observations for all firms within each quarter and size group and then averaging across size groups. Intraday periods are as defined above.
Figure 2: Empirical moments in periods with and without news arrival.
This figure plots ratios of news to no-news moment estimates of variance, volume, and illiquidity within each of three intraday periods (the Regular Market, the Pre-Market, and the After Market) during 2001-2010. Calculations and intraday period definitions are as described in Table II.
Figure 3, Panel A: Prediction error t-statistics for the overconfidence model.
This figure shows prediction error t-statistics for variance, volume, and illiquidity moments in each of three intraday periods (the Regular Market, the Pre-Market, and the After Market) conditional on news or no-news. The overconfidence version model is estimated using GMM for the 2001-2010 sample period using as in the middle column of Table II, Panel A.
This figure shows prediction error $t$-statistics for variance, volume, and illiquidity moments in each of three intraday periods (the Regular Market, the Pre-Market, and the After Market) conditional on news or no-news. The rational version of the model is estimated using GMM for the 2001-2010 sample period using as in the right column of Table II, Panel A.