The Information Content of Volatility and Order Flow — Intraday Evidence from the U.S. Treasury Market

George J. Jiang and Ingrid Lo[1]

August 2008

[1]George Jiang is from the Department of Finance, Eller College of Management, University of Arizona, Tucson, Arizona 85721-0108, Email: gjiang@email.arizona.edu. Ingrid Lo is from the Financial Markets Department, Bank of Canada, Ottawa, Canada, Email: ingridlo@bankofcanada.ca.

The Information Content of Volatility and Order Flow
— Intraday Evidence from the U.S. Treasury Market

Abstract

Motivated by findings in the existing literature on the information content of volatility and order flows, we specify a simple regime switching model to identify the "informed state" in the U.S. Treasury market. Using tick-by-tick price and order flow data of the 2-, 5-, and 10-year notes, we obtain the estimates for the probability of informed (PIN) state throughout the trading day. We show that intraday patterns of the PIN estimates are consistent with the scheduled arrival of public information, and that the PIN estimates are also highly correlated with public information shocks as measured by announcement surprises. Further, we show that PIN is positively related to trading volume and overall market depth. More importantly, significant PIN measures indicate a permanent price impact of information flow up to at least one-week horizon. Finally, we provide evidence that a higher PIN is associated with more divergence of investor opinion, and that PIN has predictive power for divergence of investor opinion in the U.S. Treasury market especially on non-announcement days.

I. Introduction

Existing literature documents that both volatility and order flow contain significant market information. The literature of modeling information flow using asset price volatility goes back to at least Clark (1973). In an arbitrage-free economy, Ross (1989) shows that price volatility is perfectly correlated with information arrival. Andersen (1996) and Andersen and Bollerslev (1997) relate information arrival to stochastic volatility and show that higher volatility is associated with the arrival of information. Ederington and Lee (1993) point to public information arrival i.e. the public announcements, as a major source of price volatility in the T-bond market. Examining the intraday patterns of the U.S. Treasury market, Balduzzi, Elton and Green (2001) and Fleming and Remolona (1999) show that in the case of macroeconomic announcements volatility surges significantly compared to non-announcement times. Engle and Li (1999) document asymmetric response of volatility to positive and negative shocks in Treasury futures market on announcement days.

In addition to the above studies, several studies also examine the role of private information (or difference in the interpretation of public information) on the trading in the US Treasury market. For example, Brandt and Kavajecz (2004), Green (2004), Menkveld, Sarkar and Wel (2008), Pasquariello and Vega (2007) use order flow to proxy for the heterogenous information revealed through trading and find evidence of informed trading in treasury bond market. The evidence suggests that heterogenous interpretation of information can occur and private information can be present on both announcement days and non-announcement days.

In this paper, we extend existing studies by examining the probability of informed state measured intra-daily in the US treasury market. First, we combine the information content of volatility and that of order flow to identify the "informed state". Second, we utilize the tick-by-tick data to infer the probability of information arrival throughout the trading day. The data used in our study is obtained from the BrokerTec electronic trading platform and contains around-the-clock trades and quotes for the on-the-run 2-year, 5-year, and 10-year notes.[1] Moreover, since macroeconomic news announcements

[1]During our sample period, the BrokerTec electronic trading platform accounts for about 60% of trading activity for these securities. While the data also contains trades and quotes of 3-year note and 3-year bond, the liquidity for these two maturities is much lower.

are mostly pre-scheduled, the intraday market transaction data allows us to investigate the interaction between information arrival and investors' difference of opinion in the treasury market. Within such framework, we further examine how information flow interact with investors' difference of opinion. In this aspect, we extend Pasquariello and Vega (2007) who examine the interaction of informed trading and dispersion of private opinion on a daily aggregate level.

To measure information flow, we specify a simple Markov switching model and incorporate order flow to model price dynamics in the US treasury market over each 5-minute interval. Using the regime switching model, we identity intradaily the probability of information arrival in the Treasury market. The advantage of using a Markov switching model is that the probability of information arrival can be measured at intraday level. This is in contrast to existing measure of probability of informed trading proposed by Easley, Hvidkjaer and O'Hara (2002) where the empirical estimation of probability of informed trading is implemented using aggregation of buy and sell transactions over daily intervals and PIN is estimated over a relatively long horizon.[2] Thus the estimates of PIN proposed in Easley, Hvidkjaer and O'Hara (2002) do not allow us to examine informed trading over shorter horizon e.g intradaily interval. The Markov switching model has been used in the literature to infer the probability of informed trading. Ahn and Melvin (2007) and Sager and Taylor (2004) use a regime switching model to identify the presence of information-based trading around important announcements in the foreign exchange market. This paper generalize and extends Ahn and Melvin (2007) and Sager and Taylor (2004) in three ways: first, we recognize the importance of private information on price discovery process in Treasury market and incorporate order flow to proxy private information flow explicitly. Second, we offer direct evidence that PIN from a Markov switching model is related to information arrival by examining the permanent price impact of PIN. Third, instead of using intra-day data on only announcement days and a limited set of control days as in Melvin and Ahn (2007) and Sager and Taylor (2004), we estimate the Markov switching model using all trading days. In this way, we can examine

[2]There is an extensive existing literature that examines information-based trading in the equity market. Studies by, e.g., Easley, Kiefer, O'Hara, and Paperman (1996), Easley, Hvidkjaer, and O'Hara (2002), propose a structural model and infer the probability of informed trading (PIN) from market transactions. They document that the probability of information-based trading is lower for high volume stocks. More importantly, they show that the estimates of PIN contain information about future stock returns.

the probability of information arrival on both announcement days and non-announcement days and examine whether the nature of information arrival matters.

We show that intraday patterns of the PIN estimates are consistent with the scheduled arrival of public information. The PIN estimates are highest at standard macroeconomics announcement times. Furthermore, the PIN estimates are also highly correlated with public information shocks. The announcement surprises is all positively related to the PIN estimates in the 2-year, 5-year and 10-year notes. Examining the relations between market activities and information arrival, our results show that absolute return, depth at the best quotes and trading volume are monotonically increasing with PIN across all maturities. The result indicates that a higher probability of information flow into the market is accompanied by a larger absolute price change and lower liquidity provision at the best quote. Spread, a widely adopted proxy for informed trading in microstructure literature, does not seem directly related to PIN. However, lagged spread is inversely related to PIN. One potential explanation is that the major source of information in the treasury bond market is macroeconomics announcements, which is prescheduled and anticipated by market participants. They therefore would have precise knowledge of timing of information arrival and could widen the spread before information arrival to protect their positions.

More importantly, we found that significant PIN measures indicate a permanent price impact of information flow up to at least one-week horizon. To test whether price changes associated with high PIN are permanent, we apply the non-parametric test in Kaniel and Liu (2006). Our results show that there is a permanent price impact of information flow up to at least 1-week horizon for 2-, 5- and 10-year notes. The price impact of PIN is significant at 1% level for announcement days and at 5% level for non-announcement days up to 1-week horizon for 2-, 5- and 10-year notes. The results are robust controlling for trading volume, market depth and spread. Furthermore, we find that the magnitude of price impact is asymmetric for announcement days and non-announcement days with significant PIN. The price impact is stronger in one-hour horizon in announcements days than non-announcement days. The subsequent price impact tends to taper off over 1-hour and 1-week horizons on announcement days, particularly following positive price change at significant PIN. However, the price impact becomes larger over one-week horizon on non-announcement days following positive price change at significant PIN. The finding holds for all three notes. The result indicates that the nature of information arrival has different implications for price dynamics.

Finally, we provide evidence that PIN not only has a positive contemporaneous relation with diverse private opinion but also has predictive power on future private opinion. There is, however, a mixed relation between PIN and subsequent diverse investor opinion on announcement days versus non-announcement days. The theoretical model of Pasquariello and Vega (2007) shows that information heterogeneity is an important factor in driving prices. With the payoff-relevant information known to all market participants in the treasury securities market, information heterogeneity takes the form of diverse private opinion or private interpretation of arrival of information on the valuation of an asset. Pasquariello and Vega (2007) find that there is more informed trading with diverse beliefs in the treasury securities market. Two related papers on private opinion dispersion are Pasquariello and Vega (2006) and Beber, Breedon and Baraschi (2008). Both papers construct proxies of diverse private opinion using monthly dispersion across professional forecast. Motivated by Naes and Skjeltorp (2006) finding on positive correlation between the variation of analyst and the slope of the order book, we use the slope of the order book to measure dispersion in private belief intradaily. A gentler slope indicates that market participants submit their orders over a wider range of quotes for a given volume, thus revealing a more diverse private opinion over the value of the treasury note. This measure has two advantages over using dispersion of professional forecast: (i) compared to the order book slope which can be measured intradaily, monthly forecast could become stale as the time passed between the forecast and decision making lengthens (ii) as long as the probability of execution is non-zero, order placed on limit order book represents beliefs backed up by real money. Our results show that the slope of the order book is in general negatively related to the probability of information arrival, indicating diverse price opinion is positively correlated with information arrival. The pattern holds for all maturities, over both announcement and non-announcement days. This finding is consistent and extends the finding of Pasquariello and Vega (2007) on daily interval to intra-day interval. Furthermore, the subsequent belief dispersion depends on the nature of information arrival. Information arrival on non-announcement days tends to predict a more diverse belief dispersion subsequently but this relation does not hold on announcement days, particularly over shorter maturity notes.

The rest of paper is structured as follows. Section II describes the data used in our study and the Markov switching model. Section III present main empirical results. Section IV presents the results on private belief dispersion. Section V concludes.

II. Data and Model

A. Data

The U.S. Treasury securities data are obtained from BrokerTec, an interdealer electronic trading platform in the secondary wholesale U.S. Treasury securities market. Since 2003, the majority of secondary trading has gone through electronic platforms with over 95% of active issue treasury occurring on electronic platforms.[3] Two platforms dominate the U.S. treasuries market: BrokerTec and E-speed. BrokerTec has a market share of 60-65% on the active issues and is more active in the trading of 2-year, 5-year and 10-year Treasury notes. There has been a strong growth in trading volume on the BrokerTec platform in recent years. The average daily trading volume of all maturities goes up from $30.9 billion in 2003, $53.0 billion in 2004, $80.2 billion in 2005, to $103.4 billion in 2006. The BrokerTec platform functions as a limit order book. Traders can submit limit orders, i.e., orders that specify both price and quantity posted on the book, or they can submit marketable limit orders, i.e., orders with a better price than or equal to the best price on the opposite side of the market, to ensure immediate execution. Limit order submitters can post "iceburg" orders, where only part of their order are visible to the market and the remaining part is hidden. All orders on the book except the hidden part of the orders are observed by market participants. The orders remain in the market until matched, deleted, inactivated, loss of connectivity, or market close. The market operates more than 22 hours a day from Monday to Friday. After the market closes at 5:30 p.m. (EST), it opens again at 7:00 p.m. (EST). The data set contains the tick-by-tick observations of transactions, order submissions and order cancellations. It includes the

[3] See "Speech to the Bond Market Association", December 8, 2004 by Michael Spencer, founder and chief executive of ICAP, one of the world's largest interdealer broker.

time stamp of the observations, the quote, the quantity entered and deleted, the side of the market and, in the case of a transaction, an aggressor indicator.

We use data from 7:30 a.m. EST to 5:00 p.m. EST since trading is more active during this time interval. This interval also contains all pre-scheduled U.S. news announcements, and it provides us with 9.5 hours of trading and 114 five-minute return observations each day. The choice of working on five-minute returns follows Fleming and Remolona (1999), Balduzzi, Elton and Green (2001), and others. Since liquidity has changed drastically over time, we restrict our sample period to the most recent years, i.e., from January 2, 2005 to December 29, 2006. Days with early closing before public holidays are also excluded as liquidity is typically low for these days. The dataset consists of over 465.5 million observations and 10.9 million transactions.

Table I provides descriptive statistics of the data. Since the order book contains the price schedule on both sides of the market, there are multiple ways to measure liquidity. We compute and report the bid-ask spread, daily trading volume (in $billions), trading duration (in seconds), daily return volatility, depth at the best quote, depth of the entire book, and hidden depth. Spread is defined both in relative terms and in ticks. Relative spread is defined as

$$\text{relative spread} = (\text{best bid price} - \text{best ask price})/\text{mid-quote} \tag{1}$$

and measured at the end of each 5-minute interval and averaged over the trading day. Tick spread is also measured at the end of each 5-minute interval and averaged over the trading day. As mentioned in Fleming and Mizrach (2008), the tick size differs for different maturities. The tick size of the 2-year and 5-year note is 1/128, whereas that of the 10-year note is 1/64. Daily return volatility is calculated as the square-root of the sum of squared log mid-quote difference sampled at 5-minute intervals

$$\text{return volatiilty} = (\sum_{i=1}^{114}(\ln p_i - \ln p_{i-1})^2)^{1/2} \tag{2}$$

where the mid-quote is defined as $p_i = (\text{best bid price} + \text{best ask price})/2$. The average (hidden) depth (in millions) at the best bid/ask is the total (hidden) observed depth at the best price on both the bid and ask side of the market measured at the end of each 5-minute interval and averaged over the trading day. The average depth and average hidden depth in the entire order book are defined similarly.

BrokerTec is a highly liquid platform over our sample period from 2005 to 2006. As shown in Table I, relative spread is smallest for the 2-year note with a sample mean of less than 0.0083% among the actively traded securities, followed by the 5-year note (0.0119%) and 10-year note (0.0179%). The tick spread is consistent with the relative spread. Trading volume is heaviest for the 2-year note ($27.45 billion per day), followed by the 5-year note ($24.69 billion per day), and 10-year note ($22.76 billion per day). In terms of trading duration, the 10-year note is most frequently traded, with an average duration of 6.59 seconds. This is closely followed by the 5-year note at 6.74 seconds. The trading duration of the most heavily traded 2-year note is on average 15.99 seconds. The result suggests that the average trade size is larger for the 2-year note than the 5-year and the 10-year note.

Return volatility is generally increasing with maturity. The trend seems related to where the depth accumulates on the order book. The mode of depth for the 2-year note locates closest to the best price, on average around 1.18 ticks away from the best price on both sides of the market. As maturity increases, depth mode locates further away from the best price: 1.67 ticks for the 5-year note and 1.53 ticks for the 10-year note. Thus normal price movements are more likely to be restricted by depth aggregated at the mode. The finding is consistent with Kavajecz and Odders-White (2004) in the equity market where accumulation of depth at a price level restricts the range of normal price changes.

The 2-year note has the deepest book both at the best price ($637.72 million) and entire book ($5,122 million). Hidden depth is low in general: hidden orders at the best price consist of less than 5% of the observed depth at the best price for the 2-year, 5-year, and 10-year notes.

Figure 1 presents the intra-day activities in the 2-year note. The intraday patterns for other bonds are similar and thus not reported for brevity. Consistent with the findings in Fleming (1997), trading volume peaks first in the 8:30 to 10:00 EST interval and goes up again from 13:00 to 14:00 EST. These two intervals overlap with major macroeconomic announcements. Trading duration shows the reverse pattern of trading volume. The time between transactions is longer at the end of the day, averaging over 40 seconds. At the most hectic interval from 8:30 to 9:00 EST, there are on average fewer than 5 seconds between transactions. Relative spread is higher at the beginning (before 8:30 EST)) and the end of the trading day (after 16:00 EST). The depth at the best price is thinner before 8:30 EST and after

15:00 EST. For the rest of the day, the book is on average over $600 millions. The level of hidden depth is higher at noon and it goes up again after 15:00 EST. This finding suggests that market participants hide more of their orders when there is less total depth in the market.

Data on macroeconomic news announcements and the survey of market participants comes from Bloomberg and Briefing.com economic calendar. We cover an extensive list of announcements. Our announcement data include the 25 announcements from Pasquariello and Vega (2006) and 8 other economic announcements: FOMC minutes, ISM service, Consumer Confidence, NY Empire State Index, Chicago PMI, Existing Home Sales, Philadelphia Fed Index, and ADP National Employment report. Lastly, we collect the release of the testimony of Semiannual Monetary Policy Report and Economic Outlook. Following Balduzzi, Elton and Green (2001) and Andersen Bollerslev, Diebold and Vega (2003), the standardized news surprise is defined as

$$S_{kt} = \frac{A_{kt} - E_{kt}}{\hat{\sigma_k}} \tag{3}$$

where A_{kt} is the actual announcement, E_{kt} is the median forecast for news k on day t and $\hat{\sigma_k}$ is the standard deviation of $A_{kt} - E_{kt}$.

B. The Markov Switching Model

We use a simple two-state Markov switching model to capture information arrival in the U.S. Treasury market. The two states represent, respectively, the lack of information flow in the market (State 1) and the presence of information flow into the market (State 2). The transition probabilities is defined as

$$p(p_t = 1|p_{t-1} = 1) = p_{11} \tag{4}$$

$$p(p_t = 2|p_{t-1} = 1) = 1 - p_{11} \tag{5}$$

$$p(p_t = 1|p_{t-1} = 2) = 1 - p_{22} \tag{6}$$

$$p(p_t = 2|p_{t-1} = 2) = p_{22} \tag{7}$$

Here the aim is to establish the presence of state-varying information flow to the Treasury market. As mentioned in the introduction, the nature of information can be public or private. As such, we combine the information contained in volatility and order flow to identify the "informed state". The modeling of information flow through asset price volatility goes back to Clark (1973). Ross (1989) shows that price volatility is perfectly correlated with information arrival in an arbitrage-free economy. Andersen (1996) and Andersen and Bollerslev (1997) relate information arrival to stochastic volatility, with higher volatility indicating arrival of information. Furthermore, Balduzzi, Elton and Green (2001), Ederington and Lee (1993) and Fleming and Remolona (1999) shows that in the case of public information arrival i.e. macroeconomic announcement, volatility surges significantly compared to non-announcement times. Thus a higher volatility is related to information arrival. Starting with Kyle's (1985) model of speculative trading, and later extensions by Foster and Viswanathan (1990), Back et al. (2000) and Pasquariello and Vega (2007), the literature has postulated that order flow carries private information of traders. The price impact of order flow is greater in the presence of informed traders. Empirical research on the Treasury market, such as Brandt and Kavajecz (2004), Green (2004), Pasquariello and Vega (2007) and Menkveld et al. (2008), confirms that order flow aggregates private information and it has a higher price impact when information flows into the market during announcement times. Incorporating these two elements of information flow, the Markov switching is given by

$$\Delta p_t = \rho_p \Delta p_{t-1} + \mu_p + \mu_{p,Inf} * I_t + \beta_{OF} OF_t + \beta_{OF,Inf} * OF_t * I_t + \varepsilon_t \tag{8}$$

with $\varepsilon_t \sim N(0, \sigma_p + \sigma_{p,Inf} * I_t)$ and I_t is equal to 1 if information arrives at the market and 0 otherwise.

Both volatility and price impact of order flow should be higher when information flows into the market, we expect $\sigma_{p,Inf}$ and $\beta_{OF,Inf}$ to be positively significant. Also, if information arrival is incorporated quickly into asset prices, the transitional probability from informed state to informed state should be lower than the transitional probability from uninformed state to uninformed state. Thus another testable hypothesis is that $p_{22} < p_{11}$ because information dissipate quickly in an efficient market and so an informed state is less likely to continue in the next period.

The model in Equation (8) generalize and extends Ahn and Melvin (2007) and Sager and Taylor (2004) Markov switching model in two major ways: first, by incorporating order flow, we control for private information flow explicitly in price dynamics whereas their work use only volatility to identify different information states. Second, instead of using only intra-day data on only announcement days and a limited set of control days as in Melvin and Ahn (2007) and Sager and Taylor (2004), we conduct the analysis using all trading days. Thus we are able to examine the probability of information arrival on both announcement and non-announcements and examine whether the nature of information arrival matters.

III. Empirical Results

A. Estimation Results and Probability of Informed (PIN) State

We estimate Markov switching model as defined in Equation (8) using data of the 2-, 5- and 10-year treasury notes note. The return, Δp_t, is defined as the change in logarithmic mid-quote over 5-minute interval. We multiply returns by a factor of 10,000 in our analysis. Order flow is the number of buy transactions minus the number of sell transactions. Table 3 reports the estimation results of the model. For all maturities, the estimates of $\sigma_{p,inf}$ and $\beta_{OF,Inf}$ are significantly positive, indicating return volatility and price impact are higher in State 2, the information arrival state. Another interesting observation is that volatility and price impact, both with and without information arrival, increases with maturities. This is consistent with Brandt and Kavajecz (2004) that the impact of order flow is larger in a less liquid market. That is, the differences are largely driven by liquidity conditions. The depth of both the 5-year and the 10-year note is 5 times smaller than the 2-year note. Also the tick size of the 10-year note is larger so that volatility will be bigger for the same transaction size. Finally, the results also show that the transitional probability of remaining in State 2 (p_{22}), the information arrival state, is lower than the transitional probability of remaining in State 1 (p_{11}), the state with no information arrival. That is, the persistence of informed state is lower than in that of uninformed state. The other pattern we observe is that p_{22} tends to increase in maturity, which is likely due to the fact that price formation or resolution of information uncertainty is slower in a less liquid market.

We next calculate the probability of information flow into the market using the estimates of the Markov switching model. The conditional probability of informed state is given by $P_{2,j} = (P_j = 2|\Delta p_j, \Delta p_{j-1})$ and is calculated using the EM algorithm as described in Hamilton (1990). We define probability of informed state PIN_t as the average of $P_{2,j}$ over 30-minute intervals. The reason of averaging over 30-minute interval is to mitigate estimation error for the probability of informed state due to market variables, such as the liquidity effects. Table 5 reports the summary statistics of PIN estimates. The sample mean of PIN increases with maturity with the sample mean of PIN lowest in the 2-year note. Since the arrival of public information is common across maturities, the differences may be due to the following two factors. First, the response to informational shock may vary for Treasury notes with different maturities. Second, it is likely that the level of private information varies across the market of different bonds. Note that our model captures the flow of both public information and private information. Considering the fact that the 2-year note is the most liquidly traded, it is not surprising that the price impact of private information tends to be smaller.

B. PIN, Information Arrival and Liquidity

We now presents initial evidence on how the probability of information flow, PIN, is related to information arrival and liquidity. Figure 2 plots the intraday patterns of PIN based on the Markov switching model as defined in Equation (8) for the 2-year note. The plots are similar for other maturities. There is a clear intraday pattern of PIN over the day, with the measure peaking around pre-scheduled macroeconomic news announcement times, such as 8:30 AM, 10:00 AM and 14:00 PM. This offers initial evidence that the PIN estimate captures the arrival of public information – macro economic announcements–in the Treasury notes market. Focusing on announcement days, the spiking of PIN at announcement times is even more distinct at the prescheduled announcement time, indicating PIN captures information arrival at pre-scheduled announcement time. On days without announcements, the intraday pattern of information arrival is less distinctive. Nevertheless, the PIN measure is higher in the morning from 8:00 to 9:00 (EST) and after lunch, coinciding with the time that more market participants entering the market. As pointed out in Brandt and Kavajecz (2004), order flow on non-announcement days still carries significant information content because market participants trade on their private asymmetric valuation of the asset. Thus a higher PIN from 8:00 to 9:00 (EST) and after lunch on non-announcement days is consistent with more traders entering the market and thus leads to a surge in private information in these times.

We next examine how PIN is related to information shocks on announcement days. Information shocks here is captured by announcement surprise. As we have over 30 announcements, it is infeasible to include all of them in the estimation. We include only 7 economic announcements deemed important in the market as in Pasquariello and Vega (2006). They are Nonfarm Payroll Employment, Retail Sales, New Home Sales, Consumer Confidence Index, ISM index, Index of Leading Indicators and Initial Unemployment Claims. We regress the probability of information arrival on the announcement surprises. More specifically,

$$PIN_t = \alpha + \Sigma_{j=1}^{J} \gamma_j |Sur_{j,t}| + \epsilon_t \tag{9}$$

where PIN_t is the 5-minute average $P_{2,j}$ over 30-minute interval. Table 5 reports the result of Equation (9). The PIN measure is highly correlated with public information shocks as measured by announcement surprises. Almost all γ_j are positively significant in all three maturities.

How is information arrival related to liquidity in the market? As a preliminary analysis, we sort PIN estimate to form 5 equal groups (quintiles) and examine absolute return ($|ret| = \Delta \ln p_t \times 1000$), the bid-ask spread, depth at the best quotes, depth behind the best quotes and trading volume. Table 6 reports the sorting results. Absolute return, depth at the best quote and trade volume are monotonically increasing with PIN across all maturities. The is consistent with findings in existing literature that information flow into the market is often associated with high return volatility and lower liquidity provision at the best quote. Depth behind best quotes in general increases with PIN for the 2-year note whereas PIN first increases and then drops for the 5-year and the 10-year notes. The lower depth behind the best quotes in the lowest PIN group is consistent with theoretical work like Glosten (1994): a lower probability of trading against informed traders lowers the cost of submitting best orders relative to orders behind the best quotes. Thus investors submit less depth behind the best quotes. The finding on lower depth behind the best quotes at the highest PIN group in the 5-year and 10-year notes indicates that a high probability of information arrival could also be related to a lower depth behind the best quotes in the market. The bid-ask spread, a common proxy for informed trading, does not seem to have a simple monotonic relation with the probability of information arrival. One possible explanation is that the major source of information in the treasury bond market is macroeconomic announcements, which is prescheduled and anticipated by market participants. They therefore would have precise knowledge of timing of information arrival and could widen the spread before information arrives to protect their positions. We explore this possibility by examining whether the lagged spread is inversely related to PIN and found evidence supporting the hypothesis. The lagged spread is highest in the group in highest PIN group, indicating that market participants do indeed widen spread to protect themselves against upcoming information.

IV. Further Analysis

A. Informed State and Permanent Price Impact

We recognize that without a structural model behind the PIN estimated from the Markov switching model described in the last section, one might argue that PIN could be measuring liquidity shocks. We address this question in three ways: first we apply a non-parametric test based in Kaniel and Liu (2006) on the price impact of information flow on bond prices. If PIN is truly informative, then a high PIN should have long-lasting price impact because information arrival should have a permanent effect on price. Otherwise, there should be price reversal. More precisely, let $R_[t-1,t] = \ln P_t - \ln P_{t-1}$ denote the logarithmic return when PIN_t is significantly greater than the median PIN over the trading day. We examine whether the return $R_{[t-1,t+j]} = \ln P_{t+j} - \ln P_{t-1}$, where j is 1 hour, 1 day or 1 week, remains in the same direction as $R_[t-1,t]$. If the price change remains in the same direction, then we have initial evidence of permanent price impact.

Let n_{PIN} be the number of same direction mid-quotes in 1 hour, 1 day or 1 week following significant PIN, and $P_{highPIN}$ be the fraction of times that PIN is significantly higher than median and n be the total number of quotes in the same direction in 1 hour, 1 day or 1 week following all PIN. Under the null hypothesis H_0 of equal informativeness under significant PIN and insignificant PIN, the probability that out of these n quote revisions n_{PIN} or more are preceded by a significant PIN is approximated by

$$1 - N[\frac{n_{PIN} - n \cdot P_{highPI}}{\sqrt{n \cdot P_{highPI}(1 - P_{highPI})}}] \tag{10}$$

where N is the standard normal cumulative distribution function. We conduct the test on the overall sample, on only announcement days and on non-announcement days.

Table 7 reports the p-values of the non-parametric test of the permanent price impact. Panel A shows the results based on all days in our sample. In all three maturities, the null hypothesis of equal informativeness under significant PIN and insignificant PIN is rejected at the 1% level in 1-hour, 1-day and 1-week horizon. Thus the results indicates that significant PIN is related to information arrival and leads to permanent price change. Panel B shows the test conducted on announcement days. The result is similar to the overall sample and the null hypothesis of equal informativeness is again rejected in all maturities at 1% level. Panel C shows the test result of non-announcement days. Permanent price impact remains significant even on non-announcement days: the two-year note have p-values below 5% for all three horizons and the p-values are below 10% for both the 5-year and 10-year note.

To control for market conditions and announcements explicitly, we estimate a Probit model to examine whether informed state predict price change in the same direction as in Kaniel and Liu (2006). More specifically, we control for macroeconomic announcements and market conditions in the following Probit regression,

$$\begin{aligned} P(state_{t+h}=1) &= f(\alpha + \rho_{ann} I(state_{t,ann}) + \rho_{noann} I(state_{t,noann}) + \beta_{depth} depth \\ &\quad + \beta_{trdqn} trdqn + \beta_{sprd} sprd + \gamma_{ann} D_{ann}) \end{aligned}$$

where $P(state_{t+h} = 1)$ if $R_{t-1,t+h}$ is in same direction as R_t, where h is 1 hour, 1 day or 1 week. Otherwise, $P(state_{t+h}) = 0$. The indicator variable $I(state_{t,ann(noann)})$ takes the value of one if $PIN_{t,ann(noann)}$ is significantly larger than median on an announcement day (non-announcement days) and 0 otherwise. The variable $depth$ is overall market depth on the book $trdqn$ is the trading volume and $sprd$ is the bid-ask spread at time t. The variable D_{ann} takes the value of 1 if it is an announcement day but 0 otherwise.

Table 8 reports the Probit regression result. Both the dummy indicating significant PIN at announcement and non-announcement days, ρ_{ann} and ρ_{nonews}, are significantly positive in predicting the same price change direction. This offers evidence that PIN captures information arrival in both announcement days and non-announcement days. The price impact is also more likely to persist when depth is low on announcement days at 1-hour horizon for all three maturities but it does not have an effect afterwards for the 2-year note and at 1-week horizon for the 10-year note. This suggests that liquidity effects gradually wears out for some maturities depending on the resilience of the order book. The depth coefficient on non-announcement days does not have a significant impact on all three maturities at both 1-hour and 1-week horizon. Turning to trading volume, price impact is significantly more likely to persist for all three maturities at all horizons when trading volume is high on announcement day. But trading volume is significantly only at 1-hour horizon for the 10-year notes on non-announcement day. Spread is not significant at predicting price impact (except at 1-hour horizon for 5-year note). Interesting, the announcement day dummy, γ_{ann} , is not significant in predicting price impact after controlling for our measure of information arrival, PIN, depth, spread and trading volume.

Having established evidence on permanent price change with significant PIN, we now examine the magnitude of price impact under significant PIN. Here we utilize Madhavan and Cheng (1997) and measure the magnitude of permanent price impact as $R_{[t-1,t+j]} = lnP_{t+j} - lnP_{t-1}$ where j=1 hour, 1 day or 1 week. We examine two scenarios with significant PIN, when return is positive i.e. $R_{[t-1,t+j]} > 0$ and when it is negative i.e. $R_{[t-1,t]} < 0$. We examine the sample mean of $R_{[t-1,t+j]}$ under both scenarios. Consistent with the findings in the non-parametric test, all permanent price impact is significant at 1% level over 1-hour, 1-day and 1 week ahead for the whole sample for all three maturities. Thus a positive (negative) price change with significant PIN remains positive (negative) 1-week ahead. The price impact tends to be larger for longer maturity notes. The positive price change with significant PIN tends to taper off over time but negative price change with significant PIN tends to widen from 1-hour to 1-day horizon and then drops at one week horizon for all three maturities. Similar to overall sample, price impact following positive price change with significant PIN on announcement days tends to drops as time passes but the price impact following negative price change actually widens overtime for the 5-year note and the 10-year note. The findings suggest asymmetric response to positive and negative price change when public information arrives. Turning to non-announcement days, the result is significant at 1% interval for price impact following positive price change with significant PIN for all three maturities. The results following negative price change is in general significant at 1% interval but it is not significant for the 1-day horizon of the 2-year note and price actually reverses for the 5-year note. Comparing announcement days with non-announcement days, the price impact of announcement day is larger at 1-hour horizon for both positive and negative $R_{[t-1,t+j]}$ in all three maturities. The subsequent price impact tends to taper off over 1-hour and 1-week horizons on announcement days following positive price change. However, the price impact becomes larger over one-week horizon on non-announcement days following positive price change. The finding holds for all notes. The result indicates that the nature of information arrival has different implications for price dynamics.

B. Informed State and Divergence of Opinion

Lastly, we examine how diverse private opinion is related the PIN estimates. Previous literature like Pasquariello and Vega (2006) and Beber, Breedon and Baraschi (2008) use dispersion across professional forecaster on announcement to create monthly measure of opinion dispersion. There are two disadvantage of using monthly forecast. First, as mentioned in the introduction, the forecast could become stale as the time between the forecast and decision making lengthens. Second, the forecast is made on announcements but not directly on the assets involved. Thus the dispersion measure in these studies at best represents a measure with noise on the opinion dispersion on the asset. Motivated by findings in Naes and Skjeltorp (2006) on positive correlation between the variation in analyst forecast and the slope of the order book, we use the slope of the order book to measure dispersion in private belief intradaily. Following the notations in Naes and Skjeltorp (2006), the slope of the order book at the end of 5-minute interval j is given by

$$SLOPE_j = \frac{SE_j + DE_j}{2} \tag{11}$$

where SE_j and DE_j is the slope of the ask side and bid side of the book respectively at 5-minute interval j. The ask side of the slope of the order book is given by

$$SE_j = \frac{1}{N_a}\{\frac{v_1^A}{\frac{p_1^A}{p_0^A} - 1} + \sum_{\tau=1}^{N_a} \frac{\frac{v_{\tau+1}^A}{v_\tau^A} - 1}{\frac{p_{\tau+1}^A}{p_\tau^A} - 1}\} \tag{12}$$

where N_a is the number of ask prices containing orders. τ denotes the tick levels where $\tau = 0$ denotes the bid-ask mid-point and $\tau = 1$ denotes the best ask quote with positive volume. v_τ^A denotes the logarithm of volume at tick level τ and p_τ^A denotes the ask quote at tick level τ. The bid ask of the market is defined in a similar fashion. With over 50% of depth concentrates on the top 4 levels of depth in all maturities, we set N equal to 4 in Equation (12). We average $SLOPE_j$ over each 30-minute interval in the trading day to obtain the slope of the order book Slp_t. A gentler slope indicates that market participants submit their orders over a wider range of quotes for a given volume, thus revealing a more diverse private opinion over the value of the treasury note.

The slope of the order book should be more directly related to dispersion in private opinion than analyst forecast particularly in treasury note market because of two reasons: first, the payoff-relevant information is known in treasury securities and so there is no asymmetric information on payoff as in equity market. The dispersion of orders on the price schedule thus represents disperse private interpretation of information or disperse private opinion on the valuation of the asset. Second, orders placed on limit order book represents beliefs backed up by real money and so should more accurately reflects private belief. Nevertheless, as a robustness check, we calculate the correlation of slope of the order book and analyst forecast dispersion of the most important announcements, Nonfarm Payroll Employment. The analyst forecast dispersion is measured by the standard deviation of across professional forecasts on Bloomberg. The slope of the order book is an average of Slp_t over the most active part of trading day from 8:00 to 15:00 (EST) during the week before the announcement. The correlation is -0.18, -0.10 and -0.04 for the 2-, 5- and 10-year note respectively, indicating that a gentler slope is associated with higher belief dispersion especially for notes of shorter maturities.

How is probability of information arrival related to dispersion of private opinion contemporaneously? We sort the intra-day PIN measure into 5 equal groups (quintiles) and examine the relationship between PIN and slope of the order book. Table 10 shows the sorting result of the 2-, 5- and 10-year note. The slope of the order book is in general negatively related to the probability of information arrival. The highest group of PIN is associated with the gentlest slope in all three maturities. The pattern holds for both announcement days and non-announcement days. This finding is consistent with and extends that of Pasquariello and Vega (2007) on daily positive relation between private information and belief dispersion to intra-day interval.

We next examine how PIN is related to subsequent diverse private opinion. We perform the sorting on PIN and examine the future slope of the order book. The relationship of PIN and the future slope of the order book is different on announcement and non-announcement days. On non-announcement days, the PIN measure is inversely related to future slope of the order book. The result holds for all maturities. This indicates that a higher probability of information arrival on non-announcement days is positively related to subsequent diverse private opinion. On the other hand, the result is less consistent for announcement days. For example, the highest PIN group of the 2-year note is associated with the second steepest slope of the order book. For the 5-year note, the future slope first rises with PIN and then falls. For the 10-year note, the future slope is inversely related to PIN. The result has two potential implications: first, the inverse relationship between PIN and the future slope on non-announcement days indicates that more pedestrian type of information arrival creates more diverse private opinion. Second, the relationship between public information arrival and subsequent private opinion depends on maturities. Public information arrival is associated with a less diverse future private opinion on 2-year note but a more diverse future private opinion in 10-year note. One possible explanation on the divergence between short and long maturity is that the impact of news on shorter maturity note is discounted over a shorter period of time till maturity and thus has less uncertainty compared to a longer maturity note.

V. Conclusion

In this paper, we exploit the intraday tick-by-tick data to study information flow and investors' difference of opinion in the US treasury market. Using a simple Markov switching model, we estimate an intraday measure of PIN. Our paper thus extends the literature which measures probability of informed trading over daily intervals. We find that information arrival is incorporated quickly into market price and private opinion tends to be more diverse when information arrives.

Our findings suggest that the estimated probability of informed state estimated from a Markov switching model effectively captures the information arrival. The measure peaks at the macroeconomic news announcement time, and is significantly related to trading volume and market depth. More importantly, we show that there is a permanent price impact up to 1-week horizon when PIN is significantly higher than the median. The findings are consistent across maturities and are robust when we explicitly control for liquidity conditions in the market and the presence of announcements.

Lastly, we utilizes the slope of the order book as private opinion dispersion measure. We find that the estimated probability of informed state is positively related to the dispersion of private interpretation of information or private opinion of valuation of assets in the treasury market. Also PIN has predictive power for divergence of investor opinion in the U.S. Treasury market especially on non-announcement days.

References

Ahn, S., and M. Melvin, 2007, "Exchange Rate and FOMC Days" *Journal of Money, Credit and Banking*,39,1245-1266

Andersen, T. G., 1996, "Return volatility and trading volume: An information flow interpretation of stochastic volatility," *Journal of Finance*, 51, 169–204.

Andersen, T., and T. Bollerslev, 1998, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *Journal of International Economic Review*,39,885905

Back, K., Cao, H. and Willard, G., 2000, "Imperfect Competition Among Informed Traders," *Journal of Finance*, 55, 2117-2155.

Balduzzi, P., E. Elton, and C. Green, 2001, "Economic News and Bond Prices: Evidence From the U.S. Treasury Market," *Journal of Financial and Quantitative Analysis*, 36, 523-543.

Beber, A., F. Breedon and A. Baraschi, 2008, "Difference in Beliefs and Currency Options Markets " *Working Paper*,HEC - University of Lausanne

Brandt, T., and K. Kavajecz, 2004, "Price Discovery in the U.S. Treasury Market: The Impact of Order Flow and Liquidity on the Yield Curve" *Journal of Finance*,59,2623-2654

Clark, B. K., 1973, "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica*, 41, 135–155.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance*,57, 21852221.

Ederington, L.H. and J.H. Lee (1993), "How markets process information: news releases and volatility, *Journal of Finance* 48, 1161-1189.

Fleming,M., 1997, "The Round-the-Clock of U.S. Treasury Securities," *Economic Policy Review*, Federal Reserve Bank of New York 3, 9-32.

Fleming, M. and E. Remolona, 1997, "What Moves the Bond Market?" *Federal Reserve Bank of New York Economic Policy Review*, 3(4), 31–50

Foster, F. and S. Viswanathan, 1990, "A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Securities Markets," *Review of Financial Studies*, 3, 593-624.

Glosten, L., (1994), Is the Electronic Open Limit Order Book Inevitable? *Journal of Finance* 49, 11271161.

Green,T. C., 2004, "Economic News and the Impact of Trading on Bond Prices," *Journal of Finance*, 59, 1201–1233.

Kaniel, R., and H. Liu, 2006, "So What Orders Do Informed Traders Use?, *Journal of Business*, 79, 1867913.

Kyle, A., 1985, "Continuous Auctions and Insider Trading," *Econometrica* 53, 13151335.

Madhavan, A. and M. Cheng, 1997 "In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets," *Review of Financial Studies*, 10, 175-203.

Menkveld, A., A. Sarkar and M. van der Wel, 2008, "Customer Flow, Intermediaries and the Discovery of Equilibrium Risk-Free Rate,," *Working paper, Federal Reserve Bank of New York.*

Naes, R., and J. Skjeltorp, 2006, "Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market" *Journal of Financial Markets*, 9, 408-432–2019.

Pasquariello, P. and C. Vega, 2007, "Informed and Strategic Order Flow in the Bond Markets," *Review of Financial Studies*, 20, 1975–2019.

Sager, M., and M. Taylor, 2004, "The impact of European Central Bank Governing Council announcements on the foreign exchange market: a microstructural analysis, " *Journal of International Money and Finance*,23,1043-1051

Table 1. Summary Statistics of Market Activities

This table reports the summary statistics of daily trading volume ($ billions), daily return volatility (%) of 5-minute returns based on the mid bid-ask quote from 7:00 a.m. to 5:00 p.m., trading durations (seconds), relative spread ($\times 10,000$) and spread in ticks, average depth at the best bid and ask ($ millions), average depth in the entire order book ($ millions), average hidden depth at the best bid and ask ($ millions), and average hidden depth in the entire book during the sample period from 2005 to 2006. Spread and depth variables are averaged over 5-minute intervals of the trading day.

Variable	Mean	Median	StDev	Max	Min	Shewness	Kurtosis
Panel A: 2-year note)							
Spread (in ticks)	1.06	1.05	0.05	1.59	0.99	4.50	39.24
Relative spread ($\times 10,000$)	0.83	0.83	0.04	1.29	0.78	5.02	47.35
Trading volume ($ billions)	27.45	26.55	10.12	79.50	6.05	0.97	5.08
Trading durations (seconds)	15.99	14.61	6.76	48.21	3.48	0.98	4.09
Return volatility (%)	0.07	0.06	0.03	0.28	0.03	2.61	13.60
Depth at the best bid and ask ($ mil)	637.72	593.14	254.17	1567.41	190.25	0.44	2.46
Hidden depth at the best bid and ask($mil)	32.64	25.77	22.56	173.68	1.82	2.04	10.21
Depth of the entire order book ($ mil)	5122.56	4227.90	2416.23	10305.34	899.38	0.34	1.77
Hidden depth of the entire order book ($ mil)	99.83	81.71	73.53	526.09	9.25	2.04	9.08
Panel C: 5-year note							
Spread (in ticks)	1.18	1.16	0.10	2.30	1.04	4.65	42.55
Relative spread ($\times 10,000$)	0.93	0.92	0.08	1.87	0.83	4.93	47.01
Trading volume ($ billions)	24.69	24.17	7.48	50.31	7.71	0.55	3.36
Trading durations (seconds)	6.74	6.02	3.13	23.94	2.20	1.41	5.97
Return volatility (%)	0.17	0.15	0.06	0.45	0.07	1.71	6.90
Depth at the best bid and ask ($ mil)	119.30	118.22	33.46	213.12	54.86	0.47	2.71
Hidden depth at the best bid and ask($mil)	6.83	5.90	4.25	39.37	0.22	1.90	10.92
Depth of the entire order book ($ mil)	1238.48	1154.73	485.39	2522.77	442.96	0.43	2.01
Hidden depth of the entire order book ($ mil)	40.36	29.48	133.01	2885.68	4.18	20.66	441.77
Panel D: 10-year note							
Spread (in ticks)	1.13	1.11	0.07	1.82	0.99	3.27	28.19
Relative spread ($\times 10,000$)	1.79	1.77	0.11	2.93	1.60	3.16	25.69
Trading volume ($ billions)	22.76	22.62	6.93	43.68	5.32	0.38	2.84
Trading durations (seconds)	6.59	5.59	3.35	22.49	2.23	1.32	4.82
Return volatility (%)	0.29	0.26	0.10	0.77	0.11	1.67	7.43
Depth at the best bid and ask ($ mil)	120.93	118.37	32.11	227.99	50.96	0.55	3.10
Hidden depth at the best bid and ask($mil)	5.50	4.82	3.24	28.60	0.88	2.12	11.88
Depth of the entire order book ($ mil)	1520.08	1376.26	657.52	3459.07	439.77	0.75	2.69
Hidden depth of the entire order book ($ mil)	36.43	31.22	24.07	233.61	2.52	2.88	20.97

Table 2. Macroeconomic News with Pre-Scheduled Announcements

This table reports the list of macroeconomic news included in our analysis. N denotes the total number of announcements during the period from January 2005 to December 2006. Day and Time denote, respectively, the weekday or day of the month and time (EST) of announcement. σ_{surprise} denotes the standard deviation of announcement surprises. $N_{|\text{surprise}|>k\sigma_{\text{surprise}}}$ denotes the number of announcements where the announcement surprise is more than k standard deviation.

News/Event	N	Day	Time	σ_{surprise}	$N_{\|\text{surprise}\|>\sigma_{\text{surprise}}}$	$N_{\|\text{surprise}\|>2\sigma_{\text{surprise}}}$
Business Inventories	24	Around the 15th of the month	8:30[a]	0.002	5	1
Capacity Utilization	24	Two weeks after month end	9:15	0.003	6	1
Change in Nonfarm Payrolls	24	First Friday of the month	8:30	59.228	9	0
Chicago PMI	24	Last business day of the month	10:00	5.094	8	1
Construction Spending	24	Two weeks after month-end	10:00	0.245	1	1
Consumer Confidence	24	Last Tuesday of Month	10:00	3.860	6	2
Consumer Credit	24	5th business day of the month	15:00	125.82	1	1
Consumer Price Index	24	Around the 13th of the month	8:30	0.002	9	0
Current Account	8	10 to 11 weeks after quarter-end	8:30	7.687	2	0
Durable Orders	24	Around the 26th of the month	8:30	0.031	9	1
Economic outlook	6	According to schedule	10:00[b]	n.a.	n.a.	n.a.
Existing Home Sales	24	On the 25th of the month	10:00	0.160	7	1
FOMC Minutes	16	Thursday following the next FOMC meeting date	14:15	n.a.	n.a.	n.a.
FOMC rate decision expected	16	According to schedule	14:10	0.000	0	0
Factory Orders	24	Around the first business day of the month	10:00	0.006	6	2
GDP Advance	8	3rd / 4th week of the month for prior quarter	8:30	0.006	1	1
GDP Final	8	3rd / 4th week of second month following the quarter	8:30	0.001	4	1
GDP Preliminary	8	3rd / 4th week of first month following the quarter	8:30	0.003	3	0
Housing Starts	24	Two or three weeks after the reporting month	8:30	124.26	6	2

Event	N	Day	Time	σ_{surprise}	$N_{\lvert\text{surprise}\rvert > \sigma_{\text{surprise}}}$	$N_{\lvert\text{surprise}\rvert > 2\sigma_{\text{surprise}}}$
ISM Services	24	On the third business day of the month	10:00	2.834	9	1
ISM index	24	First business day of the month	10:00	2.332	6	1
Industrial Production	24	Around the 15th of the month	9:15	0.003	8	1
Initial Jobless Claims	104	Thursday weekly	8:30	17.499	23	6
Leading Indicators	24	around the first few business days of the month	8:30	0.002	7	2
Monthly Treasury Budget	24	about the third week of the month	14:00	5.239	4	2
NY Empire State Index	24	15th/16th of the month	8:30	9.738	9	1
New Home Sales	24	Around the last business day of the month	10:00	92.492	6	2
PCE	24	Around the first business day of the month	8:30	0.046	5	5
Personal Income	24	Around the first business day of the month	8:30	0.003	2	1
Philadelphia Fed	24	Third Thursday of the month	12:00	7.960	10	0
Producer Price Index	24	Around the 11th of each month	8:30	0.307	2	1
Retail Sales	24	Around the 12th of the month	8:30	0.121	1	1
Semiannual Monetary Policy Report	4	February and July annually	10:00	n.a.	n.a.	n.a.
Trade Balance	24	Around the 20th of the month	8:30	3.225	7	1
ADP National Employment Report	8	2 days before Change in Nonfarm Payrolls	8:15	n.a.	n.a.	n.a.

[a] – Business inventories are announced at either 10:00 a.m. or 8:30 a.m. During 2005-2006, 17 announcements took place at 10:00 a.m. and 7 announcements took place at 8:30 a.m.

[b] – One testimony of Economic Outlook is released at 14:30 on 5th June, 2006.

Table 3. Estimation Results of the Markov Switch Model

This table reports the estimation results of the Markov switch model in equation (8):

$$\Delta p_t = \rho_p \Delta p_{t-1} + \mu_p + \mu_{p,Inf} * I_t + \beta_{OF} OF_t + \beta_{OF,Inf} * OF_t * I_t + \varepsilon_t$$

with $\varepsilon_t \sim N(0, \sigma_p + \sigma_{p,Inf} * I_t)$ and I_t is equal to 1 if information arrives at the market and 0 otherwise. Standard errors are reported in the brackets beneath the coefficient estimates.

Maturity	ρ_p	μ_p	$\mu_{p,HVol}$	σ_p	$\sigma_{p,inf}$	β_{OF}	$\beta_{OF,inf}$	p_{11}	p_{22}	likelihood
2-year										
	0.006	0.034	-0.075	0.417	1.193	0.014	0.010	0.979	0.731	
	(0.002)	(0.027)	(0.004)	(0.002)	(0.026)	(0.000)	(0.001)	(0.001)	(0.012)	-38178.0
5-year										
	0.044	0.032	0.006	0.872	2.545	0.021	0.015	0.969	0.757	
	(0.004)	(0.045)	(0.004)	(0.005)	(0.046)	(0.000)	(0.001)	(0.001)	(0.010)	-81970.7
10-year										
	0.055	0.001	0.003	1.533	3.417	0.037	0.025	0.968	0.807	
	(0.008)	(0.064)	(0.004)	(0.009)	(0.061)	(0.000)	(0.001)	(0.001)	(0.008)	-112802.5

Table 4. Summary Statistics of the Probability of Informed State (PIN)

This table reports the summary statistics of the estimates of the probability of informed state (PIN) based on the Markov switch model as in equation (8).

Maturity	Mean	Median	StDev	Maximum	Minimum	Skewness	Kurtosis	rho1	rho2	rho3
2-year	0.073	0.017	0.142	0.988	0.007	0.486	0.322	0.305	0.247	0.230
5-year	0.113	0.026	0.179	0.998	0.010	0.582	0.454	0.418	0.357	0.326
10-year	0.140	0.039	0.199	1.000	0.013	0.639	0.505	0.454	0.402	0.380

Table 5. PIN Estimates and Announcement Surprises

This table reports the regression results of PIN estimates against announcement surprises of Nonfarm Payroll (Nonfarm), Consumer Confidence Index (C.Confi.), ISM Index (ISM), Initial Jobless Claims (Ini.Jbls.), Leading Indicators (Leading), New Home Sales(NewHome) and Retail Sales (Retail) at time t. Standard errors are reported in the brackets beneath the coefficient estimates.

Maturity	α	$NonFarm$	$C.Confi.$	ISM	$Ini.Jbls.$	$Leading$	$NewHome$	$Retail$	R^2
2-year	0.073	0.555	0.219	0.341	0.219	0.045	0.149	5.951	
	(*0.000*)	(*0.005*)	(*0.004*)	(*0.004*)	(*0.002*)	(*0.004*)	(*0.005*)	(*0.100*)	*0.11*
5-year	0.117	0.623	0.259	0.276	0.275	0.073	0.170	7.853	
	(*0.000*)	(*0.007*)	(*0.007*)	(*0.007*)	(*0.003*)	(*0.007*)	(*0.007*)	(*0.158*)	*0.10*
10-year	0.145	0.563	0.251	0.343	0.266	0.003	0.191	8.132	
	(*0.000*)	(*0.009*)	(*0.009*)	(*0.009*)	(*0.004*)	(*0.008*)	(*0.009*)	(*0.196*)	*0.08*

Table 6. PIN, Volatility and Liquidity Variables

Table 6 reports the relation between PIN estimates and volatility as well as liquidity variables. We sort the PIN estimates in each 30-minute interval into 5 equal groups (quintiles). For each group, we then calculate and report the mean absolute return ($|ret|$), depth at the best quotes (dep0), depth behind the best quotes (depbhd) and trading volume (trdqn).

	PIN_t	$\|ret_t\|$	$spread_t$	$spread_{t-1}$	$dep0_t$	$depbhd_t$	$trdqn_t$
Panel A: 2-year note							
Q1(lowest)	0.0078	0.0386	0.0082	0.0081	703.88	4170.73	677.91
Q2	0.0117	0.0716	0.0083	0.0081	656.64	4321.32	1047.9
Q3	0.0173	0.0803	0.0083	0.0082	629.19	4265.82	1191.02
Q4	0.0362	0.1	0.0082	0.0082	655.46	4374.64	1571.44
Q5	0.2979	0.2017	0.0083	0.0087	615.28	4376.48	2967.56
Panel B: 5-year note							
Q1(lowest)	0.012	0.1112	0.0089	0.0089	122.71	1025.64	635.31
Q2	0.0161	0.1752	0.0092	0.009	126.5	1163.01	987.18
Q3	0.0276	0.2299	0.0094	0.0089	125.42	1161.87	1246.83
Q4	0.0927	0.2919	0.0092	0.0089	121.97	1123.31	1519.38
Q5	0.43	0.5323	0.0092	0.0103	110.27	1025.78	2353.48
Panel C: 10-year note							
Q1(lowest)	0.0162	0.2058	0.0173	0.0172	126.98	1231.61	570.83
Q2	0.0227	0.3023	0.0175	0.0174	126.71	1433.63	863.56
Q3	0.0417	0.4247	0.0178	0.0173	126.07	1492.84	1140.7
Q4	0.1324	0.5327	0.0175	0.0174	122.63	1445.43	1417.59
Q5	0.5005	0.88	0.0175	0.0187	114.65	1363.63	2219.98

Table 7. Non-parametric Test of Permanent Price Impact

This table reports the p-value ($\times 100$) of the nonparametric test of permanent price impact when PIN estimate is significantly larger than the median value over the trading day. The test is based on the following statistic:

$$1 - N[\frac{n_{PIN} - n \cdot P_{highPI}}{\sqrt{n \cdot P_{highPI}(1 - P_{highPI})}}] \tag{13}$$

where N is the standard normal cumulative distribution function, n_{PIN} is the number of same direction mid-quotes over the next 1-hour, 1-day, or 1-week following significant PIN, and $P_{highPIN}$ is the fraction of times that PIN is significant and n be the total number of quotes in the same direction in 1 hour, 1 day or 1 week following significant and insignificant PIN.

	After 1 hour	After 1 day	After 1 week
Panel A: All Days			
2-year note	0.0000	0.0000	0.0000
5-year note	0.0000	0.0000	0.3162
10-year note	0.0000	0.0000	0.0403
Panel B: Announcement Days			
2-year note	0.0000	0.0000	0.0001
5-year note	0.0000	0.0000	1.0311
10-year note	0.0000	0.0001	0.1443
Panel C: No Announcement Days			
2-year note	0.0702	0.8332	3.3010
5-year note	6.4758	1.3685	7.7311
10-year note	2.3113	4.1565	6.8959

Table 8. The Probit Model of Permanent Price Impact

This table reports the estimation results of the Probit model for the permanent price impact.

$$P(state_{t+h} = 1) = f(\alpha + \rho_{ann} I(state_{t,ann}) + \rho_{noann} I(state_{t,noann}) + \beta_{depth} depth + \beta_{trdqn} Trdqn + \beta_{sprd} sprd + \gamma_{ann} D_{ann})$$

where $P(state_{t+h} = 1)$ if $R_{[t-1,t+h]}$ is in the same direction as $R_{[t-1,t]}$, where h is 1 hour, 1 day or 1 week. Otherwise, $P(state_{t+h}) = 0$. The indicator variable $I(state_{t,ann(noann)})$ takes the value of 1 if $PIN_{t,ann(noann)}$ is significantly larger than median in ann ($noann$) an announcement day (non-announcement days) and 0 otherwise. The variable $depth$ is overall market depth on the book. $Trdqn$ is the trading volume and $sprd$ is the bid-ask spread at time t. The variable D_{ann} takes the value of 1 if it is an announcement day and 0 otherwise.

	After 1-hour			After 1-day			After 1-week		
	Est	Std	p-value	Est	Std	p-value	Est	std	p-value
Panel A: 2-year note									
Constant	-399.782	0.242	0.0000	-2163.170	0.241	0.0000	-1304.244	0.245	0.0000
ρ_{ann}	399.664	0.286	0.0000	2163.107	0.272	0.0000	1303.777	0.276	0.0000
ρ_{nonews}	399.612	0.473	0.0000	2163.893	0.465	0.0000	1305.092	0.472	0.0000
β^{depth}_{ann}	-0.045	0.015	0.0026	-0.009	0.014	0.5363	-0.016	0.014	0.2660
β^{depth}_{noann}	0.026	0.025	0.3021	-0.051	0.025	0.0384	-0.033	0.025	0.1784
β^{volume}_{ann}	0.141	0.020	0.0000	0.066	0.017	0.0001	0.047	0.016	0.0037
β^{volume}_{noann}	0.063	0.059	0.2864	0.019	0.058	0.7388	0.018	0.058	0.7528
β^{spread}_{ann}	45.658	29.184	0.1177	-2.631	24.155	0.9133	31.590	24.408	0.1956
β^{spread}_{noann}	21.537	80.340	0.7886	-74.219	79.701	0.3517	-111.799	81.400	0.1696
γ_{ann}	0.636	0.696	0.3602	0.585	0.371	0.1150	0.147	0.292	0.6156
likelihood	-961.679			-1063			-1063		
Panel B: 5-year note									
Constant	-89.233	0.178	0.0000	-197.064	0.168	0.0000	-177.586	0.169	0.0000
ρ_{ann}	89.901	0.194	0.0000	197.179	0.184	0.0000	177.521	0.185	0.0000
ρ_{nonews}	88.950	0.346	0.0000	197.025	0.327	0.0000	178.117	0.327	0.0000
β^{depth}_{ann}	-0.282	0.070	0.0001	-0.130	0.066	0.0480	-0.208	0.065	0.0014
β^{depth}_{noann}	0.014	0.114	0.9039	-0.046	0.110	0.6761	-0.007	0.109	0.9475
β^{volume}_{ann}	0.202	0.031	0.0000	0.126	0.027	0.0000	0.114	0.026	0.0000
β^{volume}_{noann}	0.092	0.082	0.2631	-0.011	0.076	0.8805	-0.113	0.076	0.1381
β^{spread}_{ann}	-17.184	9.585	0.0730	-3.461	9.341	0.7110	6.783	9.872	0.4920
β^{spread}_{noann}	67.134	51.999	0.1967	29.197	49.340	0.5540	-38.126	49.391	0.4402
γ_{ann}	-0.235	0.368	0.5229	0.052	0.317	0.8705	-0.418	0.291	0.1497
likelihood	-1153			-1347			-1374		
Panel B: 10-year note									
Constant	-66.803	0.177	0.0000	-320.440	0.168	0.0000	-458.033	0.167	0.0000
ρ_{ann}	67.380	0.203	0.0000	320.437	0.194	0.0000	458.011	0.193	0.0000
ρ_{nonews}	66.778	0.341	0.0000	320.173	0.322	0.0000	457.596	0.321	0.0000
β^{depth}_{ann}	-0.170	0.043	0.0001	-0.089	0.041	0.0291	-0.056	0.040	0.1656
β^{depth}_{noann}	-0.056	0.084	0.5076	-0.138	0.081	0.0860	-0.098	0.081	0.2260
β^{volume}_{ann}	0.175	0.030	0.0000	0.163	0.028	0.0000	0.061	0.026	0.0190
β^{volume}_{noann}	0.218	0.083	0.0088	0.050	0.076	0.5106	0.110	0.076	0.1482
β^{spread}_{ann}	-6.856	8.188	0.4024	-0.624	7.994	0.9378	0.906	7.959	0.9094
β^{spread}_{noann}	15.064	26.195	0.5652	30.646	25.055	0.2213	25.583	24.800	0.3023
γ_{ann}	-0.271	0.339	0.4247	-0.343	0.296	0.2467	-0.326	0.288	0.2587
likelihood	-1304			-1473			-1511		

Table 9. Informed State and Price Movements

This table reports returns in 1 hour ($R_{[t-1,t+1hr]}$), 1 day ($R_{[t-1,t+1dy]}$) and 1 week ($R_{[t-1,t+1wk]}$) following positive (negative) price change given PIN is significant. $R_{[t-1,t+h]}$ is the permanent price impact, defined as the logarithmic return from $t-1$ to $t+j$, where j is 1 hour, 1 day or 1 week interval following significant PIN. $R_{[t-1,t]}$ is the logarithmic return from $t-1$ to t at significant PIN, PIN^*. All logarithmic return are multiplied by 1,000. Standard errors are reported in the brackets beneath the coefficient estimates.

	$R_t > 0 \| PIN^*$			$R_t < 0 \| PIN^*$		
Bond	$R_{[t-1,t+1hr]}$	$R_{[t-1,t+1dy]}$	$R_{[t-1,t+1wk]}$	$R_{[t-1,t+1hr]}$	$R_{[t-1,t+1dy]}$	$R_{[t-1,t+1wk]}$
Panel A: All Days						
2-year	0.250	0.197	0.016	-0.248	-0.251	-0.230
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
5-year	0.557	0.510	0.477	-0.525	-0.541	-0.526
	(0.001)	(0.003)	(0.006)	(0.001)	(0.003)	(0.007)
10-year	0.890	0.843	0.730	-0.868	-1.107	-0.998
	(0.002)	(0.005)	(0.010)	(0.001)	(0.004)	(0.009)
Panel B: Announcement Days						
2-year	0.282	0.195	-0.038	-0.271	-0.322	-0.287
	(0.001)	(0.002)	(0.004)	(0.001)	(0.002)	(0.004)
5-year	0.622	0.394	0.191	-0.567	-0.843	-0.947
	(0.002)	(0.004)	(0.008)	(0.001)	(0.004)	(0.008)
10-year	0.955	0.941	0.721	-0.927	-1.210	-1.261
	(0.003)	(0.006)	(0.012)	(0.002)	(0.006)	(0.011)
Panel C: Non-announcement Days						
2-year	0.135	0.201	0.210	-0.169	-0.006	-0.034
	(0.001)	(0.006)	(0.017)	(0.001)	(0.006)	(0.018)
5-year	0.327	0.924	1.499	-0.352	-0.246	0.215
	(0.003)	(0.017)	(0.032)	(0.003)	(0.016)	(0.035)
10-year	0.649	0.480	0.764	-0.654	-0.735	-0.051
	(0.005)	(0.014)	(0.050)	(0.005)	(0.016)	(0.054)

Table 10. PIN and Asymmetric Opinion – Contemporaneous Relation

This table reports how PIN is related to dispersion in private opinion contemporaneously. We sort PIN estimates in each 30-minute interval into 5 equal groups (quintiles). For each group, we then calculate and report the mean estimates of PIN ($\overline{PIN_t}$), the number of significant PIN (N_t^*) and the slope of the order book Slp_t.

All Sample			Announcement Days			No Announcement Days		
$\overline{PIN_t}$	N_t^*	Slp_t	$\overline{PIN_t}$	N_t^*	Slp_t	$\overline{PIN_t}$	N_t^*	Slp_t
Panel A: 2-year note								
0.0078	0	13874.9	0.0079	0	13848	0.0078	0	13945.2
0.0117	0	13585.6	0.0117	0	13574.9	0.0117	0	13614.7
0.0173	56	13501	0.0173	29	13473.4	0.0173	27	13591
0.0362	154	13622.6	0.0365	80	13614.1	0.0353	74	13651.1
0.2979	1313	13406.5	0.3106	1076	13411	0.235	237	13384.3
Panel A: 5-year note								
0.012	0	9420.3	0.012	0	9351.7	0.012	0	9590.1
0.0161	5	9450.6	0.0161	0	9455	0.0161	5	9437.3
0.0276	39	9445.8	0.0274	7	9430.6	0.0283	32	9491.1
0.0927	336	9341.5	0.0945	215	9357.6	0.0869	121	9291.5
0.43	1573	9205.7	0.4404	1316	9224.9	0.3727	257	9100.1
Panel A: 10-year note								
0.0162	0	4842.2	0.0163	0	4812.6	0.0161	0	4919.9
0.0227	0	4846.8	0.0228	0	4835.3	0.0226	0	4878.8
0.0417	44	4828.8	0.0419	16	4834.8	0.041	28	4810.3
0.1324	514	4808.1	0.1337	344	4816.4	0.1286	170	4781.8
0.5005	1581	4714.4	0.5082	1330	4705	0.4572	251	4767.2

Table 11. Does PIN Predict Asymmetric Opinion?

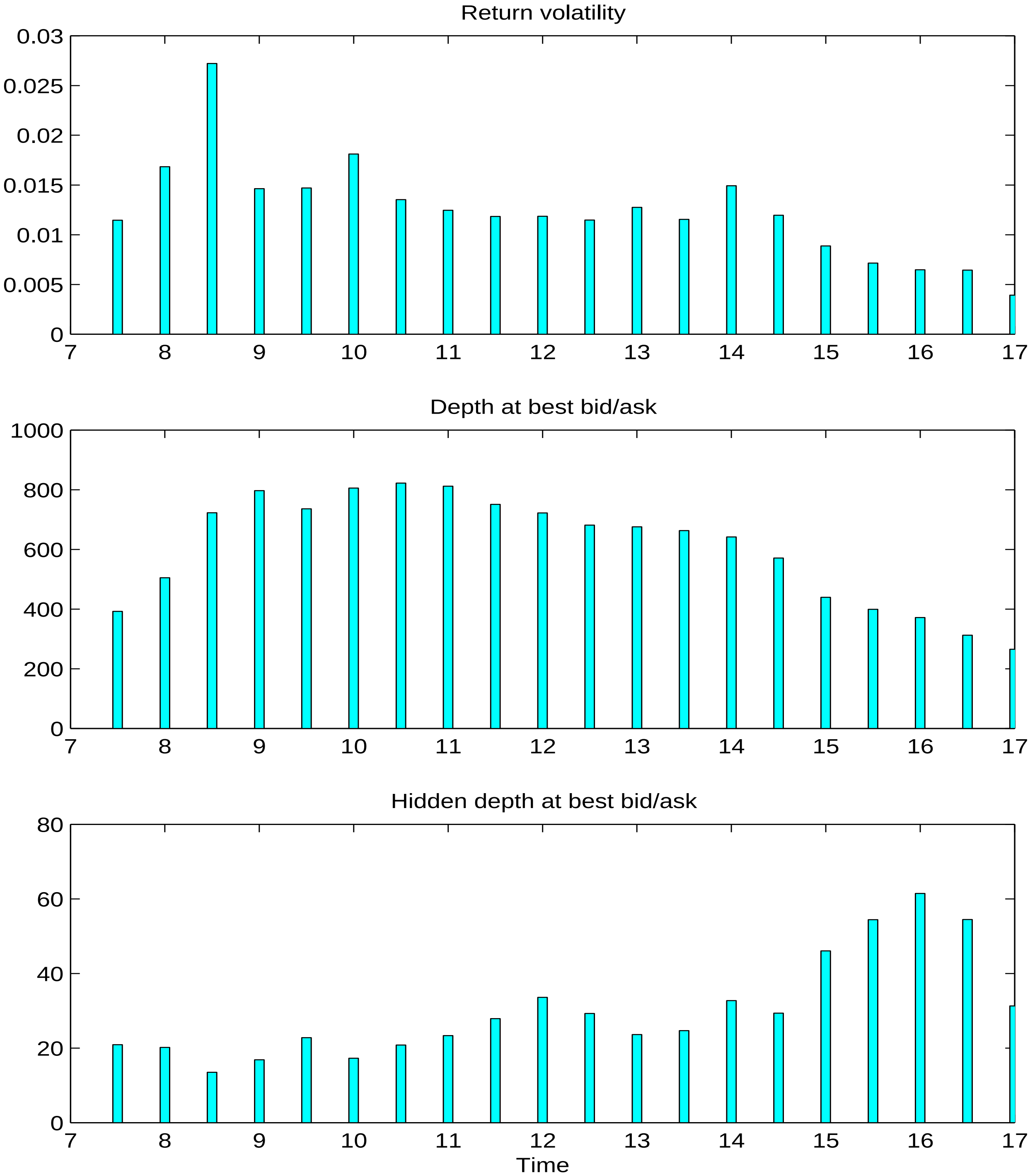
This table reports how PIN is related to future dispersion in private opinion. We sort PIN estimates in each 30-minute interval into 5 equal groups (quintiles). For each group, we then calculate and report the mean estimates of PIN ($\overline{PIN_t}$), the number of significant PIN (N_t^*) and the future slope of the order book Slp_{t+1}.

All Sample			Announcement Days			No Announcement Days		
$\overline{PIN_t}$	lag N_t^*	Slp_{t+1}	$\overline{PIN_t}$	lag N_t^*	Slp_{t+1}	$\overline{PIN_t}$	lag N_t^*	Slp_{t+1}
Panel A: 2-year note								
0.0081	0	13782.7	0.0080	0	13762.6	0.0081	0	13834.7
0.0124	0	13569.3	0.0120	0	13535.5	0.0123	0	13665.3
0.0183	73	13479.4	0.0180	36	13435.2	0.0183	37	13624.1
0.0382	161	13574.4	0.0390	83	13563.3	0.0370	78	13611.3
0.3020	1336	13584.6	0.3160	1088	13603.0	0.2381	248	13497.6
Panel A: 5-year note								
0.0125	0	9446.8	0.0130	0	9385.4	0.0124	0	9600.8
0.0174	9	9437.6	0.0170	0	9399.2	0.0173	9	9556.5
0.0300	46	9419.1	0.0300	8	9409.8	0.0306	38	9446.2
0.0980	377	9316.9	0.0990	239	9371.2	0.0942	138	9149.8
0.4339	1582	9243.6	0.4440	1321	9252.1	0.3785	261	9196.8
Panel A: 10-year note								
0.0169	0	4856.3	0.0170	0	4834.4	0.0168	0	4911.7
0.0250	5	4848.4	0.0250	2	4826.7	0.0250	3	4911.6
0.0452	51	4806.3	0.0450	18	4805.0	0.0451	33	4810.2
0.1391	554	4788.0	0.1400	372	4785.4	0.1367	182	4796.8
0.5043	1587	4741.1	0.5120	1331	4748.0	0.4634	256	4702.1

FIGURE 1

Intraday Market Activities

This figure plots market activities in each half-hour window during the day from 7:30 to 17:00. Variables include trading volume ($ millions), trading duration (seconds), relative bid-ask spread ($\times 10,000$), return volatility (%) calculated from 5-minute returns based on the mid bid-ask quote, average depth at the best bid/ask ($ millions) calculated over each 5-minute interval, and average hidden depth at the best bid/ask ($ millions) calculated over each 5-minute interval.



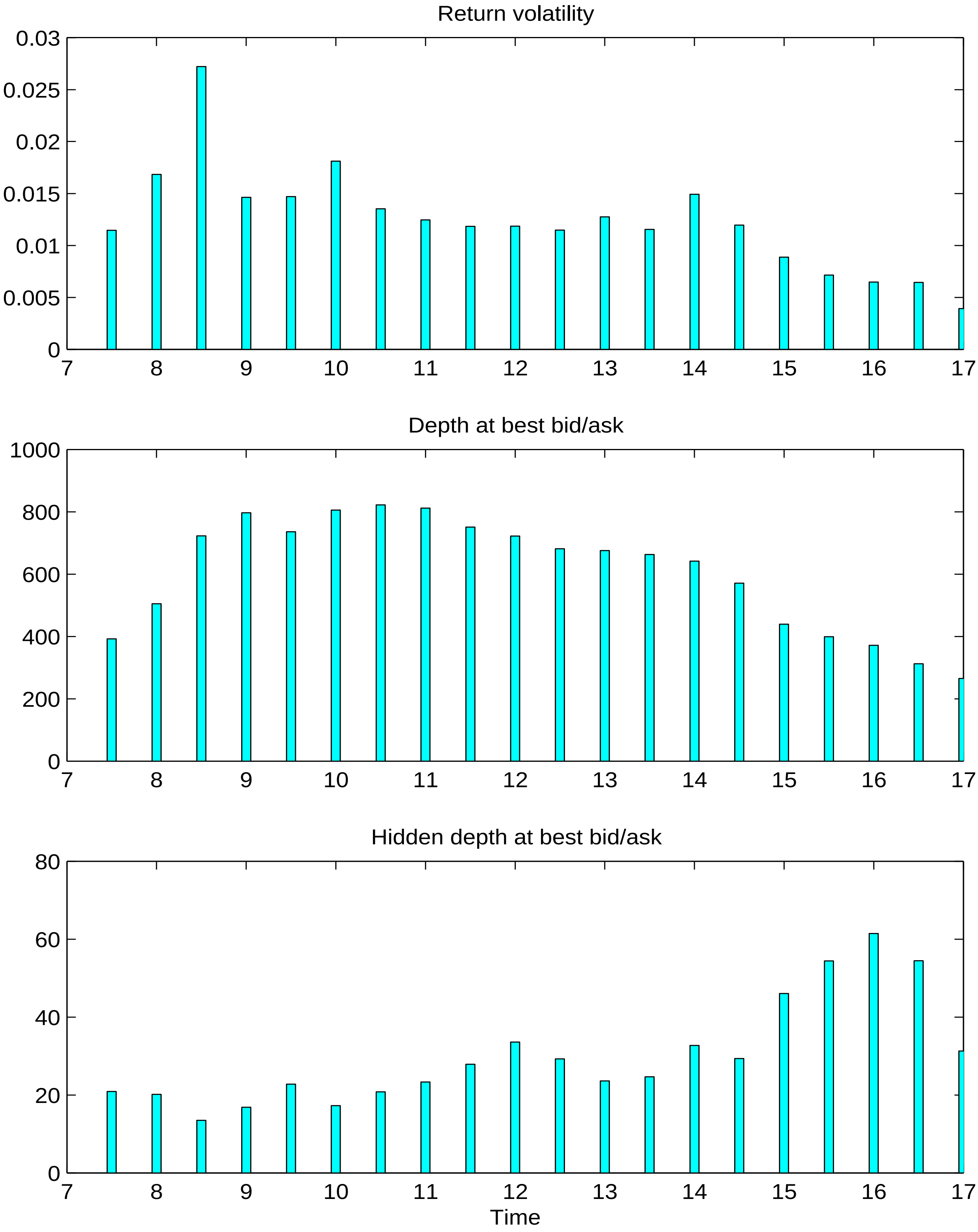

Return volatility
0.03
0.025
0.02
0.015
0.01
0.005
0
7 8 9 10 11 12 13 14 15 16 17
Depth at best bid/ask
1000
800
600
400
200
0
7 8 9 10 11 12 13 14 15 16 17
Hidden depth at best bid/ask
80
60
40
20
0
7 8 9 10 11 12 13 14 15 16 17
Time

FIGURE 2

Intraday Plots of Probability of Informed Sate (PIN)

This figure plots the average estimates of the probability of informed sate (PIN) in each half-hour window during the day from 7:30 to 17:00, based on both the basic model and extended model. The intraday patterns are plotted for (a) all days in the sample, (b) days with announcements and (c) days without announcements.