## *Essays on the Microstructure of Informed Trading and Hidden Orders in Futures Markets*

PHUENSANE, PONGSUTTI

# Essays on the Microstructure of Informed Trading and Hidden Orders in Futures Markets

Pongsutti Phuensane

Durham University Business School

Durham University

A thesis submitted for the degree of

*Doctor of Philosophy*

2016

I herewith declare that I have produced this thesis without the prohibited assistance of third parties and without making use of aids other than those specified from other sources, and which have been identified as such. This thesis has not previously been presented in an identical or similar form to any other English or foreign examination board.

The Ph.D. work was conducted from January 2014 under the supervision of Prof. Dr. Julian M. Williams at Durham University.

Pongsutti Phuensane

Durham, United Kingdom

# Acknowledgements

This thesis is the result of my research as a doctoral student at the Durham University Business School, Durham University between January 2014 and May 2016. During my research, I had the privilege of cooperating with a number of people who I would like to express my sincere gratitude and warmest thanks.

First of all, I would like to sincerely thank my supervisor, Prof. Dr. Julian M. Williams for his patience, support and advice on my research topic on informed trading, high-frequency trading and algorithmic trading in the market microstructure area. I would like to especially thank him for introducing me to High Frequency Trading and a Matlab platform. I also thank Dr. Ayse Yazici, my second supervisor and Dr. Chulwoo Han my previous second supervisor, for many helpful discussions and comments. I appreciate the feedback I received from the discussants at the Spring 2015 Conference of the Multinational Finance Society.

My special thanks go to Dr. Juliana Malagon and Dr. Arz'e Karam for advising on chapter four of this thesis (examiners and discussants for this chapter), and Dr. Jing Nie for being a great friend whose encouragement animated me to embark on a Ph.D. thesis. Furthermore, I am indebted to my institutional sponsor, Faculty of Management Sciences, Khon Kaen University, Thailand, and in the particular finance department. My gratitude also goes to the open source community, especially LaTeXand MathWorks for generously making available programs, file exchanges and packages for research.

Most importantly, I would like to sincerely thank my family for their love and support throughout my Ph.D. journey.

# Abstract

The thesis seeks a better understanding of the market microstructure topic, evaluates an informed trading measurement tool that is available for econometricians, and attempts to measure the impact of the presence of hidden liquidity to market quality. The two favorable prevailing informed trading detection, PIN, and VPIN models are used to investigate the informed trading activity around LIBOR manipulation in the LIBOR reference futures market- Eurodollar futures. These two models show a significant performance as an early warning system, however, there is not statistically significant differences relative to the event in the long-run variation.

In addition to the chapter seven of this thesis, I examine the relationship between market quality and hidden liquidity, uncovering a strong positive effect from hidden order to market quality proxies.

# Contents

# List of Figures

# List of Tables

# Nomenclature

As far as possible a single coherent notation was used. Nevertheless, sometimes a certain variable may have dierent meanings.

| Symbol | Description |
|---|---|
| $\alpha$ | Probability of information event |
| $A$ | Ask price |
| $b_t$ | Buy order |
| $b$ | Bad News |
| $B$ | Bid price |
| $\delta$ | Probability of bad news |
| $g$ | Good news |
| $\mathbb{D}$ | Variance operator |
| $\mathbb{E}$ | Expectation operator |
| $F$ | Distribution function |
| $\widetilde{HDV}$ | Signed-Hidden Order Volume |
| $i$ | Trading period |
| $I$ | Trading days |
| I | Informed trader |
| $\kappa$ | A certain cost |

*Continued on next page*

| **Symbol** | **Description** |
|---|---|
| $\lambda$ | Illiquidity parameter |
| $\mathbb{L}$ | Lag operator |
| $\mathscr{L}$ | Log-Likelihood function |
| $L$ | Likelihood function |
| $\mu$ | Arrival rate of informed traders |
| $M$ | Data set of $(B_i, S_i)_{i=1}^{I}$ |
| $\nu$ | Residual |
| $n$ | No News |
| $\mathcal{N}$ | Normal Distribution |
| $OI$ | Order imbalance |
| $\pi$ | Profit |
| $P$ | Posterior expectation |
| $\mathbb{P}$ | Probability |
| $PIN$ | Probability of Informed Trading |
| $\widetilde{QR}$ | Quote Revision |
| $S$ | Value of asset |
| $\tilde{S}$ | Value of asset at the end of period |
| $\sigma$ | Variance |
| $s_t$ | Sell order |
| $\Sigma$ | Summation operator |
| $\tau$ | Risk aversion parameter |
| $\Theta$ | Parameter vector of $(\alpha, \delta, \varepsilon, \mu)$ |
| $\tau$ | Time bar |
| $t$ | The beginning of period |

*Continued on next page*

| **Symbol** | **Description** |
|---|---|
| $T$ | Time |
| $\widetilde{TD}$ | Signed-Trade Direction |
| $\widetilde{TV}$ | Signed-Trade Volume |
| $u$ | Order from uninformed traders |
| $\tilde{u}$ | Uninformed traders order |
| $U$ | Uninformed trader |
| $\mathscr{U}$ | Utility function |
| $\varepsilon$ | Arrival rate of uninformed traders |
| $V^B$ | Buy volume |
| $V^S$ | Sell volume |
| $x$ | Order from informed traders |
| $\theta$ | Payoff of the risky asset |
| $y$ | Market maker Price |
| $Y$ | Independent variable |

# Chapter 1

# Introduction

## 1.1 Market Microstructure

The Market Microstructure has become increasingly important on a financial markets efficiency and integrity which is devoted to theoretical, empirical, and experimental research. This study is concerned with the details of how exchange occurs in markets , including the role of information in the price discovery process; the definition, measurement, control, and determinants of liquidity and transaction costs; and their implications for the efficiency, welfare, and regulation of alternative trading mechanisms and market structures (Madhavan [2000]).

### 1.1.1 Price Formation Process

Whilst many economists of market microstructure theory focus on how specific trading mechanisms affect the price formation process, O'Hara [1995] focuses on the study of the process and outcomes of exchanging assets under a specific set of rules. Later, Easley et al. [1996] developed new models on the theoretical framework of Easley and O'Hara [1992] to estimate the level of information asymmetry in the market, called the probability of informed trading (PIN).

My first empirical analysis (chapter 4 and 5) evaluates the effectiveness of the probability of informed trading- PIN (Easley et al. [1996]) as the PIN has been studied as an early warning of unusual activity in financial markets. It is a measure of asymmetric information on a trading event between informed and uninformed trading, which produces a mixed discrete-time and continuous-time sequential model of the trading mechanism, in which trades emerge when three classifications of economic agents – namely market makers, informed traders and uninformed traders collaborate. The model assumes Poisson arrival rates for informed and uninformed traders and daily Bernoulli probabilities on the occurrence of news, and types of news including, good, bad or no news (Wei [2013]). This work creates a new empirical practice for evaluating the PIN by investigating the change around documented episodes of recorded manipulation of the LIBOR by using Eurodollar futures data. Also, I examine the PIN around maturity events to investigate its variation in Eurodollar futures trading over their lifecycle, from inception to maturity. The results indicate that the PIN could have been used as an early warning of unusual activity in the LIBOR reference rate, and anecdotally I can demonstrate that on specific dates identified by the CFTC and FSA the recursively estimated PIN reaches a peak over 1,000 basis points higher for certain near-delivery contracts than for others further away in the forward curve.

In chapter 6, the PIN and Volume-Synchronized Probability of Informed Trading (VPIN) are examined and compared. Then, I investigate the effectiveness of PIN and VPIN in determining changes in the information structure and order flow of a futures market around documented episodes of recorded manipulation of the reference rate, from the various publicly available regulatory reports. The fundamental mechanism behind VPIN was developed from PIN, that can be found in Easley et al. [2008]. In Easley et al. [2010] this model is developed by introducing time variations of the arrival rate of informed and uninformed traders in GARCH style by estimating the expected Order Imbalance for each of block traded. I find a very strong connection between PIN, VPIN, and time to maturity of the contract that is not fully explained by the time variation in activity in the market. However, an event study using both a new bootstrap approach and asymptotic standard error on the VPIN and PIN respectively around documented LIBOR manipulation cases has mixed results. For certain events we can see a substantial change in the average detected levels of PIN and VPIN; however, a cross-sectional analysis of all reported cases up to mid-2015, indicates no significant change in the PIN and VPIN for the contracts in our sample.

In chapter 7, I examine the impact of hidden order to other interested variables in the market quality study. In this chapter, I adapt the OLS analysis from Hasbrouck and Saar [2013] by considering the impact of hidden liquidity on market quality. In addition, I extend Hasbrouck [1991] and Dufour and Engle [2000] bivariate VAR model to the interaction between quote revisions, signed-trades, signed-trade volume and hidden liquidity. Then the impulse response function (IRF) of the VAR model is used to estimate the interaction impact between these variables. The IRF also shows the speed of decline for the response of the observed variables to a one standard deviation shock in other variables. In this analysis I have start with the new algorithm of hidden order detection for the limit order book using E-mini S&P500 data. My algorithm shows 43% of the trade volume is involved with invisible liquidity. It is found that price impact fell and market quality is improved with the presence of hidden order both during high and low-frequency trading periods. I use this measure to study the association between the hidden order and other observed market environments. The analysis finds aggressive hidden order activity when trading volume is increased.

### 1.1.2 High Frequency Trading & Big Data Analysis

For the last decade, the financial markets environment has been different in fundamental ways. Speed is one of the most important factors due to information gathering and the actions prompted by this information have created high volatility in the market (Hasbrouck and Saar [2013]). In comparison to this High-frequency trading evolution and traditional human control, I can pinpoint human operating time frames of minutes to time scales of microseconds of computer algorithms which respond at a pace 100 times faster than it would take for a human trader to blink. This high-frequency trading environment shows the fastest trade updated for Eurodollar futures is 500 microseconds (chapters 4, 5 and 6) and fastest trade updated for E-Mini S&P 500 futures is 60 microseconds (chapter 7). However, the average trade updated is 45 and 2.5 seconds for Eurodollar and E-Mini S&P 500 respectively. At these speeds, only the microstructure matters (O'Hara [2015]).This thesis utilizes the HFT comprehensive data set for trade, quote and limit order book (LOB) pulled from Thomson Reuters Tick History (TRTH) taped in millisecond time stamp.

As mentioned in the previous section, my data is recorded in millisecond time stamp which resulted in a tremendous data size compared to the recent literature. The raw data (uncompressed CSV files) comprises 489 GB of best bid, best offer, and trade data for all 40 Eurodollar futures contracts and 642 GB of the limit order book and trade data for 32 E-Mini S&P 500 futures contracts, directly retrieved from the CME tapes. The raw tapes were streamed into a new format 'hdf5' which provided a high-integrity medium for this amount of data. With compression, the raw Eurodollar futures data 489 GB was reduced to 89 GB and 642 GB of E-Mini S&P 500 was reduced to 143 GB of compressed hdf5 data stored in separate files by maturity date, and then stored on a solid state drive. In total, the data prepared for each data set was around three months, from downloading the raw data from TRTH to data cleaning procedures.

Table 1.1 presents an example of the Limit Order Book data of the E-Mini S&P 500 on May 05, 2013 (ESU3; Chapter 7), time between 22:43:06.283 and 22:43:18.809. This example visualizes the HFT data recorded in millisecond time stamp (tick-by-tick), on each observation row including 31 columns of bids, 31 columns of ask and 4 columns of trade. These sets of data include date vector for bid and ask, a ten deep level of bid price, bid volume, the number of traders for bid, ask price, ask volume, the number of traders for ask. Finally, there are four columns of date vector, price, volume and number of traders for trade. Also, Figure 1.1 illustrates the information in Limit Order Book of the E-Mini S&P 500 on September 6, 2013 (ESU3) for the ten deep level of prices, volumes, and number of traders.

These tremendous data sets of the Eurodollar and E-Mini S&P futures contain a variety of data types to uncover the hidden intraday patterns or abnormal trading. Whilst the PIN analysis usies tick-by-tick to estimate the daily PIN (Chapter 4, 5), the daily tick-by-tick data is separated into 50 blocks to calculate the VPIN (Chapter 6). Also, for chapter 7, the hidden order, the VAR, and the IRF are estimated using purely tick-by-tick data. This big data analytic has significant importance to my empirical analysis because the data of futures

**Table 1.1: The Limit Order Book data of the E-Mini S&P 500 on May 05, 2013 (ESU3), time between 22:43:06.283 Hours and 22:43:18.809 Hours**

| Ask | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time | Price | | | | | | | | | | Volume | | | | | | | | | | Number of Traders | | | | | | | | | |
| 05May13 | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 | L10 | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 | L10 | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 | L10 |
| 22:43:06.283 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 38 | 263 | 263 | 287 | 262 | 267 | 262 | 235 | 235 | 236 | 2 | 9 | 9 | 9 | 8 | 7 | 6 | 5 | 5 | 6 |
| 22:43:06.283 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 43 | 263 | 263 | 287 | 262 | 267 | 262 | 235 | 235 | 236 | 3 | 9 | 9 | 9 | 8 | 7 | 6 | 5 | 5 | 6 |
| 22:43:06.283 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 43 | 263 | 263 | 287 | 262 | 262 | 262 | 235 | 235 | 236 | 3 | 9 | 9 | 9 | 8 | 6 | 6 | 5 | 5 | 6 |
| 22:43:06.283 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 43 | 263 | 290 | 287 | 262 | 262 | 262 | 235 | 235 | 236 | 3 | 9 | 10 | 9 | 8 | 6 | 6 | 5 | 5 | 6 |
| 22:43:06.283 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 43 | 263 | 290 | 287 | 262 | 262 | 235 | 235 | 235 | 236 | 3 | 9 | 10 | 9 | 8 | 6 | 5 | 5 | 5 | 6 |
| 22:43:06.456 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 10 | 263 | 290 | 287 | 262 | 262 | 235 | 235 | 235 | 236 | 2 | 9 | 10 | 9 | 8 | 6 | 5 | 5 | 5 | 6 |
| 22:43:06.456 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 10 | 152 | 290 | 287 | 262 | 262 | 235 | 235 | 235 | 236 | 2 | 8 | 10 | 9 | 8 | 6 | 5 | 5 | 5 | 6 |
| 22:43:08.809 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 10 | 152 | 290 | 287 | 262 | 262 | 235 | 235 | 235 | 236 | 2 | 8 | 10 | 9 | 8 | 6 | 5 | 5 | 5 | 6 |
| 22:43:18.809 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 10 | 263 | 290 | 287 | 262 | 262 | 235 | 235 | 235 | 236 | 2 | 9 | 10 | 9 | 8 | 6 | 5 | 5 | 5 | 6 |
| 22:43:18.809 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 43 | 263 | 290 | 287 | 262 | 262 | 235 | 235 | 235 | 236 | 3 | 9 | 10 | 9 | 8 | 6 | 5 | 5 | 5 | 6 |
| 22:43:18.809 | 1604.75 | 1605 | 1605.25 | 1605.5 | 1605.75 | 1606 | 1606.25 | 1606.5 | 1606.75 | 1607 | 44 | 263 | 290 | 287 | 262 | 262 | 235 | 235 | 235 | 236 | 4 | 9 | 10 | 9 | 8 | 6 | 5 | 5 | 5 | 6 |

| Bid | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time | Price | | | | | | | | | | Volume | | | | | | | | | | Number of Traders | | | | | | | | | |
| 05May13 | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 | L10 | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 | L10 | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 | L10 |
| 22:43:06.283 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 48 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 4 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:06.283 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 48 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 4 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:06.283 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 48 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 4 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:06.283 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 48 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 4 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:06.283 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 48 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 4 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:06.456 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 48 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 4 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:06.456 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 48 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 4 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:08.809 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 49 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 5 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:18.809 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 49 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 5 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:18.809 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 49 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 5 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |
| 22:43:18.809 | 1604.25 | 1604 | 1603.75 | 1603.5 | 1603.25 | 1603 | 1602.75 | 1602.5 | 1602.25 | 1602 | 49 | 300 | 300 | 297 | 273 | 235 | 225 | 225 | 225 | 671 | 5 | 10 | 10 | 9 | 9 | 5 | 5 | 5 | 5 | 7 |

Note: This table presents an example of the Limit Order Book data of the E-Mini S&P 500 on May 05, 2013 (ESU3), time between 22:43:06.283 Hours and 22:43:18.809 Hours. During these twelve seconds, there are eleven bids and asks updated with a ten deep levels of the bid price, a bid volume, the number of traders for bid, ask price, ask volume and number of traders for ask. In total on May 05, 2013 which is the low-frequency trading period, the data recorded 66 columns and 6,830 rows. However, for the highest trading frequency on this contract appears on September 6, 2013, with the observation of 66 columns and 4,441,626 rows.

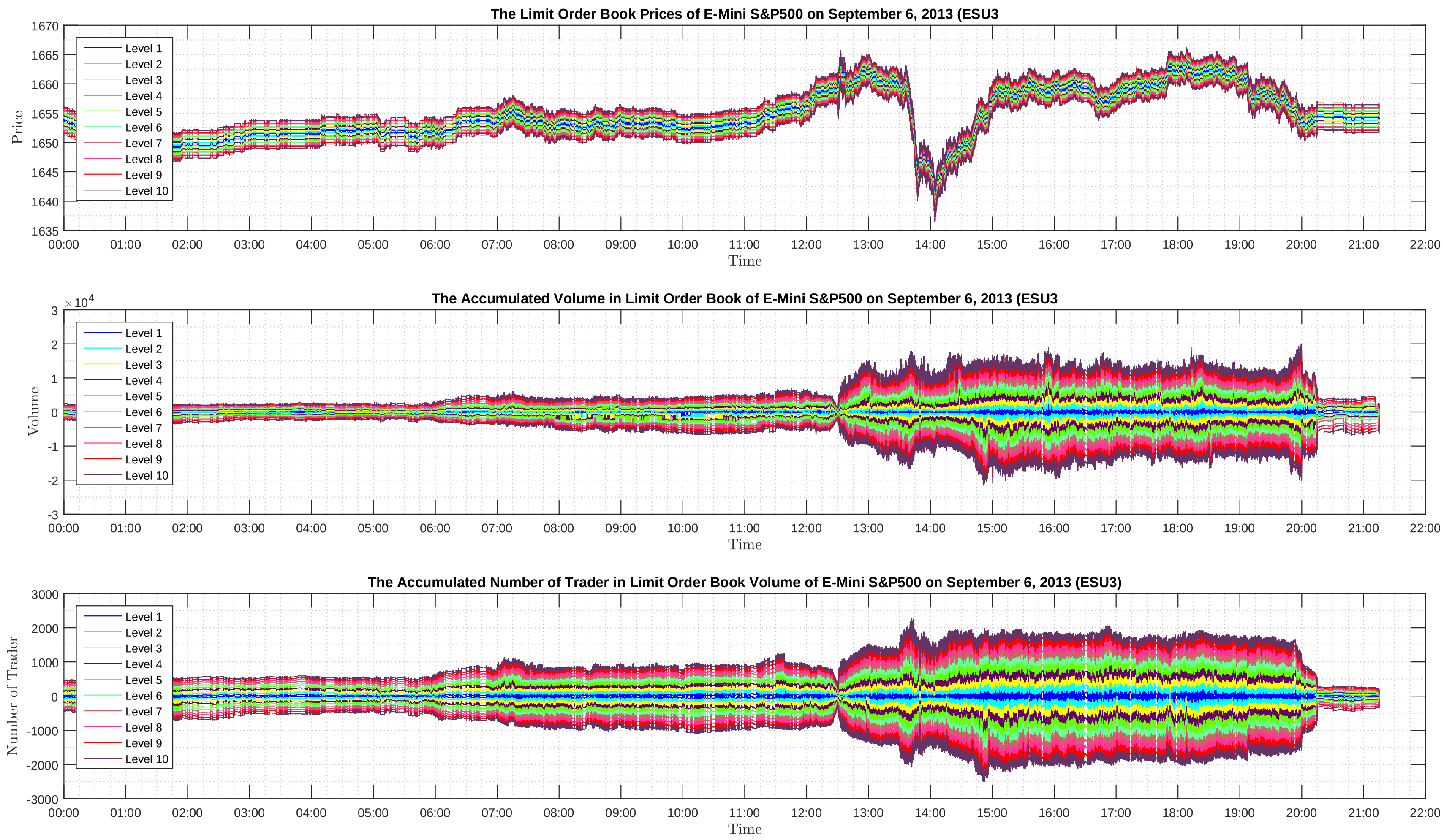



**Figure 1.1: The Limit Order Book Data of E-Mini S&P500 on September 6, 2013 (ESU3)**

Note: This figure presents the LOB data of E-Mini S&P500 on September 6, 2013 (ESU3). The figure presents 10 deep levels of LOB data for both bid and ask. The first subfigure presents the limit order price, second presents the limit order volume and the last subfigure present the number of traders in LOB.

trading for both Eurodollar and E-Mini S&P 500 is enormous in the terms of size with a different type of data. More importantly, this financial data changes very fast, and somehow this type of data does not fit the structure of the primitive models. This is why sometimes the daily or a longer period data may lead to the poor performance of the financial model. Therefore, to gain the value of the TRTH database and the precision of the model, this thesis also preferred to use tick-by-tick data.

## 1.2 The Topics Examined

Below provides a synopsis of each topic in the dissertation. The underlying methodology and results are presented in the individual chapters. A bibliography closes the dissertation.

### 1.2.1 Preliminary Analysis of Probability of Informed Trading (PIN)

One of the main topics of the market microstructure that has become increasingly popular is information asymmetry with forms of the standard theoretical models (see Glosten and Milgrom [1985]; Kyle [1985]; Easley and O'Hara [1987]; Easley et al. [1997b]). Easley et al. [1996] develops the probability of informed trading or PIN model to estimate the level of information asymmetry in the market. This model is based on the arrival rate of informed and uninformed traders. These traders arrival in the market following a Poisson process with probability of two types of news, good or bad. This chapter presents the preliminary results of the PIN model from Easley et al. [1996] in actually detecting informed behavior around a toxicity event. For this study, I use a data set of Eurodollar futures market as the pricing mechanism of this futures contract which is based on the LIBOR. I find that the PIN reacts strongly to certain types of events and the variation is statistically significant relative to the events. The result of the PIN is always higher for the Eurodollar futures market compared to the PIN in equity markets (circa 0.4 to 0.9 versus 0.1 to 0.5; see Easley et al. [1996]; Aslan et al. [2011]; Abad and Yagüe [2012]).However, other PIN studies of interest rate futures have found something similarly high (see Kim et al. [2014]).

This topic is addressed in chapter 4.

### 1.2.2 Motivating Evidence of LIBOR Manipulation and The PIN Analysis

One of the most striking evidences of the failure in financial regulation is represented by the London Interbank Offered Rate (LIBOR). Since May 2008, a huge scandal focusing on a possibility of criminal wrongdoing by a number of the most trusted international banks revealed manipulation of the benchmark interest rate known as the LIBOR. This scandal became as matter of fact on June 2012 when Barclays agreed to pay fines of $360 million and £144.5 for having rigged the LIBOR.

This chapter provides the crucial evidence of LIBOR manipulation including a communication evidence between interest rate derivative traders and LIBOR submitter s described in the CFTC and FSA documents. Also, I provide statistical evidence of LIBOR manipulation including LIBOR quotes and cross-sectional p-value correlation for banks' quote on the LIBOR submission. Furthermore, the chapter applies PIN with the LIBOR manipulation cases recorded in the regulatory reports. The objective of this empirical exercise is to examine the effectiveness of the PIN model from Easley et al. [1996] in actually detecting informed behavior around a LIBOR manipulation event. For this study, I use a data set of Eurodollar futures market as the pricing mechanism of the futures which is based on the LIBOR. to clearly understand why I use the Eurodollar futures as a data set to study the PIN around the LIBOR manipulation, this chapter provides the number of communication requested on LIBOR manipulation related to a number of currencies. From this evidence, it can be seen that the second most popular was the 3M-LIBOR which is the benched mark for the Eurodollar futures market. Additionally, I then compute the PIN around the maturity date as a normal event in the futures contract and investigate the variation of PIN around these events. Therefore, focused on a short period, the variation of PIN around LIBOR manipulation indicates that the PIN is a good early warning signal. However, the general long-run variation of the PIN was not statistically significant relative to both LIBOR manipulation and the maturity event.

This topic is addressed in chapter 5.

### 1.2.3 Order Flow Toxicity and Informed Trading Around Know Market Manipulation Events: Evidence from Interest Rate Futures

A fundamental of PIN has been extended to a volume adjusted PIN, denoted VPIN , developed by Easley et al. [2010]; Easley et al. [2012]. However, there is a key debate ongoing to whether the current high profile measures of the probability-of-informed trading (PIN) and 'market-toxicity' (measured by a volume adjusted PIN denoted VPIN) actually provide a substantive measurement of the phenomena in question. Andersen and Bondarenko [2013] criticizes the fact that the VPIN is highly sensitive to the volume of order in each volume bucket. In a series of follow-up papers Andersen and Bondarenko [2014c], Easley et al. [2014] and again Andersen and Bondarenko [2014b] have debated the relative metrics of the VPIN approach. However, my results of VPIN fall somewhere between Andersen and Bondarenko [2014a] and Easley et al. [2014].

This chapter follows a term-structure analysis to examine PIN and VPIN around the LIBOR manipulation and maturity event. I find very mixed results. Both PIN and VPIN vary systematically and in a statistically significant pattern in respect to the term structure of the futures contracts. PIN varies in a v-shaped pattern, with long (2000 to 3500 days) and short maturity (0 to 500 days) contracts having significantly higher PIN and VPIN measurements than intermediate contracts (which are actually the most heavily traded). However, when moving to documented cases of market manipulation in the reference rate, the results are ambiguous. There are definitive cases when the PIN (and to an extent certain flavors of the VPIN) shift systematically around a relevant documented LIBOR manipulation event. However, when building cross sectional averages across events, we find no significant evidence of systematic shifts in either the PIN or VPIN metric.

This topic is addressed in chapter 6.

### 1.2.4 Term-structure analysis of hidden order in the limit order book: evidence from the E-Mini S&P 500

Another fundamental of market microstructure study is a market mechanism and what is the new information effect on the market. The standard source of information for market participants is pieces of information, such as price and volume from trades and quotes data. The quotes data come from two type of order, market order and the limit order. The limit order book data normally appears on the trading screen which provides mostly 5 to 10 levels of price and volume from above (below) best offer (best bid). Currently, there is a new standard feature of the electronic limit order book in most electronic trading platforms, which allows traders to visible, entirely invisible or partially invisible their orders, called hidden order, invisible order and max show. Therefore, this chapter studies how the hidden order is incorporated into the financial market, and I revisit and adapt the Hasbrouck and Saar [2013] ordinary least squares (OLS) model to my hidden order study. Hasbrouck' OLS examines the effect of the $RunsInProcess$ and market quality, however, my OLS evaluates the relationship between hidden order and the market quality. Furthermore, I am able to extend the earlier works in market microstructure utilizing VAR models for estimation coefficients between quote revisions and trades (see Hasbrouck [1991], Dufour and Engle [2000], Zebedee [2001], Chung et al. [2005], Viljoen et al. [2014]). Whilst the market microstructure VAR model studies the relationship between quote revisions and trade direction, I extend this to quote revisions, signed-trade direction, signed-trade volume and signed-hidden liquidity. Comprehensive results are found, the coefficients estimated from the OLS between hidden order and market quality shows similar results for low and high-frequency trading periods. Furthermore, the impact of $HD$ is greater during low frequency than high-frequency trading periods, and it is always positive. In addition, the impulse response analysis is also presented in this. Significantly, I provide a comprehensive innovation of signed-hidden order detection algorithm for E-mini S&P500 using limit order book data with Volume-Weighted Average Price trade direction indicator. I also provide my hidden order detection algorithm in this chapter.

This topic is addressed in chapter 7.

# Chapter 2

# CME And Futures Data

## 2.1 Thomson Reuters Tick History Data (TRTH)

A key component of the empirical analysis in this thesis has been conducted based on tick-by-tick data from Thomson Reuters Tick History (TRTH). This unique data provide a high-performance tick container to store real-time and historical tick messages recorded in microsecond timestamps. The TRTH allows the researcher to collect trades, quote, and limit order book data (LOB) for most of the financial market. In doing so, in chapters four, five and six, the Eurodollar futures trading data is able to be easily matched between trade data and quote data at any given point in time. Furthermore, in chapter seven, I synchronize trade data and limit order book data from the E-Mini S&P 500. The limit order book data provided by TRTH allowed me to collect price and volume for ten levels above the last offer and ten levels below the last bid. Another benefit of TRTH is Reuters Instrument Codes (RICs) which is the unique and proprietary identification system. The RICs is applied in variety financial markets among every asset class traded globally. Therefore, this code reduced time for filtering and identifies the data set that is used in this thesis.

Additionally, in chapters six and seven, I have removed the weekend data for all trade, quote and LOB data to eliminate abnormalities activity. This activity creates an extreme value of, for example, a wide-spread, an extreme price impact, and high level of an order imbalance. Also, I have removed irregular value from our analysis, such as negative spread, NaN, and zero value. The details on cleaning procedures are provided in the following section.

## 2.2 The Chicago Mercantile Exchange (CME)

The dataset used in this research comes from the Chicago Mercantile Exchange (CME). The CME is owned by the CME Group which was merged between the New York Mercantile Exchange (NYMEX) and COMEX. The Merc (CME), CBOT, NYMEX and COMEX. Currently, the CME is the largest exchange for futures and options contracts open interest (number of contracts outstanding) of any futures exchange in the world. The CME offers the widest range of futures and options based on global benchmark products across major asset classes, such as, interest rates, equity index, foreign exchange, energy, agricultural commodities, metals, weather and real estate [1].

The transaction flow at the CME starts from the quote vendor systems receiving market data from market participants. The CME provides three different trading platforms, including the CME Globex, CME Direct, and CME ClearPort platforms. Then, price and volume are settled through matching, pricing & market integrity system. At the end of a trading session, all buyers and sellers

[1] CME Globex Reference Guide, CME Group, accessed April 14, 2016, <http://www.cmegroup.com/globex/files/GlobexRefGd.pdf>)

are required to clear their trade through the CME clearing house. As the worlds largest exchange for futures and options, the CME clearing house monitors and processes more than one billion trades each year which is worth more than $1,000 trillion. In general, the CME opens 24 hours from Sunday to Friday, 18.00 - 17.00 New York time/Eastern Time (17.00 - 16.00 Chicago Time/Central Time) with a one hour break each day beginning at 17.00 (16.00 Central Time).

In order to conduct any kind of trading, there has to be a venue to bring buyers and sellers together. Before technology and computers transformed the way of trading, the transaction took place at a much slower place and traders got in touch more physically on the trading floor. However, since the development of electronic trading, the buyers and sellers are now able to communicate via computerized trading machines (CME [2005]).

### 2.2.1 The Open outcry

The open outcry or pits trade is a face-to-face auction operating during regular trading hours, this method operates by floor traders who wear different colored jackets presented as their identities (traders, runner, brokers employees). These floor traders use complex hand signals to communicate with other market participants for trading information. Trading hours for interest rate futures and options on open outcry operating during normal business hours start from 7.20 until 14.00 from Monday to Friday side-by-side with the CME Globex electronic platform. However, the E-Mini S&P 500 are operating only on the CME Globex.

Figure 2.1 shows the CME trading floor, where traders stand in the pits making bids and offers on a specific commodity in particular locations where commodity trading may take place. The pits are surrounded by workstations or desks where traders receive the orders from clients around the world. Each pit is separated into small desks regarding the different contract delivery months. During business hours, lots of market participants wear gold jackets as it designates employees of various financial firms by a different type of badge. The CME members wear red, CME employees wearblue, out-trade clerks wear pale green. These people communicate trading information from the pits to the phone by means of hand signals, then the order is carried to and from members by runners.

### 2.2.2 The CME Globex electronic trading platform

The CME Globex platform was the first global electronic trading system for futures and options. This platform offers real-time market data for derivatives in all major asset classes such as, interest rates, equity indexes, FX, agriculture, energy, metals, weather and real estate. The Globex provides a 10 deep level of limit order book for futures and 3 deep levels for options. The CME operates the Globex electronic trading platform, which is approximately 80% of total volume of the exchange trading through this platform. Whilst the electronic trading platform is increasingly popular, the CME Group will shutter its New York trading floor after December 30, 2016, due to low trading volume. However, open-outcry options trading will continue on CME's floor in Chicago [1].

The trading session on the CME Globex platform operates 24 hours from 18.00 until 17.00 on the next day at New York time/ET or 17.00 to 16.00 at Chicago Time/CT with a one hour break each day beginning at 17.00 (16.00

[1] T., Polansek, CME to close New York options pits, shutting down the trading floor, accessed April 14, 2016, <http://www.reuters.com/article/usa-cme-options-idUSL2N17G24B>)

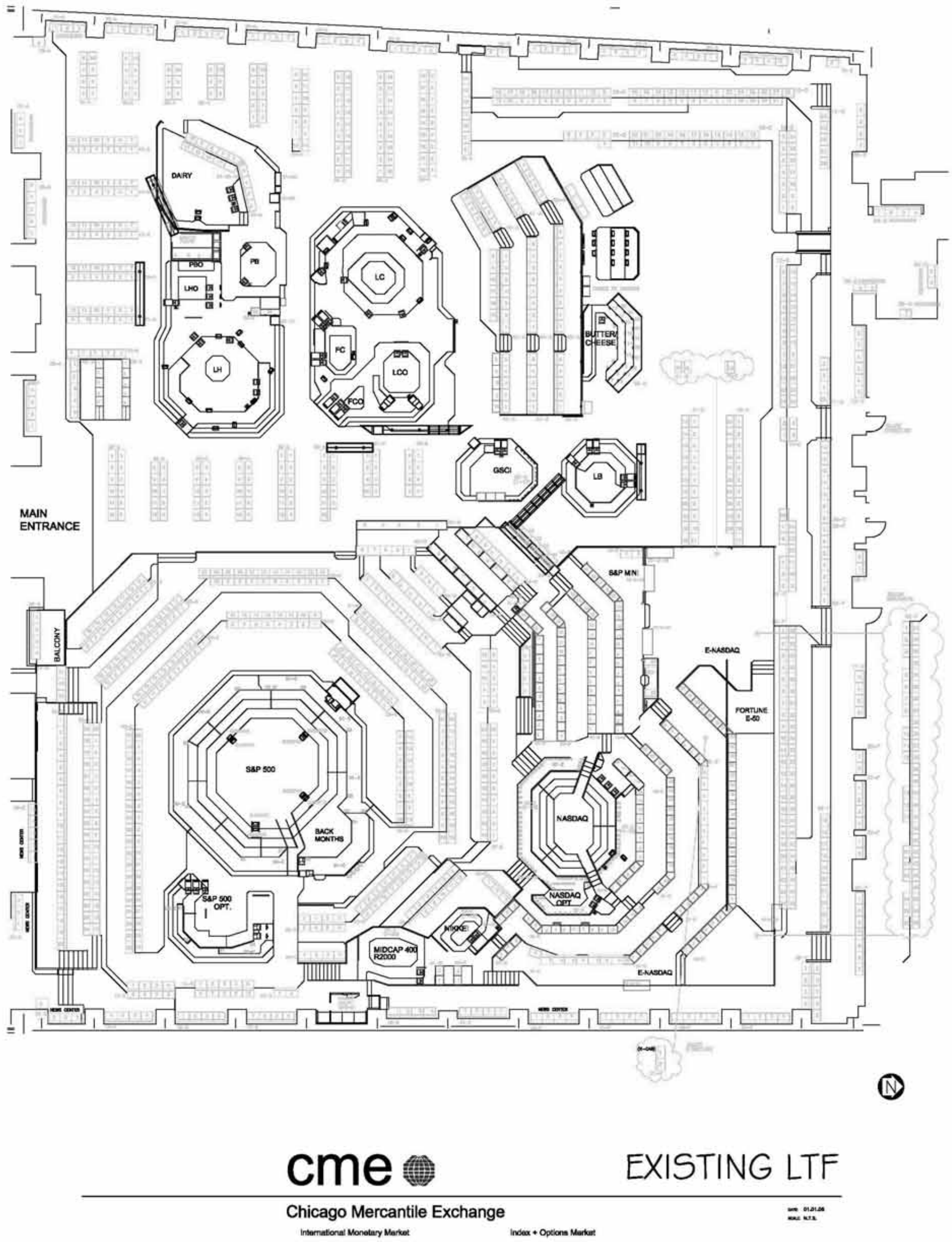



**Figure 2.1: CME Trading Venues for Open Outcry (Source: CME [2005])**

CT). The products that trade on the Globex platform are classified into three trading groups based on their hour of availability (CME [2005]). First, Side-by-Side contracts, this means the product that trades in both CME Globex platform and is also traded via open outcry on the trading floor for a portion of day. Second, Electronic-only is the product that trades only on the CME Globex platform. Third, After-Hours Electronic are the contracts that trade on the CME Globex platform only after the product stops trading on the trading floor.

The basic futures orders are buy order and sell order, however, there are different types of specific buy and sell order provided by the CME Globex platform. This platform supports a broad type of order functionality and order qualifiers, including- Market order, Limit Order, Market order with protection, Market to limit, Stop limit, Stop order, Stop order with protection, Minimum quantity, and Display quantity.

**Market order** is an order that is to be filled at the best available price immediately upon receipt by the broker

**Limit Order**, the Globex Limit Order Book (LOB) offers a 10-deep level of limit order for futures and 3-deep level for options. A limit order allows buyer and sellers to define a specific minimum and maximum price to pay or accept (the limit price), this order will remain on the book until the order is executed, canceled or expires.

**Market order with protection**, this order is similar to the market order plus a protection range to protect the risk of having missed-price at extreme prices. The protection range is normally at 50 % above or below the current best bid and offer, respectively.

**Stop Limit Order** is the order that rests in a limit order and is triggered and executed when the trigger price is traded on the market. This order is similar to a Limit order with specified limit price, however, the order is executed at all price levels between the trigger price and the limit price. For a buy order, the trigger price is lower than the limit price and higher than the last traded price. For a sell order, the trigger price is higher than the limit price and lower than the last traded price. If the order is not fully executed, the un-executed order remains on the market at the limit price.

**Stop order with protection** is a stop limit order with protection to avoid the risk of being executed at extreme prices. The protection point range is the trigger price plus (minus) 50% of the Non-Reviewable Trading Range for that product for the limit price for a buy (sell) stop with protection.

**A Minimum quantity** is the order that specifies a minimum quantity that can be immediately executed only if this certain order can be immediately matched. If at least the minimum quantity cannot be filled, then the entire order is canceled; however, if the minimum quantity or more is filled, then the remaining quantity is placed on the book.

**Display Quantity, Max Show**, this type of order allows market participants to visible only a specific portion of their order to the marketplace. When the displayed order has been executed, another portion equal to the displaced quantity is then displayed as a new order; for instance, if a trader wants to place a Buy order for 100 quantity with a 10 displayed. Therefore, no more than Buy

10 is exposed to the market at any time, and the remaining quantities are booked but are not displayed. Each time the quantity of 10 instruments is executed, the next 10 remaining orders enter the market as a new order at the bottom of the book.

## 2.3 The Eurodollar Futures Data

The data used in this research are introduced in this section, together with a preliminary processing of Eurodollar Futures data. As previously stated, Eurodollar futures contracts are traded on the CME's Globex platform and CME pit (open-outcry) trades. Data for both electronic and open-outcry are directly obtained from the CME tapes for the period January 1, 1996 to January 1, 2014. Pit trades are quotes from CME by the ED code and GE for the Globex code; the amalgamated tapes are classified under the ED moniker. The volume ratio between the Globex and open-outcry is between three and four orders of magnitude over the sample, so separating the pit trades from the electronic trades currently appears to be less interesting. The CME tapes data for the 44 Eurodollar futures contracts available from the Thomson Reuter database via the Tick History system. We have conducted an analysis on the 4 monthly contracts, however the data is very sparse and the volumes are very small (up to 5 orders of magnitude for busy versus busy days) compared to the 40 quarterly contracts.

The data of the Eurodollar Futures trading are collected using the RICs (Reuter's Information Codes), which are quoted by financial institutions in the US Dollar exchanges. These 40 Eurodollar futures contracts are all Eurodollar transactions trading on the CME's Globex platform; they contain a maturity code, starting and ending dates, bid/ask price, bid/ask volume, trade prices and trade volume. The data pulled from the RIC information system, is a Computerized Trade Reconstruction (CTR) reported to the Commodity Futures Trading Commission (CFTC) on instruments in the tag range set, and are given a specific code, i.e. EDU0, EDU1,. . . , EDU9. The first two letters "ED" indicate Eurodollar futures, and the third letter "H, M, U, Z" indicates the delivery month of March, June, September and December respectively. The last character in the Eurodollar code, 0 to 9, represents the delivery year for 10-year future contracts from 2000 to 2009 or 2010 to 2019. For instance, in the case of EDH0, EDM1, EDU2 and EDZ3, their delivery months and years are March 2000, June 2001, September 2002 and December 2003 respectively (see Table 2.1).

Table 2.1 indicates the settlement year, the total volume of bid/asks, total number of bid/asks, total volume of trade and total number of trades for each Eurodollar futures contract. The data include all update-by-update inside quotes and trades for a total sample size of about 1 billion rows of prices and volumes in millisecond time-stamps and other trade information. This is believed to be an unprecedented microstructure data set as shown in Table 2.1 and Figure 2.2.

Of the 40 Eurodollar future contracts, the largest bid volume is the EDZ3 contract at \$83.84 quadrillion. EDH4 has both the highest ask volume at \$92.80 quadrillion and the largest trade volume of \$120 trillion. Meanwhile, the biggest number of bids and asks belongs to EDU5 with approximately 25 million for both bids and asks, while EDH9 and EDZ9 has the largest number of trades with about 1.95 million for both series of contracts. The total value of bid/asks and trade from January 1, 1996 to January 1, 2014 was \$1.75 quintillion, \$1.76 quintillion and \$2.82 trillion, respectively. Furthermore, the total number of bid/asks and

**Table 2.1: Eurodollar Futures maturity codes, quotes volume and trades volume.**

| Maturity code | Delivery year | Bid volume (billion) | Ask volume (billion) | Trade volume (billion) | Bid updated (million) | Ask updates (million) | Trades (million) |
|---|---|---|---|---|---|---|---|
| EDH0 | 2000,2010, 2020 | 23.60 | 23.90 | 0.05 | 21.32 | 23.14 | 1.69 |
| EDH1 | 2001,2011, 2012 | 52.70 | 52.10 | 0.06 | 22.34 | 23.04 | 1.51 |
| EDH2 | 2002,2012, 2022 | 40.10 | 42.60 | 0.07 | 23.41 | 23.58 | 1.48 |
| EDH3 | 2003,2013, 2023 | 56.10 | 56.60 | 0.05 | 24.12 | 24.49 | 1.31 |
| EDH4 | 2004,2014 | 79.30 | 92.80 | 0.05 | 23.11 | 24.33 | 1.50 |
| EDH5 | 2005,2015 | 57.20 | 58.80 | 0.06 | 21.61 | 25.40 | 1.81 |
| EDH6 | 1996,2006,2016 | 31.10 | 30.40 | 0.07 | 18.33 | 25.97 | 1.82 |
| EDH7 | 1997,2007,2017 | 42.00 | 40.10 | 0.06 | 13.62 | 23.03 | 1.38 |
| EDH8 | 1998,2008,2018 | 34.40 | 33.90 | 0.11 | 12.23 | 21.25 | 1.48 |
| EDH9 | 1999,2009,2019 | 20.10 | 20.10 | 0.09 | 18.20 | 22.97 | 1.95 |
| EDM0 | 2000,2010, 2020 | 27.30 | 27.70 | 0.06 | 21.51 | 23.02 | 1.67 |
| EDM1 | 2001,2011, 2012 | 56.60 | 57.60 | 0.06 | 22.78 | 23.42 | 1.49 |
| EDM2 | 2002,2012, 2022 | 39.10 | 39.50 | 0.06 | 23.67 | 23.93 | 1.48 |
| EDM3 | 2003,2013, 2023 | 61.90 | 66.10 | 0.04 | 23.35 | 23.87 | 1.30 |
| EDM4 | 2004,2014 | 83.30 | 80.20 | 0.05 | 23.00 | 24.26 | 1.55 |
| EDM5 | 2005,2015 | 45.20 | 48.30 | 0.08 | 20.87 | 25.56 | 1.92 |
| EDM6 | 1996.2006,2016 | 28.20 | 28.10 | 0.08 | 17.09 | 25.52 | 1.76 |
| EDM7 | 1997,2007,2017 | 40.00 | 40.00 | 0.07 | 12.40 | 22.11 | 1.31 |
| EDM8 | 1998,2008,2018 | 31.30 | 30.90 | 0.10 | 13.69 | 21.55 | 1.55 |
| EDM9 | 1999,2009,2019 | 17.00 | 17.00 | 0.08 | 19.56 | 23.29 | 1.93 |
| EDU0 | 2000,2010, 2020 | 30.80 | 30.10 | 0.07 | 22.79 | 22.86 | 1.63 |
| EDU1 | 2001,2011, 2021 | 52.80 | 67.20 | 0.02 | 23.47 | 23.34 | 1.49 |
| EDU2 | 2002,2012, 2022 | 38.30 | 38.30 | 0.06 | 24.55 | 24.33 | 1.44 |
| EDU3 | 2003,2013, 2023 | 77.30 | 74.90 | 0.05 | 23.94 | 23.84 | 1.39 |
| EDU4 | 2004,2014 | 75.70 | 73.80 | 0.06 | 24.60 | 24.40 | 1.64 |
| EDU5 | 2005,2015 | 39.50 | 40.70 | 0.08 | 25.73 | 25.71 | 1.93 |
| EDU6 | 1996,2006,2016 | 31.20 | 31.00 | 0.08 | 25.00 | 24.98 | 1.60 |
| EDU7 | 1997,2007,2017 | 40.60 | 39.60 | 0.09 | 21.35 | 21.30 | 1.32 |
| EDU8 | 1998,2008,2018 | 27.60 | 29.20 | 0.01 | 22.05 | 21.88 | 1.61 |
| EDU9 | 1999,2009,2019 | 15.90 | 16.30 | 0.07 | 19.56 | 23.29 | 1.93 |
| EDZ0 | 2000,2010, 2020 | 42.80 | 43.70 | 0.07 | 21.32 | 23.14 | 1.69 |
| EDZ1 | 2001,2011, 2021 | 46.80 | 46.00 | 0.07 | 22.34 | 23.04 | 1.51 |
| EDZ2 | 2002,2012, 2021 | 44.80 | 45.50 | 0.06 | 23.41 | 23.58 | 1.48 |
| EDZ3 | 2003,2013, 2023 | 83.80 | 82.10 | 0.06 | 24.12 | 24.49 | 1.31 |
| EDZ4 | 2004,2014 | 67.00 | 70.00 | 0.07 | 23.11 | 24.33 | 1.50 |
| EDZ5 | 2005,2015 | 36.20 | 36.10 | 0.09 | 21.61 | 25.40 | 1.81 |
| EDZ6 | 1996,2006,2016 | 36.80 | 36.50 | 0.08 | 18.33 | 25.97 | 1.82 |
| EDZ7 | 1997,2007,2017 | 38.10 | 37.90 | 0.12 | 13.62 | 23.03 | 1.38 |
| EDZ8 | 1998,2008,2018 | 24.00 | 23.70 | 0.11 | 12.23 | 21.25 | 1.48 |
| EDZ9 | 1999,2009,2019 | 18.10 | 18.30 | 0.06 | 18.20 | 22.97 | 1.95 |
| Total | | 1,754.60 | 1,767.60 | 2.82 | 831.36 | 946.91 | 63.83 |

Note: Eurodollar Futures maturity codes, quotes volume and trades volume from January 1, 1996 to January 1, 2014. The volume of bid/asks is reported in terms of 100 billion contracts. The trade volume is reported in millions of contracts.

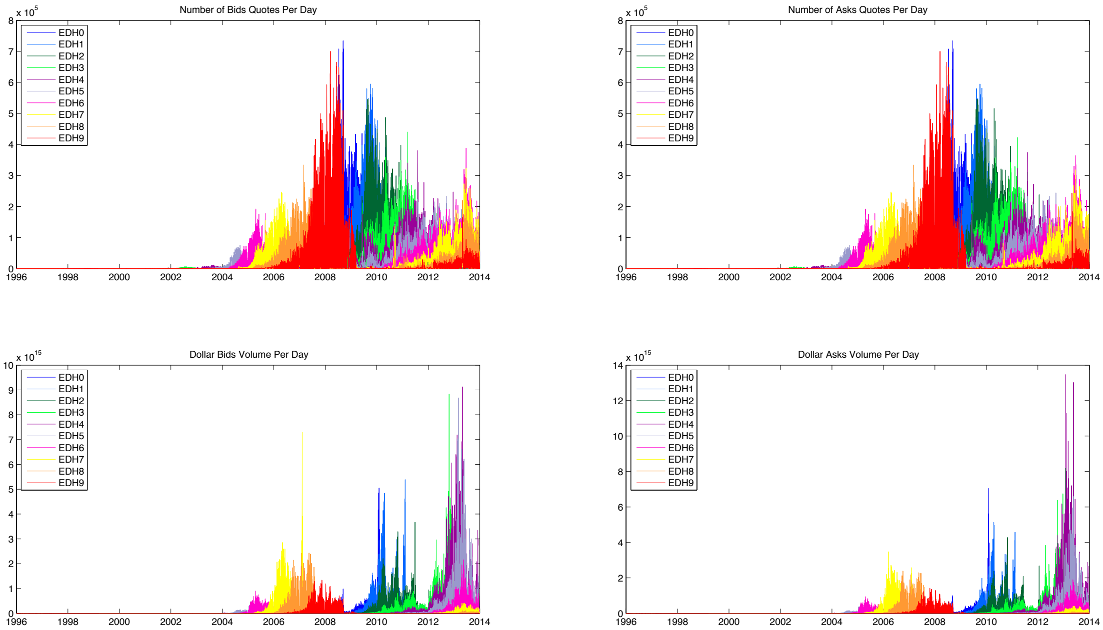



**Figure 2.2: Number of bid/ask quotes and volume of quotes per day for EDH contracts EDH1, EDH2,. . . , EDH9 between 1996 and 2013.**

trades was approximately 831.36 million, 946.91 million and 63.83 million respectively.

Some information about the logistics of dealing with such a vast amount of data can now be provided. The raw data (uncompressed CSV files) comprised 489GB of best-bid, best offer and trade data, directly retrieved from the CME tapes. The raw tapes were streamed into a new format ‘hdf5’ which provided a high-integrity medium for this amount of data. Price and volume data were stored in a number format in order to reduce the amount of storage space as much as possible (e.g. prices were stored using IEEE single-precision numbers; volumes were stored at 16 bit integers). With compression, the raw 489GB was reduced to 89GB of compressed hdf5 data stored in separate files by maturity date, and then stored on a solid state drive. In the interests of speed, the data were replicated across three separate drives.

Given the scale of this significant data, the PIN calculations were computed for each day by multiple instances of Matlab, on a large multiprocessor workstation with 0.5TB of RAM and virtual memory. The total computation time in each empirical analysis was around four to six weeks.

### 2.3.1 ED-Data Cleaning Procedures

In the market microstructure topic, data collection is a ubiquitous function especially in high frequency data - not only for record keeping, but also to provide a mixture of data analysis protocol which is the main part of empirical research (Hellerstein [2008]). Despite the fact that data collection is important for this research, data quality remain inescapable. Even though the ED Trades and Quotes data is provided by a trustworthy data base, the TRTH data base, a precise data cleaning technique is also important for correcting data processing task. The cleaning of this data is carried out following steps S1-S5. The steps are applied with all trades and quotes data.

Step 1: Delete entries a bid, ask or trade price lower than 80 and NaN value. This procedure tries to reduce an unexpected price or mis-price from trades and quotes data as the ED price covered around 95 to 100.

Step 2: Delete entries a bid, ask or trade volume equal to zero and NaN value. This technique is to eliminate nonessential values in this time series data.

Step 3: Delete duplicate data from both Trades and Quotes data. This organization and sorting technique helps to reduce costs time and improve efficiency in analysis process.

Step 4: Retain entries missing values and match with equivalent time vector for each variable.

Step 5: Finally, trades and standing quotes is matched at the time of the trade time stamp.

In practice the Eurodollar futures data is incredibly clean. Most outliers are found in the 1996 to 2005 period and are easily recognizable as the price is generally quoted either near zero (circa 0.90) or near 1000, (circa 900) due to miss recording of a pit trade.

## 2.4 The E-Mini S&P 500

The largest trading volume in the CME is the E-mini S&P500. The E-mini trades exclusively on the CME Globex trading platform in a fully electronic limit order market, 24 hours a day from Sunday to Friday 17.00 - 16.00 (Chicago Time/CT) with 15-minute technical maintenance break each day. The E-mini futures contract is in denominations 10 times smaller than the original floor-traded S&P 500 index futures contract. The contracts are cash settled against the value of the underlying S&P 500 equity index at expiration dates in March, June, September, and December of each year. Trades can occur up to 8.30 on the $3^{rd}$ Friday of the contract month. The notional value of one E-mini contract is $50 times the S&P stock index, the minimum price increment or tick size is 0.25 index point or $12.50. Since its introduction, the E-mini has become a more favorable instrument to hedge exposure to baskets of U.S. stocks or to speculate on the direction of the entire stock market. The E-mini contract attracts the highest dollar volume amongst U.S. equity index products. Kirilenko et al. [2015] states that the E-mini features offer both pre-trade and post-trade transparency. The pre-trade transparency is provided by transmitting to the public in real time, price and quantity of buy and sell orders resting in the central limit order book up or down ten tick levels from the last transaction price. The post-trade transparency is provided by transmitting to the public the price and quantity of executed transactions.Furthermore, Hasbrouck [2003] shows that the transparency of the E-mini has become a price discovery market for the U.S. Equity Indices.

This work uses publicly available data from CME's Globex. The limit order book data and the transaction records are collected by using the RICs (Reuter's Information Codes) from the Thomson Reuters Tick History Service (TRTH). I collected two sets of records from TRTH data, first transaction records, and second the quotes updated records for buy and sell orders resting in the central limit order book up or down ten tick levels from the last transaction price. The sample was constructed to capture all activity in the limit order book across E-mini S&P 500 futures trading. I began by identifying all common delivery times for each contract. At this point, I had to study the product code. The code for this contract start with "ES" which mean E-mini S&P 500. The third letter "H, M, U, Z" indicates the delivery month: March, June, September and December respectively whilst the number 0 to 9 is the last digit of the year of maturity, up to ten years. For example, the contract that has an expiry date on the third Friday of March 2014 is the ESH4 and on the third Friday of June 2015 is ESM5. In total, this thesis includes 32 E-mini contracts from July 2, 2008 to June 19, 2015 which included ESU8, ESZ9, ESH0,..., ESM5. I exclude half-days adjacent to holiday closures.

### 2.4.1 ES-Data Cleaning Procedures

At this point, I acquired trade and limit order book data for ten levels from below(above) best bid(offer) from TRTH, which included price, volume, and number of traders. Both trades and limit order book data records are time stamped to microseconds, although in practice, the accuracy is to one millisecond. Even though the E-mini S&P 500 trade and limit order book data is provided by a trustworthy database, a precise data cleaning technique is also important to correct the data. The cleaning procedure of this data is carried out following steps S1-S5. This step is applied to all trades and limit order book data.

Step 1: Delete entries quotes with bids(offers) that are greater(smaller) than offers(bids) or mis-priced from trades and quotes data.

Step 2: Delete entries a bid, ask or trade volume equal to zero and NaN value. This technique is to eliminate nonessential values in this time series data.

Step 3: Delete duplicate data from both trades and quotes data. This organization and sorting technique helps to reduce costs and time, and improve efficiency in the analysis process.

Step 4: Retain entries missing values and match with equivalent time vector for each variable.

Step 5: Finally, trades and standing quotes is matched at the time of the trade time stamp.

The E-mini S&P 500 is a futures contract on the S&P index, whose normal time to maturity affects trading activity. Trading in futures, traders have to trade against the expiration period, with trading activity increasing for a shorter time to maturity compared to a longer period. Therefore, this work focuses on a term-structure of the E-mini S&P500 instead of the whole trading period. As a result, I separate data into 18 different data sets from one to eighteen weeks to the maturity. The main reason for this is that the trading frequency for longer than eighteen weeks to maturity is between two and three orders of magnitudes lower than the trading frequency for shorter periods. For instance, the average trade updated on the eighteenth week is about 5,000 trades, however, the average trade updated between one to thirteen weeks before maturity varies between 1.1 to 2.8 million trades. This trades number is more than 200 times larger than the average trade updated on the eighteenth week. Furthermore, focusing from eighteen weeks to maturity is because the E-mini has a quarterly expiration and mostly trades occurred within a twelve-week period. Obviously, index futures are different from other futures contracts, such as Eurodollar futures in which normal trading activity for index futures is located roughly within one year to maturity, peaking at a week to maturity. This short term trading activity cannot compare with the ten year trading periods of Eurodollar futures, as normally traders are used to managing, hedging, or speculating their trading position against a long or short term interest rate with anything from an overnight rate to a ten-year interest rate.

Table 2.2 provides an average quote volume in the limit order book, trade volume, quotes updated and trade updated from July 2, 2008 to June 19, 2015 presented as a term-structure from eighteen to one week to maturity. A summary of trades volume and trades updated in this table confirms that trading activity on this index futures is mostly located between fourteen and two weeks to maturity, as I mentioned in the previous section. Trading volume significantly increases from 0.067 million on the sixteenth week to 7.510 million in the fourteenth, which is more than 100 times larger. Also, this similar pattern appears on the bid, ask

**Table 2.2: Averge number of quotes volume within limit order book, trade volumes, quotes updated and trade updated for E-Mini S&P 500 presented as a term-structure from eighteen to one weeks to maturity.**

| Week to Maturity | Bid Volume (billion) | | | | | Ask Volume(billion) | | | | | Trade Volume (million) | Bid Updated (million) | Ask Updated (million) | Trade Updated (million) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L_1 | L_2 | L_3 | L_4 | L_5 | L_1 | L_2 | L_3 | L_4 | L_5 | | | | |
| 1 | 1.434 | 3.112 | 3.539 | 3.693 | 3.594 | 1.391 | 3.043 | 3.490 | 3.635 | 3.546 | 0.801 | 3.793 | 3.793 | 0.158 |
| 2 | 4.944 | 10.198 | 12.202 | 12.681 | 12.817 | 4.949 | 9.960 | 11.797 | 12.301 | 1.248 | 4.598 | 9.582 | 9.582 | 1.118 |
| 3 | 5.803 | 11.567 | 13.740 | 14.746 | 15.120 | 5.802 | 11.660 | 13.890 | 14.920 | 15.359 | 5.716 | 12.486 | 12.486 | 1.444 |
| 4 | 4.066 | 8.364 | 10.048 | 10.566 | 10.862 | 4.144 | 8.450 | 10.117 | 10.796 | 11.210 | 4.105 | 8.789 | 8.789 | 1.034 |
| 5 | 5.586 | 12.002 | 14.696 | 15.594 | 16.103 | 5.435 | 11.950 | 15.018 | 16.206 | 16.958 | 5.700 | 14.458 | 14.458 | 1.543 |
| 6 | 6.571 | 14.292 | 18.156 | 19.370 | 20.260 | 6.635 | 14.533 | 18.667 | 20.288 | 21.201 | 7.199 | 17.368 | 17.368 | 1.913 |
| 7 | 6.775 | 14.633 | 18.214 | 20.817 | 21.735 | 6.744 | 14.429 | 18.060 | 20.748 | 21.841 | 7.777 | 20.179 | 20.179 | 2.196 |
| 8 | 5.907 | 12.218 | 15.100 | 17.818 | 18.866 | 5.835 | 12.308 | 15.547 | 18.428 | 19.584 | 7.460 | 19.261 | 19.261 | 2.120 |
| 9 | 4.458 | 9.469 | 11.950 | 13.634 | 14.177 | 4.382 | 9.438 | 11.976 | 13.784 | 14.487 | 5.714 | 14.004 | 14.004 | 1.643 |
| 10 | 4.601 | 9.625 | 12.412 | 15.196 | 16.343 | 4.574 | 9.503 | 12.282 | 15.109 | 16.369 | 7.379 | 17.755 | 17.755 | 2.127 |
| 11 | 7.846 | 16.901 | 21.548 | 24.495 | 25.460 | 8.028 | 16.994 | 21.638 | 24.592 | 25.738 | 10.338 | 25.644 | 25.644 | 2.840 |
| 12 | 6.796 | 13.799 | 16.613 | 19.417 | 20.154 | 7.009 | 14.078 | 16.883 | 19.545 | 20.210 | 7.689 | 19.209 | 19.209 | 2.043 |
| 13 | 6.400 | 13.131 | 16.091 | 19.205 | 19.928 | 6.272 | 12.874 | 15.851 | 19.054 | 19.791 | 8.909 | 19.608 | 19.608 | 2.464 |
| 14 | 4.962 | 9.618 | 11.617 | 13.633 | 14.292 | 5.097 | 9.768 | 11.731 | 13.838 | 14.679 | 7.510 | 17.637 | 17.637 | 2.160 |
| 15 | 1.697 | 4.020 | 4.709 | 5.074 | 5.451 | 1.688 | 4.100 | 4.827 | 5.197 | 5.591 | 2.389 | 7.642 | 7.642 | 0.649 |
| 16 | 0.280 | 0.543 | 0.586 | 0.591 | 0.611 | 0.234 | 0.521 | 0.572 | 0.574 | 0.591 | 0.067 | 2.093 | 2.093 | 0.019 |
| 17 | 0.175 | 0.362 | 0.392 | 0.396 | 0.410 | 0.196 | 0.369 | 0.391 | 0.393 | 0.406 | 0.028 | 1.510 | 1.510 | 0.008 |
| 18 | 0.063 | 0.159 | 0.180 | 0.186 | 0.191 | 0.100 | 0.172 | 0.182 | 0.185 | 0.194 | 0.016 | 1.125 | 1.125 | 0.005 |

Note: This table presents an average number of quotes volume in limit order book, trade volume, quotes updated and trade updated from July 2, 2008 to June 19, 2015 presented as a term-structure from eighteen to one week to maturity.

and trade updated. Trading activity reaches a peak at eleven weeks to maturity for both trade and quotes in the limit order book. The average weekly trading volume for this week is 10.338 million, trade updated is 2.840 million and quotes updated is 25.644 million. After ten weeks to maturity, trade and quotes updated slightly decrease and finally plunge in the last week of trading period or one week to maturity. For the last week of the trading period, the average weekly trading volume is 0.801 million volumes, which is 13 times smaller compared with the average weekly trading volume at eleven weeks to maturity, which is the peak trading period. Additionally, for one week to maturity, trade and quotes updated are 0.158 million and 3.793 million respectively, which is 18 times smaller for trade updated and 7 times smaller for quotes updated compared with the highest trading frequency period.

Table 2.3 reports the descriptive statistic for E-Mini S&P500 futures reported by the number of weeks to maturity from July 2, 2008 to June 19, 2015 presented as a term-structure from eighteen to one week to maturity. Noticeably, the spread in this paper is calculated by using the VWAP technique from limit order book data. The average weekly spread for eighteen to one week to maturity varies between 1.72 and 2.19 basis-points and the average standard deviation of this spread is 1.2 basis-points; the maximum standard deviation of spread appears on the seventeenth week at 6.9 basis-points, then drops to just about 1.2 after sixteen weeks to maturity. The maximum average weekly return appears on the last week to maturity at about 0.084 basis-points and the minimum is -0.079 for sixteen week to maturity. The maximum standard deviation of the return appears on the seventeenth week at 0.45 basis-points, then drops to just about 0.06 basis-points from fifteen weeks to maturity. The pattern of standard deviation of return is similar to the pattern of standard deviation of spread.

**Table 2.3: The descriptive statistics for E-Mini S&P500 futures reported by the number of weeks to maturity from July 2, 2008 to June 19, 2015 presented as a term-structure from eighteen to one week to maturity.**

| Number of Week to Maturity | Mean spread (BPS) | Median spread (BPS) | Min spread (BPS) | Max spread (BPS) | Std spread (BPS) | Skew spread | Kurt. Spread | Mean return (BSP) | Min return (BSP) | Max return (BSP) | Std return (BSP) | Skew return | Krut. Return |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.826 | 1.823 | 1.208 | 3.299 | 1.285 | 1.093 | 31.500 | 0.084 | -0.069 | 0.072 | 0.082 | 0.026 | 14.195 |
| 2 | 1.869 | 1.863 | 1.306 | 2.716 | 1.199 | 0.433 | 7.231 | 0.017 | -0.060 | 0.055 | 0.059 | -0.003 | 14.772 |
| 3 | 1.878 | 1.867 | 1.219 | 3.133 | 1.220 | 0.795 | 11.703 | -0.016 | -0.078 | 0.082 | 0.058 | -0.010 | 18.181 |
| 4 | 1.744 | 1.736 | 1.213 | 2.716 | 1.170 | 0.655 | 10.710 | -0.036 | -0.074 | 0.075 | 0.063 | 0.002 | 17.468 |
| 5 | 1.725 | 1.721 | 1.178 | 2.613 | 1.158 | 0.274 | 6.038 | 0.009 | -0.057 | 0.057 | 0.062 | 0.003 | 13.998 |
| 6 | 1.726 | 1.721 | 1.204 | 2.591 | 1.120 | 0.279 | 5.763 | -0.079 | -0.054 | 0.056 | 0.061 | 0.028 | 15.107 |
| 7 | 1.730 | 1.723 | 1.088 | 2.859 | 1.188 | 0.758 | 12.829 | 0.005 | -0.072 | 0.073 | 0.062 | 0.038 | 16.730 |
| 8 | 1.744 | 1.737 | 1.083 | 2.847 | 1.257 | 0.584 | 11.974 | -0.035 | -0.074 | 0.073 | 0.062 | -0.015 | 15.898 |
| 9 | 1.745 | 1.739 | 1.140 | 2.674 | 1.171 | 0.365 | 6.870 | 0.039 | -0.070 | 0.058 | 0.063 | -0.015 | 15.200 |
| 10 | 1.743 | 1.739 | 1.195 | 2.572 | 1.133 | 0.189 | 6.075 | -0.023 | -0.063 | 0.065 | 0.062 | 0.005 | 14.597 |
| 11 | 1.723 | 1.718 | 1.134 | 2.771 | 1.179 | 0.504 | 9.836 | -0.004 | -0.070 | 0.066 | 0.066 | -0.024 | 17.593 |
| 12 | 1.734 | 1.729 | 1.151 | 2.816 | 1.242 | 0.648 | 10.408 | -0.075 | -0.075 | 0.071 | 0.069 | -0.011 | 14.304 |
| 13 | 1.770 | 1.766 | 1.120 | 2.784 | 1.212 | 0.403 | 9.766 | 0.033 | -0.064 | 0.066 | 0.067 | -0.003 | 17.184 |
| 14 | 1.778 | 1.776 | 1.115 | 2.755 | 1.126 | 0.269 | 9.256 | 0.019 | -0.066 | 0.069 | 0.068 | 0.011 | 14.866 |
| 15 | 1.796 | 1.790 | 1.271 | 3.147 | 1.123 | 1.294 | 32.621 | 0.046 | -0.107 | 0.105 | 0.125 | 0.003 | 12.176 |
| 16 | 1.938 | 1.864 | 1.409 | 6.308 | 3.360 | 7.678 | 162.499 | 0.069 | -0.198 | 0.206 | 0.284 | 0.059 | 12.913 |
| 17 | 2.197 | 2.008 | 1.447 | 7.062 | 6.960 | 5.562 | 89.492 | 0.029 | -0.227 | 0.229 | 0.456 | -0.026 | 9.390 |
| 18 | 2.018 | 1.954 | 1.468 | 8.401 | 4.014 | 6.050 | 145.804 | 0.056 | -0.223 | 0.235 | 0.437 | -0.021 | 8.622 |

Note: This table presents descriptive statistic for E-Mini S&P500 futures reported by number of week to maturity from July 2, 2008 to June 19, 2015 presented as a term-structure from eighteen to one week to maturity.

## 2.5 Regulation

Futures markets in the United State are currently regulated federally by the U.S. Commodity Futures Trading Commission (CFTC). The CFTC was established in 1974 after being under Federal regulation since 1920 (www.cftc.gov). This committee is responsible for issuing the license for futures exchange and approving the contracts (Hull [2012]). More importantly, the CFTC has the mission to look after the public interest and to promote market integrity from a variety of market abuses. This means it is responsible for ensuring that price and information communicated to the public is served to all market participants. This body has to deal with the public interest to investigate and take an action against inappropriate trading in the futures market.

One of the best examples of CFTC action happened in the aftermath of the 2008 financial crisis involved in a part of interest rate derivatives that leaded to the reform of in this market known as the Dodd-Frank trading rules. The brief history of the investigation into market manipulation for the benchmark of interest rate derivatives is provided in chapter 5.

Another U.S. regulator is the Securities and Exchange Commission (SEC; www.sec.gov), this body was created after congress passed the Securities Act in 1933. The responsibility of SEC is similar to the CFTC, the body has a mission to protect investors and maintain fair trading for all investors. The SEC needs to ensure that all investors should have access to certain basic information about their investment as presented in an efficient market theory. Normally, the SEC conducts securities exchange, securities brokers and dealers, investment advisors, public companies, and mutual funds to maintain fair dealing and protect market participants against fraud.

The United Kingdom market was regulated by the Financial Services Authority (FSA; fsa.gov.uk) which is an independent non-government organization. The FSA was established in December 2001 under the Financial Services and Aarket Act 2000 (FSMA). This body is a limited company financed by the financial service industry, however, the FSA is accountable to the treasury minister and the U.K. parliament. Similar to the CFTC and the SEC, FSA regulates most U.K. financial services markets, exchange and firms to meet its standards. However, after 2012 the FSA was separated into two regulatory authorities; The Financial Conduct Authority (FCA) and the Prudential Regulation Authority (see figure 2.3). The mission for these two bodies is to protect investors and market participants from fraud, market abuse, or other financial crime. Similar to the CFTC, the highest fines was for the repeated misleading of LIBOR settlements known as the LIBOR scandal. Since 2012 the imposed fines from FCA worth £744 million, the biggest fines handed to the Deutsche bank accounted for £277 million, the UBS accounted for £160, the Rabobank accounted for £105, the RBS accounted for £87.5, the Barclays accounted for £59.5 and the Lloyds accounted for £105.

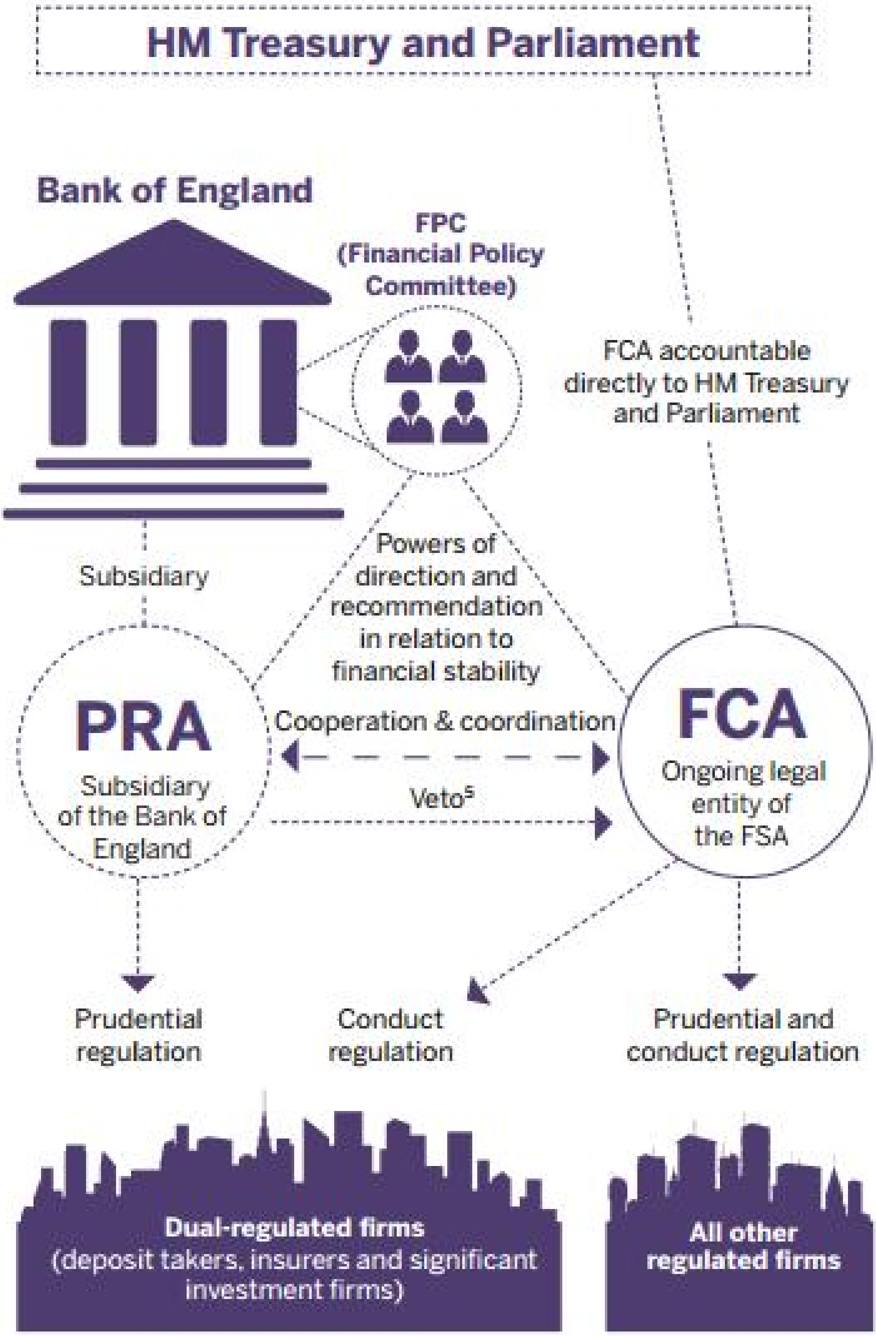



**Figure 2.3: The new system for regulating financial services in the UK (Source: FCA [2012])**
Note: This flow chart presents the new system of financial regulating service in the UK. This new system involves a number of bodies such as, the FCA, the PRA, the Bank of England, the Financial Policy Committe (FPC), and Her Majesty's Treasury.

# Chapter 3

# A Review of Microstructure Models

This chapter summarizes the major information-based trading models with an expansion to understanding how these models incorporate price behavior.

### 3.0.1 The Rational Expectation Model

This section provides a constructive explanation of the rational expectations (RE) model that nests the standard Grossman and Stiglitz [1980], Hellwig [1980], Admati [1985] models. The foundation from these RE models has been expanded into different economic phenomena: information acquisition in financial markets (Verrecchia [1982]), the operation of information markets (Admati and Pfleiderer [1986, 1987, 1990]), multi-asset information learning (Admati [1985]), insider trading (Leland [1992]), asset-payoff and supply (Breon-Drish [2015]), price volatility (Gao et al. [2013]).

The main RE model follows the original one-period model from the Grossman and Stiglitz [1980] framework. The RE model depends on strong assumptions that all random variables are jointly normally distributed and all agents have absolute risk aversion (CARA) preference with a risk aversion parameter (Breon-Drish [2015]). This model considers the market of risk free and risky assets over a single period of time $T$, $T = 0$ to 1. At the beginning of a trading period the price of the risky asset is $S$ and the payoff is unknown until the end of the trading period or time = 1 represented by random variable $\theta$. The information of $\theta$ is defined as the value of the assets into two components: $\theta = \kappa + \nu$, where $\kappa$ is a certain cost of $\theta$ on condition of information from informed traders. The second component $\nu$ is the residual that is uncorrelated with $\kappa$. Also, the $\kappa$ is uniformly distributed and the $\nu$ is independently distributed with zero mean and finite variance $\sigma_\nu^2$ (Grossman [1988]). Following Gao et al. [2013], the a price of risk free assets is normalized to 1, and the payoff is certain at $1+\varphi$, so, with our loss $\varphi = 0$. Also, the RE model assumes that informed traders can observe $s$, however, uninformed traders generally observe only the price.

For this model, the proportion of informed traders is a random number $\lambda$ with a distribution $F(\cdot)$ with probability $\mathbb{P} > 0$. As the model assumes that trader is CARA, the utility derived by a trader $i$ from the return $\pi_i = (\theta - S)x_i$ of buying $x_i$ units of the risky asset at price $S$ is $\mathscr{U}_i(\pi) = \mathbb{E}[(\theta - S)x_i|F_i] - 0.5\rho\mathbb{D}[(\theta - S)x_i|F_i]$, where $\mathbb{E}[\cdot]$ is the expectation operator, $\mathbb{D}[\cdot]$ is the variance operator, $\rho > 0$ is the risk aversion coefficient, and $F_i$ is an information set for trader $i$.

At the equilibrium, the demand schedules for an informed (I) trader, a position $X_i$ in risky asset yields a utility:

$$\begin{aligned}\mathscr{U}_I &= \mathbb{E}[(\theta - S)X_I|\kappa, S] - 0.5\rho\mathbb{D}[(\theta - S)X_I|\kappa, S] \\ &= X_I\mathbb{E}[(\theta - S)|\kappa, S] - 0.5\rho X_I^2\mathbb{D}[(\theta - S)|\kappa, S].\end{aligned}$$

The demand schedules of utility maximization is:

$$X_I = \frac{\mathbb{E}[\theta|\kappa, S]}{\rho\mathbb{D}[\theta|\kappa, S]}.$$

where $\mathbb{E}[\theta|\kappa, S] = \mathbb{E}[\theta|\kappa] = \kappa$, and $\mathbb{D}[\theta|\kappa, S] = \mathbb{D}[\theta|\kappa] = \sigma_\nu^2$. Therefore, informed trader's demand schedule defined as:

$$X_I = \frac{1}{\rho\sigma_\nu^2}\kappa - \frac{1}{\rho\sigma_\nu^2}S. \tag{3.0.1}$$

As, $\kappa$ and $S$ are linear function, then $X_I = X_{I(\kappa,S)}$.

Gao et al. [2013] show that the demand of an uninformed (U) trader can be calculated using the same method, however, the expectation of the payoff at time 1 depends on only the price. Therefore, the demand schedule for the uninformed trader can be defined as:

$$X_U = \frac{[\theta|S] - S}{\rho\mathbb{D}[\theta|S]}. \tag{3.0.2}$$

the rational expectation equilibrium is a set of trades, conditional on the information that traders have, $X_I(\kappa, S)$ for informed traders and $X_U(S)$ for uniformed traders, and a measurable price functional $S(\kappa, \lambda)$. For market clearing, for all pairs $(\kappa, \lambda)$, the price $S(\kappa, \lambda)$ equates to the supply of the risky asset to total demand, then

$$\lambda X_I(\kappa, S(\kappa, \lambda)) + (1 - \lambda)X_U(S(\kappa, \lambda)) = 1. \tag{3.0.3}$$

At equilibrium, the REE assumes all uninformed traders believe that the number of informed traders is completely irrelevant to the price. Therefore, the price function $S(\cdot)$ depends on $\kappa$ and fully revels the signal $\kappa$, then all traders will behave as if they are informed. Hence, the equilibrium price can be found equating demand to supply, which we have $X_U = X_I = 1$. Substituting $X_I = 1$

in Eq. 3.0.1, then an alternative price function to write is

$$S = \kappa - \rho\sigma_{\nu}^{2}. \tag{3.0.4}$$

The rational expectation equilibrium is defined by the set of functions of $X_I, X_U$ and $S$, however, this is not the unique equilibrium for the economy. Therefore, the price does not reveal the informed's signal to uninformed traders

### 3.0.2 The Kyle [1985] Model

In another market microstructure context that consider when trader has superior information to derived the security price or how the prices adjust to the full information value is introduced in Kyle [1985]. This model proposes that the new equilibrium price partially reflects the new full information with three types of market participants: market makers, noise traders, and informed traders. Kyle proposed a single period model which at the value of an asset is a random variable : $S \sim \mathcal{N}(p_0, \sigma_0^2)$, at the end of period asset value of $\tilde{S} \sim \mathcal{N}(p_0, \Sigma_0)$.

There are also noise or uninformed traders who trade for exogenous reasons and submit a market order for $\tilde{u}$ quantity, where $\tilde{u} \sim \mathcal{N}(0, \sigma_u^2)$. Also, Kyle assumes that at the beginning of trading period informed trader can get information of the price of the asset from the value of $S$, and he knows the value of the asset at the end of the period is equal to $\tilde{S}$ but does not know the quantity $\tilde{u}$;however, an informed trader chooses to submit a market order to maximize their profit at $x$ quantity. At this point, the market maker observes the net order flow $y = u + x$ and sets a price $p$, however, the market maker cannot distinguish which part of the order comes from noise or informed traders. At the equilibrium, the market maker earns a zero expected profit, so he takes a position $-(u+x)$ to clear the market and earn $-(\tilde{S}-p)(u+x)$. The market maker sets $p$ following the function of $u+x$, thus $P(x+u) = \mathbb{E}[\tilde{S}|u+x]$.

The informed trader chooses to submit quantity $x$ that depends on $\tilde{S}$ to maximize his profit $\tilde{\pi}$ at the end of a period. Thus, the market maker sets the equilibrium price as:

$$\max_x \mathbb{E}[\tilde{\pi}|\tilde{S}] = \max_x \mathbb{E}[(\tilde{S} - P(x+\tilde{u}))x|\tilde{S}] \tag{3.0.5}$$

This model assumes at the equilibrium that the market maker price is a linear function given by the posterior expectation; $P(y) = \mu + \lambda y$. So, the maximize profit of the informed trader is $\mathbb{E}[\pi] = \max_x \mathbb{E}[(\tilde{S}-P(x+\tilde{u}))x|\tilde{S} = S] = \max_x \mathbb{E}[(S-\mu-\lambda(x-\tilde{u}))x|S] = \max_x (S-\mu-\lambda x)x$, when the last step follows from the fact that $\mathbb{E}[\tilde{u}] = 0$.

Then, maximizing the expected profit, the solution to informed trader for optimal trade is

$$x = \arg\max_{x\in\mathbb{R}} \mathbb{E}[\pi] = \frac{S-u}{2\lambda} = \alpha + \beta v, \tag{3.0.6}$$

where $\alpha = -\frac{u}{2\lambda}$ and $\beta = \frac{1}{2\lambda}$, if $\lambda > 0$.

To solve the linear market maker pricing and trade order parameters, we assume that the market maker sets the market price in order to earn zero profits at $p = \mathbb{E}[\tilde{S}|u+x]$. The expected order flow is $\mathbb{E}(y) = \tilde{u} + x = \tilde{u} + \alpha + \beta\tilde{S}$, where $S$ and $y$ are jointly normally distributed, that is, $\mathbb{E}[\tilde{S}|y]$. Therefore, the market maker should follow the maximum likelihood estimator to optimal pricing rule that equals to $\mu + \lambda y$ where $\mu$ and $y$ minimize, $\min_{\mu,\lambda} \mathbb{E}[(\tilde{S} - P(y))^2] = \min_{\mu,\lambda} \mathbb{E}[(\tilde{S} - \mu - \lambda y)^2] = \min_{\mu,\lambda} \mathbb{E}[(\tilde{S} - \mu - \lambda(\tilde{u} + \alpha + \beta\tilde{S}))^2] = \min_{\mu,\lambda} \mathbb{E}[(\tilde{S}(1-\alpha\beta) - \alpha\tilde{u} - \mu - \lambda\alpha)^2]$.

Following assumptions $\mathbb{E}[S] = p_0, \mathbb{E}[(S-p_0)^2] = \Sigma_0, \mathbb{E}[u] = 0, \mathbb{E}[u^2] = \sigma_u^2$ and $\mathbb{E}[uS] = 0$, then $\min_{\mu,\lambda} \mathbb{E}(1-\lambda\beta)^2(\Sigma_0+p_0^2)+(\mu+\lambda\sigma)^2+\lambda^2\sigma_u^2-2(\mu+\lambda\alpha)(1-\lambda\beta)p_0$. The first condition respect to $\mu$ and $\lambda$ are $\mu = -\lambda\alpha + p_0(1-\lambda\beta)$, thus

$$\lambda = \frac{\beta\Sigma_0}{\beta^2\Sigma_0 + \sigma_u^2}. \tag{3.0.7}$$

Use $\alpha = -\frac{\mu}{2\lambda}$ and $\beta = \frac{1}{2\lambda}$, we have $\mu = p_0$ and $\lambda = \frac{\sqrt{\Sigma_0}}{2\sigma_u}$.

At equilibrium, the market maker trade price is

$$p = p_0 + \frac{\sqrt{\Sigma_0}}{2\sigma_u}(\tilde{u} + \tilde{x}), \tag{3.0.8}$$

where the informed order is

$$x = \frac{(\tilde{S} - p_0)\sigma_u}{\sqrt{\Sigma_0}}. \tag{3.0.9}$$

From equation 3.0.9, we can see that the informed order is greater or more active in the magnitude provided by the volatility of the order from uninformed trader $\sigma_u$. Substituted 3.0.9 into 3.0.8, we get $p = p_0 + \frac{\sqrt{\Sigma_0}\tilde{u}}{2\sigma_u} + \frac{(\tilde{S}+p_0)}{2}$. Thus, only one-half of private information $\frac{1}{2}\tilde{v}$ is reflected to $p$, therefore the equilibrium price is not fully reveled by informed trader's information. The expected profit of the informed trader, unconditional on knowing the value of $\tilde{S}$ at the beginning of trading period is

$$\mathbb{E}[\tilde{\pi}] = \frac{\sigma_u(\tilde{S} - p_0)^2}{2\sqrt{\Sigma_0}} \tag{3.0.10}$$

Since the market maker sets the trade price in condition to earn zero profit, the expected gain for informed traders is the expected loss from noise traders, not the market maker. As the expected profit from informed order is a linear in noise volatility, this can be assumed that informed trader hide their order with the orders from uninformed trader to hide the position. Consider the illiquidity parameter $\lambda = \frac{\sqrt{\Sigma_0}}{2\sigma_u}$ which presents the value that the market maker rises the price when the net order flow $y = u + x$ increases by one unit. Therefore, the $\lambda y = \sqrt{\Sigma_0}\frac{y}{2\sigma_u}$ is liquidity risk scaled by volatility of security, and $\frac{y}{\sigma_u}$ is similar to the percentage of volume. Hence, the amount of order that raises the price by 1 dollar equals $\frac{1}{\lambda}$ which is measured by the market depth or market liquidity. Intuitively, the greater number of noise traders, the greater profits of informed traders that gain from the loss of uniformed traders. However, with a greater number of uninformed traders, an individual loss is less.

### 3.0.3 The Glosten and Milgrom [1985] Model

This section describes the basic sequential trading with the superior information model proposed by Glosten and Milgrom [1985]. This model assumes that some traders have information about price of a security($S$) while others do not. Suppose that $S$ will rise at time T, so $S_T = \{\underline{S}, \overline{S}\}, \underline{S} \leq \bar{S}$. Uninformed traders (U) know that $S_T = \underline{S}$ or $S_T = \overline{S}$, however, they do not know whether $S_T = \underline{S}$ or $S_T = \overline{S}$ will occur. Following the assumption of this model, informed traders (I) arrive in the market at rate $\mu$, these traders assumed to be informed (I) because they know certainty the terminal value $S_T$.

Next, De Prado [2011] states that the probability when the security price at time $T$ equals $\underline{S}$ is $\mathbb{P}(S_T = \underline{S}) = \delta$, and the security price at time $T$ equals $\overline{S}$ is $\mathbb{P}(S_T = \overline{S}) = 1 - \delta$. If a trade occurs at $t \in [0, T)$, the probability's distribution when $(\{\underline{S}, \overline{S}\}, \{I, U\}, \{Buy, Sell\})$: for the buy's unconditional probability is $\mathbb{P}(Buy) = \frac{1+\mu(1-2\delta)}{2}$, and for the sell's unconditional probability is $\mathbb{P}(Sell) = \frac{(1-\mu)(1-2\delta)}{2}$. Therefore, $\mathbb{P}(Buy) + \mathbb{P}(Sell) = 1$. Subsequently, the nature of event occurs with $\delta_t = 0.5 \Rightarrow \mathbb{P}(Buy) = \mathbb{P}(Sell)$, $\mathbb{P}(S_T = \underline{S})$ and $\mathbb{P}(S_T = \overline{S}) = 1 - \delta$ at $t = 0$ (See table 3.1).

**Table 3.1: Probability scenarios under event certainty (De Prado [2011])**

| $S_T = \underline{S} \leq S_t$ | Buy, $\underline{S}$ | Sell, $\underline{S}$ |
|---|---|---|
| I | 0 | $\delta(1+\mu)$ |
| U | $\frac{1}{2}(1-\mu)\delta$ | $\frac{1}{2}(1-\mu)\delta$ |

| $S_T \leq \overline{S} = S_t$ | Buy, $\overline{S}$ | Sell, $\overline{S}$ |
|---|---|---|
| I | $(1-\delta)(1+\mu)$ | 0 |
| U | $\frac{1}{2}(1-\mu)(1-\delta)$ | $\frac{1}{2}(1-\mu)(1-\delta)$ |

For the time interval $t \in [0, T)$, a market maker recognizes a price level at which he intends to enter whether long position (Bid, B) or short position (Ask, A). whilst, the market maker doesn't know that other market participants at period $t$ are informed or not, he can update his belief about the value of $S_T$ as trades are revealed, so $\delta_{t+1}(Buy) = \mathbb{P}(S_T = \underline{S}|Buy) = \mathbb{P}(S_T = \underline{S}, Buy)/\mathbb{P}(Buy) = \delta_t(1-\mu)/1+\mu(1-2\delta)$. If $S_T = \underline{S} \leq S_t$ informed traders lower their expectation, since $E[S_T|Buy] = \underline{S}(1-\mu)\delta_t + \overline{S}(1+\mu)(1-\delta_t)/1-\mu(1-2\delta_t)$ and he increases his expectation on $S_T$, since $E[S_T|Sell] = \underline{S}(1+\mu)\delta_t + \overline{S}(1-\mu)(1-\delta_t)/1+\mu(1-2\delta_t)$.

Then, B is $\delta_{t+1}(Sell) = \mathbb{P}(S_T = \underline{S}|Sell) = \mathbb{P}(S_T = \underline{S}, Sell)/\mathbb{P}(Sell) = \delta_t(1+\mu)/1-\mu(1-2\delta_t)$. Therefore the spread is $A - B = 4(1-\delta_t)\delta_t\mu(\overline{S} - \underline{S})/1 - (1-2\delta_t)^2\mu^2$. At this point, the market participant's expectation for a profit is zero. Therefore, in the case of $(S_T) = \underline{S} \leq S_t$ informed trading $(\mu)$

occurs when $A = E[S_T|Buy]$ with Informed (I) traders profit is gained by direct wealth transfer from the loss incurred by Uninformed (U) trades ($\varepsilon$). We can therefore write $(A - E[S_T|I, Buy])\mathbb{P}(I|Buy) = -(A - E[S_T|U, Buy])\mathbb{P}(U|Buy)$, since by construction $\mathbb{P}(I|Sell) + \mathbb{P}(U|Sell) = 1$. This argument can be made for B when $(S_T) = S_t \leq S_t$.

### 3.0.4 The Easley et al. [1996] Model - *Probability of Informed Trading (PIN)*

The PIN is a measure of asymmetric information on a trading event between informed and uninformed trading, which was developed on the theoretical framework of Easley and O'Hara [1992], although the original PIN approach was introduced by Easley et al. [1996]. They produced a mixed discrete-time and continuous-time sequential model of the trading mechanism, in which trades emerge when three classifications of economic agents – namely market makers, informed traders and uninformed traders collaborate. Abad and Yagüe [2012], explain that on a trading day, a market maker is competitive or strategic trader, while informed and uninformed traders are risk neutral with probability of profit or loss at 0.5. Also, informed and uninformed traders arrive in the market with their unique arrival rate with a view to a trading game between the liquidity provider and trader. Despite this, the PIN is not a direct measurement observation, but rather a parameter of a microstructure model which is estimated by the numerical maximization of the likelihood function. This is explained by Easley et al. [1996], as considering the likelihood of the arrival of three different types of information on a trading day: no news, good news, and bad news. This set of options for market agents can be summarized in a standard tree diagram representing the trading process. This diagram gives the structure of the trading process, where $\alpha$ is the probability of an information event, $\delta$ is the probability of a low signal, $\mu$ is the rate of informed trade arrivals, and $\varepsilon$ is the rate of uninformed buy and sell rate arrivals.

The original PIN is a measure by Easley et al. [1996] of asymmetric information, fitted via maximum likelihood estimation directly from measurements of the order flow. The tree diagram in Figure 3.1 presents the dynamics of the game and frames the original PIN type model in terms of the information structure of the ED futures market.

The tree diagram in figure 3.1 describes the trading process of Easley et al. [1996]. Individual trading periods of a single risky asset on trading days are indexed $i = 1, 2, \ldots, I$ time is continuous within a trading day, and is indexed by $t \in [0, T]$. The competitive market makers are ready to buy or sell the asset at bid and ask posted price during the trading period. Information events are independently distributed at the beginning of each trading day, and occur with probability $\alpha$. However, informed traders will only trade on information event days; they will buy an asset if they receive a signal of good news with probability $1 - \delta$ and sell if they receive a signal of bad news with probability $\delta$. As $\varepsilon_b$ and $\varepsilon_s$ are the selling and buying rate of uninformed traders, which are supposed to have the same intensity (0.5), therefore the probability of uninformed traders is $\varepsilon_s = \varepsilon_b = \varepsilon$. In this way, the set of parameters are reduced to only four: $\alpha, \delta, \mu$ and $\varepsilon$.

On a trading day, trade occurs from both uninformed and informed traders who trade on condition of a received signal. Both types of uninformed buyers and sellers arrive in the market at rate $\varepsilon$. Competitive informed traders who are risk neutral will arrive on the trading days on which the information events have

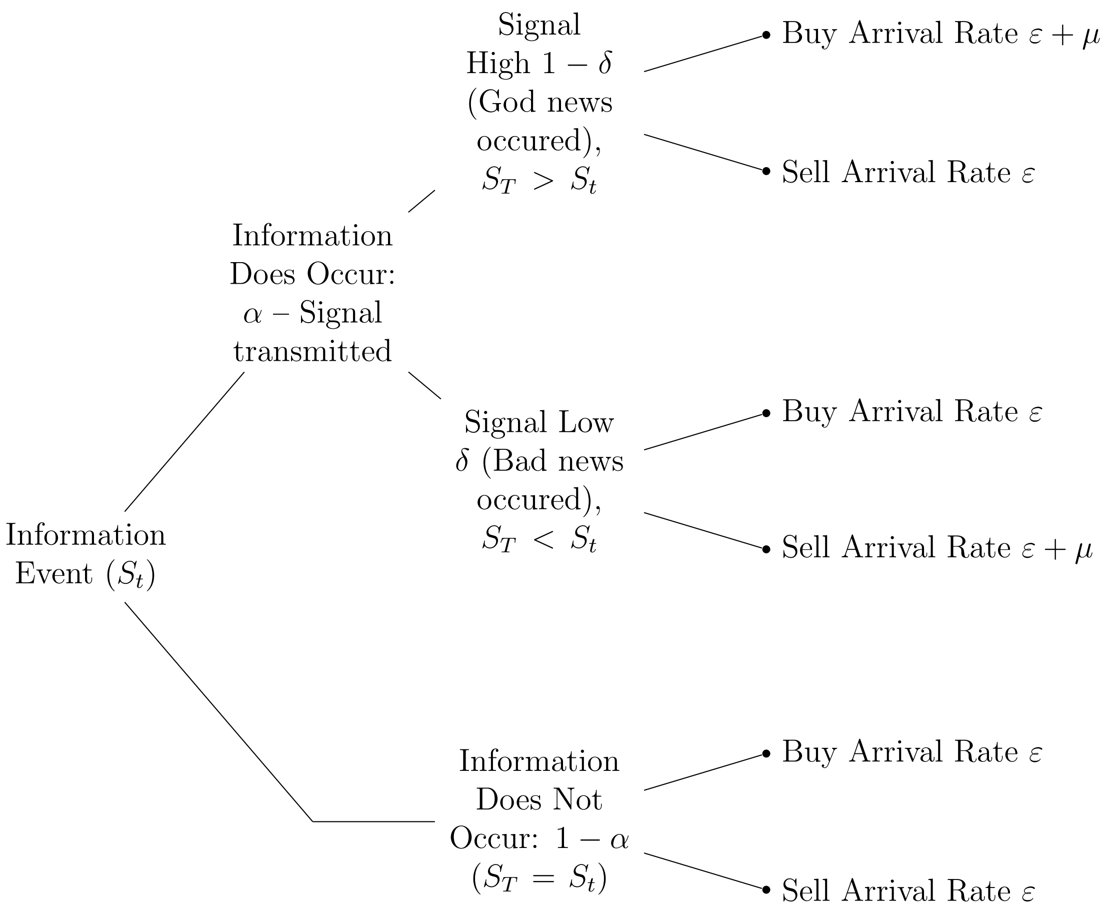



**Figure 3.1: Tree diagram of the sequential trading progress [Easley et al., 1996].**

occurred. If they receive a signal of good news, they will arrive to buy orders; conversely, they will submit sell orders if they receive a bad signal or news. The arrival rate for this process is $\mu$. These arrival processes are assumed to be independent. Then, on good event days, the arrival rate for buy orders is $\varepsilon + \mu$ and for sell orders is $\varepsilon$. On the other hand, if there is any bad news or bad signal on any trading day, the arrival rate of sell orders is $\varepsilon + \mu$ and buy orders arrive at a rate of $\mu$. Finally, if there is no news or no signal on that day, only uninformed traders arrive for both buy and sell orders at arrival rate $\varepsilon$.

In each period the news arrival contains one of the three types of information. The market maker knows that there is some probability attached to each branch, and he knows the order arrival process for this branch. However, the market maker does not know which of the three branches has been naturally selected. Since he cannot directly observe which type of branch has been selected, Easley et al. [1996] assumes that the market maker is a Bayesian who uses the arrival of trade and the rate of trading to update his beliefs about the nature of the information event on the trading day. Let $\mathbb{P}(t) = (\mathbb{P}_n(t), \mathbb{P}_b(t), \mathbb{P}_g(t))$ be a liquidity provider's belief about the occurrence of information event "no news" (n), "bad news" (b), and "good news" (g) at time $t$. Then, his prior beliefs at time 0 is $\mathbb{P}(t_0) = (1 - \alpha, \alpha\delta, \alpha(1 - \delta))$.

To determine quotes at time $t$, the market maker updates his prior belief on the condition of an arrival order of the relevant type. For instance, the bid at time $t$, $B(t)$, is the expected value of the asset conditional both on the history of the process prior to the arrival of order at time $t$ and on the fact that someone wants to sell the asset. Let $(s_t)$ denote the event that a sell order arrives at time $t$. Let $\mathbb{P}_n(t|s_t)$ be the market maker's updated belief vector conditional on the history prior to time $t$ and on the event that a sell order arrives at time $t$.

Following Bayes' rule, the market maker's posterior probability on no news at time $t$, if an order to sell arrives at $t$, is

$$\mathbb{P}_n(t|s_t) = \frac{\mathbb{P}_n(t) \cdot \varepsilon}{\varepsilon + \mathbb{P}_b(t)\mu}\,, \tag{3.0.11}$$

the posterior probability on bad news is

$$\mathbb{P}_b(t|s_t) = \frac{\mathbb{P}_b(t) \cdot (\varepsilon + \mu)}{\varepsilon + \mathbb{P}_b(t)\mu}\,, \tag{3.0.12}$$

and the posterior probability on good news is

$$\mathbb{P}_g(t|s_t) = \frac{\mathbb{P}_g(t) \cdot \varepsilon}{\varepsilon + \mathbb{P}_b(t)\mu}\,. \tag{3.0.13}$$

Comparable to the case of buy orders ($b_t$) the posterior probability on good news at time $t$ is

$$\mathbb{P}_g(t|b_t) = \frac{\mathbb{P}_g(t) \cdot (\varepsilon + \mu)}{\varepsilon + \mathbb{P}_g(t)\mu}\,, \tag{3.0.14}$$

the posterior probability on bad news is

$$\mathbb{P}_b(t|b_t) = \frac{\mathbb{P}_b(t) \cdot \varepsilon}{\varepsilon + \mathbb{P}_g(t)\mu}\,, \tag{3.0.15}$$

and the posterior probability on no news is

$$\mathbb{P}_n(t|b_t) = \frac{\mathbb{P}_n(t) \cdot \varepsilon}{\varepsilon + \mathbb{P}_g(t)\mu}\,. \tag{3.0.16}$$

At the end of the trading on any period, the full information value of the asset is realized. If it is good news on the trading period ($i$) the informed trader knows that the value of the asset at the end of this period is worth $\bar{S}_i$, similarly it is $\underline{S}_i$ if it is bad news on period $i$ and the asset on day $i$ is worth $S_i^* = \delta \underline{S}_i + (1+\delta)\bar{S}_i$, (so, $\bar{S}_i > S_i^* > \underline{S}_i$) if there is no news at all.

To complete the bid or ask at time $t$, the liquidity provider updates his position on the condition of arrival order according to information type. At time $t$ the expected value of the asset, conditional on the history of trade prior to time $t$, is

$$\mathbb{E}[S_i|t] = \mathbb{P}_n(t)S^* + \mathbb{P}_b(t)\underline{S}_i + \mathbb{P}_g(t)\bar{S}_i \,, \tag{3.0.17}$$

where $S_i^* = \delta \underline{S}_i + (1+\delta)\bar{S}_i$ is the prior expected value of the asset.

The bid is the expected value of the asset conditional on someone wanting to sell the asset to a liquidity provider. Thus, it is

$$B(t) = \mathbb{E}[S_i|t] - \frac{\mu \mathbb{P}_b(t)}{\varepsilon + \mu \mathbb{P}_b(t)} (\mathbb{E}[S_i|t] - \underline{S}_i) \,. \tag{3.0.18}$$

Similarly, the ask is the expected value of the asset conditional on someone wanting to buy the asset from a liquidity provider. Thus, it is

$$A(t) = \mathbb{E}[S_i|t] + \frac{\mu \mathbb{P}_g(t)}{\varepsilon + \mu \mathbb{P}_g(t)} (\bar{S}_i - \mathbb{E}[S_i|t]) \,. \tag{3.0.19}$$

If there is no news on a trading day, thus there are no informed traders ($\mu = 0$), then both bid and ask are equal to the expected value of the asset. Generally, both informed and uninformed traders will be in the market, so $A(t) > \mathbb{E}[S_i|t] >$ B(t).

The bid–ask spread at time $t$ is denoted by $\Sigma(t) = A(t) - B(t)$. This spread is

$$\Sigma(t) = \frac{\mu \mathbb{P}_g(t)}{\varepsilon + \mu \mathbb{P}_g(t)} (\bar{S}_i - \mathbb{E}[S_i|t]) + \frac{\mu \mathbb{P}_b(t)}{\varepsilon + \mu \mathbb{P}_b(t)} (\mathbb{E}[S_i|t] - \underline{S}_i) \,. \tag{3.0.20}$$

The first term in the bid–ask spread equation is a probability of a buy order based on information and the second is the term for sells. The spread for the initial quotes in the period, $\Sigma$, has a particularly simple form in the natural case in which good and bad events are equally likely. That is, if $\delta = 1 - \delta$, then

$$\Sigma = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} (\bar{S}_i - \underline{S}_i) \,. \tag{3.0.21}$$

The key component of this model is the probability that an order is from an informed trader, which is called the PIN:

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \,, \tag{3.0.22}$$

where $\alpha\mu + 2\varepsilon$ is the arrival rate for all orders and $\alpha\mu$ is the arrival rate for information-based orders. Therefore, the PIN is a measure of the fraction of orders that arise from informed traders relative to the overall order flow, and the spread equation shows that it is the key determinant of spreads. These equations clarify that liquidity providers need to correctly estimate the PIN in order to identify the optimal levels at which to enter the market. An unanticipated increase in PIN will result in losses to those liquidity providers who do not adjust their prices.

It is difficult to estimate the parameter vector $\Theta = (\alpha, \delta, \varepsilon, \mu)$ because it cannot be directly observed in either the occurrence of information events or the associated arrival of uninformed and informed traders. In fact, in terms of measuring the daily arrival rate of sell $(s_t)$ and buy $(b_t)$ it is possible to infer these values using maximum likelihood and assuming that the trading process follows a Poisson process [Karyampas and Paiardini, 2011].

The Easley et al. [1996] PIN approach considers the likelihood of order arrivals on a day of known type. In this model, the likelihood of observing sell $S$ and buy $B$ orders on each type of information occurs on a no event day with probability $1-\alpha$, a bad event day with probability $\alpha\delta$ and a good event day with probability $\alpha(1-\delta)$. Thus, the likelihood is

$$\begin{aligned} L((B,S)|\Theta) =&(1-\alpha)e^{-\varepsilon_b}\frac{(\varepsilon_b)^B}{B!}e^{-\varepsilon_s}\frac{(\varepsilon_s)^S}{S!} \\ &+\alpha\delta e^{-\varepsilon_b}\frac{(\varepsilon_b)^B}{B!}e^{-(\varepsilon_s+\mu)}\frac{(\varepsilon_s+\mu)^S}{S!} \\ &+\alpha(1-\delta)e^{-(\varepsilon_b+\mu)}\frac{(\varepsilon_b+\mu)^B}{B!}e^{-\varepsilon_s}\frac{(\varepsilon_s)^S}{S!}\,. \end{aligned} \tag{3.0.23}$$

Since only one type of information occurs on trading days, the maximum likelihood estimator of the information event parameters $\alpha$ and $\delta$ will be either 0 or 1. However, these parameters can be estimated from the daily number of buy and sells, assuming that the days are independent. The likelihood of observing the data $M = (B_i, S_i)_{i=1}^{I}$ across $I$ periods is just the product of the daily likelihood function

$$L(M|\theta) = \prod_{i=1}^{I} L(\Theta|B_i, S_i)\,. \tag{3.0.24}$$

The PIN estimates are computed by maximizing the parameter vector $\theta$ from any data set $M$, which is normally taken to be from the daily trades and quotes.

### 3.0.5 The Easley et al. [2012] Model - *Volume-Synchronized Probability of Informed Trading (VPIN)*

The fundamental mechanism behind VPIN was developed from PIN that can be found in Easley et al. [2008]. Easley et al. [2010] develop this model by introducing time variations of the arrival rate of informed and uninformed traders in GARCH style by estimating the expected Order Imbalance or OI at time $\tau$, $\mathbb{E}\left[\left|V_\tau^{*B} - V_\tau^{*S}\right|\right]$ as an approximate of PIN($\alpha\mu$), then compared with total number of trades $\mathbb{E}\left[\left|V_\tau^{*B} + V_\tau^{*S}\right|\right]$ which equals $\alpha\mu + 2\varepsilon$ in the PIN model.

Easley et al. [2010] develop this model to measure the volume time in an attempt to match the speed of arrival of new information to the marketplace which called the Volume-Synchronized Probability of Informed Trading (VPIN), provides a simple metric for measuring order toxicity in a high-frequency environment while the PIN is difficult to work with the HFTs. The different system of time between PIN and VPIN is the PIN work on clock time while the VPIN works on volume time. The PIN works on a daily order imbalance between trading volume from informed traders and total order under independent process and price efficiency while the VPIN works on constant amount of volume or volume bucket. The volume bucket divides a trading section into a comparable information period, in this way a trading period is reduced to a small trading section with buy and sell volume that is classified by normal distribution and standard price changes to determine the percentage of buy and sell volume.

A volume bucket is a collection of trades with total volume $V$. In this thesis I calculate buy and sell volume ($V_\tau^B \text{and} V_\tau^S$) using tick-by-tick data and also compromising with volume bars.

$$V_\tau^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i.Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right)$$
$$V_\tau^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i.\left[1 - Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right)\right] = V - V_\tau^B\,, \tag{3.0.25}$$

where $t(\tau)$ is the index of the last volume in the bar included in the $\tau$ volume bucket. $Z$ is the cumulative distribution function(CDF) of the standard normal distribution. $\sigma_{\Delta P}$ is the estimate of the standard derivation of price changes between time bars.

Let $OI_\tau = |V_\tau^B - V_\tau^S|$ be the order imbalance in volume bucket $\tau$ the measure is,of course, an approximation to actual order imbalance as it is based on the probabilistic volume classification. First, how $\mathbb{E}[OI_\tau]$ relates to the rate of trading by showing that it is unaffected by a simple rescaling of trading. Suppose that each time bar's volume is rescaled by a factor of $\beta > 0, V_i^* = \beta V_i$,and that volume imbalance is homogeneously distributed within the bucket. Then expected number of time bars required to fill a bucket will be inversely proportional to $\beta$, $\frac{t(\tau)-t(\tau-1)}{\beta}$. The expected order imbalance, $\mathbb{E}[OI_\tau]$, unaltered ,

$$\mathbb{E}[OI_\tau^*] = \mathbb{E}\left[\left|V_\tau^{*B} - V_\tau^{*S}\right|\right] = \frac{1}{\beta}\mathbb{E}\left[\left|\beta V_\tau^B - \beta V_\tau^S\right|\right] = \mathbb{E}[OI_\tau]\,. \tag{3.0.26}$$

For each period the expected trade imbalance is $\mathbb{E}[|V_\tau^S - V_\tau^B|] \approx \alpha\mu$ and the expected total number of trades is $\mathbb{E}[V_\tau^B + V_\tau^S] = \alpha\mu + 2\varepsilon$. Volume bucketing allows us to estimate this specification very simply. In particular, imagine that we divide the trading day into equal-sized volume buckets and treat each volume bucket as equivalent to a period for information arrival. That means that $V_\tau^B + V_\tau^S$

is constant, and it is equal to $V$, for all $\tau$. We then approximate expected trade imbalance by average trade imbalance over $n$ buckets.

From the values computed above, we can write the *volume-synchronized probability of informed trading*, the VPIN flow toxicity metric, as

$$\text{VPIN} = \frac{\Sigma_{\tau=1}^{n} |V_\tau^S - V_\tau^B|}{nV}\,, \tag{3.0.27}$$

Since $\Sigma_{\tau=1}^{n} |V_\tau^S - V_\tau^B| \approx \alpha\mu$, then $\text{PIN} = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}$, this can be shown that VPIN is an approximate of PIN (Easley et al. [2010]).

# Chapter 4

# Preliminary Analysis of Informed Trading

## 4.1 Introduction

This chapter provides the preliminary analysis of probability informed trading. The PIN proposed by Easley et al. [1996] has been employed in a wide range of market microstructure applications. To estimate this model, the number of buy and sell-initiated trades is needed. However, in many markets this requirement is not publicly available such as, CME, NASDAQ, NYSE. Normally, the trade classification is usually inferred between trade and quote data by comparing the trade price to previous trade prices or to prevailing quotes (Boehmer et al. [2007]). The widely used trade classification algorithm is tick-rule, quote-rule and Lee and Ready [1991] algorithm. Whilst the PIN algorithm is substantially precise, the poor performances of these trade-classifications may have a great impact on the result of the PIN (Ke [2014]). However, this chapter follows Easley et al. [1996] by using the Lee&Ready algorithm.

The chapter is a preliminary empirical study on the PIN, so the standard PIN approach and the Lee&Ready algorithm are used to analyze the informed trading in the Eurodollar futures market. Therefore, I examine the performance of PIN by stimulating the PIN on Eurodollar futures market to investigate how it reacts to major toxicity events. The result shows that the PIN appears to have dropped around the time of the sub-prime crisis (circa 2007) and then bounced back afterwards, and sharply dropped again with the collapse of Lehman Brothers and then bounced back once again indicating that systematic market uncertainty has played a role in information asymmetry. Significantly, the result of the PIN in this thesis is always higher for the Eurodollar futures market than those recorded in the equity market (circa 0.4 to 0.8 versus 0.1 to 0.5; see Abad and Yagüe, 2012; Aslan et al., 2011; Easley et al., 1996); however other PIN studies of interest rate derivatives have found similarly high, if not higher PIN measurements, in particular Kim et al. [2014] find the PIN for the Eurodollar market to be substantially higher on average than that detected in the equity and FX markets.

This paper is organized as follows §(4.2) reviews the current research on detecting informed trading. §(4.3)provides information about trade classification related to the PIN. §(4.4) explores the PIN and presents the empirical results. Finally, §(4.5) provides the preliminary result of PIN analysis in the Eurodollar futures market.

## 4.2 Background and Motivation

Markets are mechanisms that process information. When forward pay-offs follow a semi-martingale process, the act of buying and selling claims on these pay-offs should approach a fair bet. The act of buying and selling assets in a continuous Walrasian is a mechanism by which the fragments of information possessed by each trader are disseminated into the market as a whole. Classical models of markets, such as those that underlie the capital asset pricing model and most base-models of derivatives pricing, assume that, while information provided by traders in the market may not be homogeneous, individual traders make homogeneous statistical assumptions about the quality of their private signals. Thus, the semi-martingale assumption of the evolution of the fundamental value and the resulting price of the asset is preserved. However, most models of market microstructure assume that traders both receive heterogeneous signals and have heterogeneous information about the quality of those signals.

The study of market microstructure in the presence of information asymmetries in the financial securities market has received considerable attention from academics and practitioners, and in their seminal work Grossman and Stiglitz [1980] explain the relationship between market efficiency and pieces of information. They argue that it is impossible for markets to be perfectly efficient in terms of information because if they were, the return for gathering information would be nil, in which case there would be little reason to trade and the markets would eventually collapse. Alternatively, the degree of market inefficiency determines the effort investors are willing to expend to gather and trade information; hence a non-degenerate market equilibrium will only arise when there are sufficient opportunities for profit, i.e. inefficiency, to compensate investors for the cost of trading and information-gathering. The profits earned by these industrious investors may be viewed as economic rents that accrue to those willing to engage in such activities.

Hellwig [1980] outlines some theoretical models of how asset prices evolve when a subset of traders hasprivate information. He explains how a competitive market serves to disseminate information between the market participants. The communication process in a market is usually described by the phrase that the equilibrium price reflects all the available information in the market and communicates it to participants. This area of study has expanded in the last two decades and in terms of approaches, the noisy rational expectation equilibrium of Admati [1985] uses price information for a model of multi-asset risk.

Building on work published in the preceding year on the industrial organization of futures markets, Kyle [1985] constructs a dynamic model of insider trading with sequential auctions, structured to resemble a sequential equilibrium to examine the informational content of prices, the liquidity characteristics of a speculative market, and the value of private information to an insider. Glosten and Milgrom [1985] propose that the trading of informed traders who have superior information leads to a positive bid–ask spread, even when the specialist is risk neutral and makes zero expected profits. Although, the informed trader is risk neutral, the profit that he gets is the loss from uninformed trader. The work of Admati and Pfleiderer [1988] developed a research model in which traders determine when to trade and whether to become privately informed about assets' future returns. The study was inspired by three questions, the first of which was why does trading tend to be concentrated in particular time periods within the trading day? Secondly, why are returns (or price changes) more variable in some periods and less variable in others? Thirdly, why do the periods of higher trading volume also tend to be the periods of higher return variability? Thus, the model development is based on research in which traders determine when to trade and whether to become privately informed about assets' future returns. Easley and

O'Hara [1987] investigate the effect of trade size on security prices when informed traders prefer to trade larger volumes at any price, so that market makers' trading strategies should depend on the size of the trade. As a result, market makers pricing strategies must also depend on trade size with large trades being made at less favorable prices. This model provides an explanation for the price effect of block trading and demonstrates that both the size and sequence of trades matter when determine the price-trade size relationship.

Easley and O'Hara [1992] developed a model that observes the informed trading of a sequential trade model from the learning process of order flow similar to that of Glosten and Milgrom [1985] and Easley and O'Hara [1987] in which arriving orders are updated in a probabilistic fashion, independent of any parameters. This model was nicknamed the PIN and has subsequently proved to be very popular because of its ease of implementation. It is used to measure the impact of different portions of information from traders on market liquidity and price formation. The PIN can explain the link between the existence of information, trading time and security prices. The time affects the prices, trading can move price quotes, and the time between trades affects the spread; for instance, the spread decreases as the time increases and whether informed trader trade or do not trade will correspond to the volume. Informed traders will reveal their private information for trade; they will buy (sell) if they observe good (bad) news. Therefore, this method is built on a pattern of buy and sell orders, which is interpreted as stemming from information arrivals that appear in the market.

Easley et al. [1996] investigate the differences in spreads for active and infrequently traded stocks by using an information-based trading approach with a sample of stocks listed on the New York Stock Exchange (NYSE). The findings show that the probability of informed trading is lower for active than inactive stock; however, although high volume of stock tends to have more informed traders, the increase of uninformed traders has a greater effect. Easley et al. [1996] proceed with a mixed continuous and discrete time-sequential trade model of the trading process in which the trading day's trade arises from a group of informed traders and uninformed traders, and also price movements arise from the quotes. Informed trading occurs on the trading day and the decisions are made based on their private information. This information trade-based approach is different in application from that of Hasbrouck [1991] who examines the information in trade innovations as a vector autoregressive.

Possible explanations of price processes and the number of trades influenced by information content are demonstrated by Easley et al. [1997a,b] when they examine the trade and price process in the electronic market order flow using high-frequency data, as well as by Brown et al. [1999]. Easley et al. [1997b] criticize Kyle [1985] and Glosten and Milgrom [1985], both of which identify the behaviour of uninformed traders who are assumed to be serially independent; however, if uninformed traders act strategically, then these microstructure models are misspecified. The Easley et al. [1997b] model enables uniformed traders to become buyers and sellers following independent or dependent processes. The model includes variations in trade sizes to estimate informational versus trade volumes. The results of these studies show that both large and small traded stocks are widely affected by private information; however, large trades have approximately twice the information content of small trades and uninformed traders who make decisions based on their recent trading are more likely to buy (sell) when the last trade was a (buy) sell.

The PIN has been widely adopted in a variety of studies in empirical financial literature; for instance, Easley et al. [2001] use NYSE common stocks that had two to one splits in 1995 to investigate how stock splits affect trading in the presence of uninformed and informed traders. After stock splits, there is a slight increase in uninformed trading and a tendency to execute trades using market orders. This evidence is consistent with the increase in uninformed trading so that the problem of adverse selection is not materially reduced. Also, the parameter estimates after stock splits suggest that it does not significantly change the stock information environment. Moreover, the results of the Heidle and Huang [2002] information-based study of the probability of informed trading in dealer and auction markets show that associated changes in the probability of trading with informed traders are related to changes in the bid–ask spread; also, there is more probability of an informed traders' confrontation in dealer markets than in auction markets. Grammig et al. [2001] extend the Easley et al. [1996] model to allow for a simultaneous estimation of two parallel markets in order to analyse whether the number of anonymous traders is related to information-based trading using the German stock market to answer the question, which trading platform do traders prefer? They find that the probability of informed trading is significantly lower in floor-based trading systems than in anonymous electric trading systems.

Easley et al. [2002] and Aslan et al. [2011] expand the main financial literature that focuses on asset pricing to theoretical market microstructure pricing models in the presence of asymmetric information. They investigate the effect of the role of private information-based trading on asset returns. Easley et al. [2002] uses individual NYSE-listed stocks between 1983 and 1998 to measure the probability of information-based trading incorporated with a Fama and French asset pricing framework. The main outcome of their study is that information does affect asset prices; specifically, where there is a 10% difference in the probability of informed trading between two stocks, this leads to a 2.5% percent difference in their expected return. Similarly, Aslan et al. [2011] investigate the link between private information-based trading and accounting and asset pricing by developing a proxy for the PIN (PPIN) to investigate the role of information risk in asset pricing over longer time periods. They conclude that informed trading captured by the PPIN is both statistically and economically significant for asset returns; for instance, firms with higher PPINs have higher returns, and this conclusion is robust in every asset pricing structure explored.

The structure of private and public information and post-earnings announcement drift by using the PIN has been extensively analysed by Vega [2006] using a comprehensive public news database. The results of this study are that informed traders tend to learn about the true value of an asset and stocks linked with a high PIN from public surprising news; conversely, low media coverage causes an insignificant drift. Also, most small stocks realise a greater post-announcement drift than large stocks, which tend to be more transparent. Another analysis involving the PIN appears in Ascioglu et al. [2008] who try to fill the gap between the investment issue in corporate finance literature and liquidity in the market microstructure. This study provides strong evidence that asymmetric information decreases firm investment and increases the sensitivity of investment expenditure to fluctuation in internal funds. It also illustrates that the strength of the results of the PIN is consistent with the evidence in the market microstructure, that the PIN is by far the most relevant and direct measure of asymmetric information and it should be one of significant factor for asset pricing in the Fama and French framework. Another study involving the corporate investment decision is that of

Chen et al. [2007], the results of which show that the measure of the amount of private information in stock prices and the probability of informed trading have a strong positive effect on the sensitivity of corporate investment to stock prices. Also, Brockman and Yan [2009] study the stock value with the relationship of the block holders and the amount of specific information. The results strongly support that block ownership and firm-specific information have a significant impact on the stock return.

Studies of the market microstructure in the presence of the PIN that impacts the price of risky assets were undertaken by Kang [2010], Idier and Nardelli [2011] and Chen and Zhao [2012]. Firstly, Kang [2010] investigates the relationship between the probability of information-based trading and the January effect, the so-called "January PIN effect". He finds that the mean stock returns decrease with the PIN in January, contrary to other calendar months, and that this effect is more significant for small stocks. As stated in his work, the January PIN effect is associated with selling pressure in December, especially of small stock. As a result, this seasonal effect leads to a negative relationship between the PIN and returns, and when the price bounces back in January, low-PIN stocks will exhibit a larger return within a small stock group. Secondly, Idier and Nardelli [2011] apply the PIN model developed by Easley et al. [1996] to analyse the role and impact of information-based trading on the Euro overnight market rate. They find that the PIN sharply declined after the reform of the Eurosystem's operational framework in March 2004 and the recent financial market turmoil. Finally, Chen and Zhao [2012] use the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX) to investigate the effect of informed trading (PIN) and information uncertainty when determining the price momentum. The results show that momentum trading strategies based on buying high-uncertainty good-news stocks and short selling high-uncertainty bad-news stocks only work well when applied to stock with a high PIN. In contrast, this pattern is not exhibited with low-PIN stocks, even when the uncertainty level of those stocks is high.

In view of all that has been mentioned so far, one may suppose that the studies of market microstructure in the presence of informed trading by using a probability of informed trading approach is a stabilized measurement tool for asymmetry information. While the PIN approach is quite popular for securities trading, only a tiny portion of the literature is involved with the interest rate derivative market. Thus, the gap in these previous studies offers a great opportunity to test some of the current measures of informed trading such as the PIN approach of Easley et al. [1996] with a very liquid interest rate derivative market.

## 4.3 PIN and Trade Classification Algorithms

Since the classification of trades is one of the fundamental issues for analysis of the PIN, it is essential to identify the direction of trading in terms of buying or selling. An important feature of the ED futures contract database is that it only records transactions that include Trades and Quotes, in particular price and quantity, but is not able to classify the trades as buy or sell. In order to determine whether a trade is initiated by a buy-side or sell-side, it is necessary to use trade classification techniques developed in the literature.

I mentioned in the introduction of this chapter that the accuracy of PIN belongs to the trade classification. So, how well does the trade classification work? To answer this question, therefore, the comparison of standard trade classifications such as, tick rule, quote rule and, Lee&Ready algorithm are provided based on the existing works. For example, Ellis et al. [2000] report the accuracy of these trade classifications, they find that the tick rule, the quote rule and the Lee&Ready algorithm correctly 77.66%, 76.40% and 81.05% of trades, respectively. Odders-White [2000] report the Lee&ready successively 85% of trades. Chakrabarty et al. [2012] report tick rule correctly classifies 90.80% of trade volume. To illustrate how trade classification effects PIN, Boehmer et al. [2007] report that the misclassification of Lee&Ready algorithm leads to a large bias of 18% of the estimate PIN for their sample stock in the NYSE. These different results of trade classification cause the difficulty of PIN computation. However, the PIN computation in this thesis follows the original work of Easley et al. [1996] by using the Lee and Ready [1991] algorithm for trade classification. Therefore, a brief description of trade classification approaches is provided in the following section.

### 4.3.1 Tick Rule

The tick rule is widely used for trade classification and captures the natural intuition that buyers pay a higher price and sellers get a lower price. Therefore, it assumes that trades are buys if the trade price is higher than the previous one; on the other hand, if the trade price is lower than the previous one, it is assumed that the trade is a sell. If the trade price remains stable compared to the previous price, the trade is assumed to be the same as the previous trade. Price data (Level 1 data) is needed for this classification and every trade can be classified.

Easley et al. [2012] criticizes this approach because the tick rule focuses only on price movements. The questions about the accuracy of this classification are especially relevant in HFT. In the HFT, where the price moves up and down at super-speed succession within the order book, trades tend to be misclassified. Another problem is misclassification from hidden orders when there is no price movement. For instance, if the direction of the previous trade was a buy followed by trade with a hidden sell order at midpoint the same as the buy price, this trade will later be incorrectly classified as a buy.

### 4.3.2 Quote Rule

The quote rule classification demands more data than the tick rule, since it requires the best bid and ask quotes. This method classifies a trade as a buy when it occurs above the best bid and ask midpoint and conversely, it will be classified as a sell when it occurs below the midpoint. When the trade appears at the midpoint it is unclassified. Despite the fact that this method requires more data than the tick rule and should therefore provide more accurate results in classifying the trade, it is problematic given that a large number of trades appearing at mid-quote are misclassified [Easley et al., 2012]. This incorrect classification should increase in HFT when quotes change more frequently than trades execute, which means that a trade may be classified as ask at the current trade when, in fact, it took place as a bid. Equally, this incorrect classification occurs in less liquid trading with thinner order books.

### 4.3.3 Lee–Ready Algorithm

Another widely used technique to classify the direction of trade is the Lee and Ready [1991] algorithm, which is a combination of the tick rule and the quote rule. Lee and Ready examined a sample of the trading of 150 NYSE companies in 1988 when criticizing the tick rule, and the Lee–Ready algorithm also identifies some problems with the quote rule. The original Lee–Ready algorithm specified a 5 s delay in comparing trades to quotes. The following rules apply when using this method to classify trades:

1. For a trade, if the inside (best bid (ask)) quote has not changed within the preceding 5 s and the price is equal to the bid (ask), the trade is then classified as a sell (buy).
2. When the current quotes appear for less than 5 s, the trade direction is compared with the previous quotes.
3. When the trade price is outside the inside spread, so that the current trade price is lower (higher) than the best bid (ask), the trade is classified as a sell (buy).
4. If the transaction is at the midpoint the Lee–Ready algorithm uses a tick rule to classify the trade.
5. Tick rule: when the price is between the spread but not at the midpoint, the trade is classified as a sell (buy) when the price is closer to the bid (ask).

The Lee–Ready algorithm is the main classification for the PIN because of its accuracy. Lee and Radhakrishna [2000] reported a 93% success rate when they applied this method to 144 NYSE stocks traded between November 1990 and January 1991 using a TORQ database. Finucane [2000] reported an 84.4% accuracy of 337,667 trades for all 144 firms in the NYSE TORQ database, while Odders-White [2000] found 85% accuracy using a TORQ database of 318,364 transactions. Ellis et al. [2000] used a data set of 313 NASDAQ firms that contained 2,433,019 trades and Lee–Ready correctly classified 81.1% of them. However, there is some evidence that the Lee–Ready classification is less accurate for HFT, as Chakrabarty et al. [2012] found when using TAQ data from INET for

342 stocks in NASDAQ, since the results showed that the Lee–Ready algorithm had misclassified 31% of all trades.

## 4.4 Preliminary Analysis and Implications

My empirical analysis follows the PIN approach as described in section 3.0.4. The PIN is a measure of asymmetric information on a trading event between informed and uninformed trading, which was developed on the theoretical framework of Easley and O'Hara [1992], although the original PIN approach was introduced by Easley et al. [1996]. They produced a mixed discrete-time and continuous-time sequential model of the trading mechanism, in which trades emerge when three classifications of economic agents – namely market makers, informed traders and uninformed traders collaborate. Abad and Yagüe [2012], explain that on a trading day, a market maker is a competitive or strategic trader, informed and uninformed traders are risk neutral with probability of profit or loss at 0.5. Also, informed and uninformed traders arrive in the market with their unique arrival rate with a view to a trading game between the liquidity provider and trader. Despite this, the PIN is not a direct measurement observation, but rather a parameter of a microstructure model which is estimated by the numerical maximization of the likelihood function. This is explained by Easley et al. [1996], as considering the likelihood of the arrival of three different types of information on a trading day: no news, good news, and bad news. This set of options for market agents can be summarized in a standard tree diagram representing the trading process. This diagram gives the structure of the trading process, where $\alpha$ is the probability of an information event, $\delta$ is the probability of a low signal, $\mu$ is the rate of informed trade arrivals, and $\varepsilon$ is the rate of uninformed buy and sell rate arrivals.

**Table 4.1: Comparison between averages PIN estimates across various studies and markets.**

| Author | Asset | Sample Period | Min–max | Mean |
|---|---|---|---|---|
| Easley et al. [1996] | NYSE – 90 stocks | Oct. 1, 1990 to Dec. 23, 1990 | 0.120–0.342 | 0.197 |
| Idier and Nardelli [2011] | Euro overnight interbank rate, Money market | Dec. 2000 to Mar. 2008 | 0.200–0.580 | 0.480 |
| Easley et al. [2011] | E-mini S&P 500 (CME) | Jan. 1, 2008 to Oct. 30, 2010 | 0.205–0.830 | 0.393 |
| Easley et al. [2012] | E-mini S&P 500 (CME) | Jan. 1, 2008 to Oct. 30, 2010 | 0.205–0.830 | 0.393 |
| | T-Note (CBOT) | Jan. 1, 2008 to Oct. 30, 2010 | 0.200–0.800 | 0.401 |
| | EUR/USD(CME) | Jan. 1, 2008 to Oct. 30, 2010 | 0.150–0.780 | 0.327 |
| | Brent Crude Oil (ICE) | Jan. 1, 2008 to Oct. 30, 2010 | 0.200–0.770 | 0.384 |
| | Silver (COMEX) | Jan. 1, 2008 to Oct. 30, 2010 | 0.200–0.840 | 0.411 |
| Abad and Yagüe [2012] | Spanish Stock Exchange – 15 stocks | Jan. 1, 2009 to Dec. 31, 2009 | 0.104–0.501 | 0.227 |
| Kim et al. [2014] | Intra day trading – Eurodollar Futures (CME) | Jan. 3, 2005 to Dec. 29, 2006 | 0.760–0.970 | 0.880 |
| Yan and Zhang [2014] | NYSE/AMEX stocks that have data in the ISSM and TAQ databases | Jan, 1, 1983 to Dec. 31, 2004 | 0.177–0.227 | 0.201 |
| **My PIN** | 40 Eurodollar Future contracts | Jan. 1, 1996 to Jan 1, 2014 | 0.369–0.992 | 0.688 |

Note: This table presents a sample result of PIN between other PIN and the PIN in my research. The result from this table indicates that the PIN in the ED market is always high and higher than the equities market, however, the PIN result from my work coincides with those of Kim et al. [2014] on the Eurodollar futures market.

I compute the PIN measure in the following sections using various rolling windows to record time series variation in the Eurodollar Futures PIN. The estimated PIN averages over 0.5 for the majority of the sample and this is high relative to comparable equity market studies, see Table 4.1. However, the results coincide with those of Kim et al. [2014] on the CME Globex trades within Eurodollar futures samples and Easley et al. [2012] for the US Dollar Treasury note.

Although the PIN estimates range between 0.1 and 0.8 in the equities market, in the derivatives market the range is higher, with minimum and maximum values of 0.20 and 0.97, respectively. Also, in the derivatives market the min–max spread is higher than in the equities market, particularly in the interest rate derivatives market. The highest mean of PIN appears in Kim et al. [2014] with value of 0.880 ;however, the highest actual value of PIN is shown in this work with the maximum value of 0.992. Interestingly, the value of PIN in Eurodollar Futures is higher than the equity market and always high as in this work the PIN is mostly higher than 0.5 and in Kim et al. [2014] the PIN varies between 0.76 and 0.97. Given the higher PIN that appears in the Eurodollar market in both these cases, it would be interesting to investigate this in the future.

### 4.4.1 Eurodollar and PIN Analysis

The data used in this research is the Eurodollar futures data traded on the CME's Globex platform (see section 2.3). The data are directly from the CME tapes for the period January 1, 1996 to January 1, 2014. In total, I included the 40 quarterly ED contracts that are available from the Thomson Reuter database via Tick history system. The PIN calculations were computed for each day by multiple instances of Matlab, on a large multiprocessor workstation with 0.5TB of RAM and virtual memory. The total computation time was around four weeks. The data for the 40 Eurodollar future contracts is separated into two different time-frames, the first of which was from January 1, 1996 to July 31, 2007, and the second, from August 1, 2007 to January 1, 2014. The descriptive statistics for each ED futures contract in the two periods are shown in Tables 4.2 and 4.3.

In terms of the first period, the highest mean bid/ask spread belongs to EDU1, and EDZ0 has the highest mean trades returns. The smallest mean bid–ask spread belongs to EDH6, and EDU6 has the lowest mean trade returns. As for the second period, EDM7 presents both the largest mean spread and mean returns, while EDZ9 shows the lowest mean spread, EDZ9 shows the lowest mean spread, and EDM9 presents the lowest mean return.

#### 4.4.1.1 Maximum Likelihood and PIN Estimation

The parameters of the PIN term structure model are estimated in this section. In fact, the trade process depends on four parameters – namely the probability of an information event; the probability of bad or good news; the arrival rate of informed traders; and the arrival rate of uninformed traders, presented as $\alpha, \delta, \mu$ and $\varepsilon$ respectively. These parameters determine the probability of informed trading that is observed by order imbalance on information-based trading in the Eurodollar Futures market.

These four parameters of the trading process for each Eurodollar future series in a total of 40 data sets are estimated by maximizing the likelihood function, as described in Section 4. The mean estimate of the parameters and their standard deviation for each series are provided in Tables 4.4 and 4.5. Table 4.4 presents the results obtained from a preliminary analysis of the mean estimated parameters by Eurodollar futures series contracts.

The mean value of the information event parameter $\alpha$ for Eurodollar contracts is 0.6444, which indicates that, on average, 3/5 of trading days are information driven. The second information parameter in the PIN approach is $\delta$, which is the probability of bad news, so that the signal is low (resp. high) during informed days with a probability $\delta$ (resp. $1 - \delta$). These results indicate that a high signal is observed with an estimated probability $\delta$ of 0.3517, and $1 - \delta$ is 0.6483. This means that the order is more likely to be buy than sell when an information event occurs. Thus, traders tend to believe that information in trading day-driven orders reveals an excess liquidity in demand on the Eurodollar futures market more

**Table 4.2: The descriptive statistics for 40 Eurodollar Futures from 01/01/1996 to 31/07/2007 (BSP = basis point = $1e^{-5}$).**

| ED codes | Mean spread (BPS) | Median spread (BPS) | Mode spread (BPS) | Max. spread | Min. spread | Std spread (BPS) | Skew spread | Kurt. spread | Mean returns ($\frac{1}{1000}$ (BSP)) | Median returns | Mode returns | Max. returns | Min. returns | Std returns (BSP) | Skew returns | Kurt. returns |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EDH0 | 1.5 | 1.1 | 1.1 | 0.05 | 0 | 3.3 | 42.31 | 2,718 | 9.4 | 0 | 0 | 0.01 | –0.02 | 2.0 | –10 | 2,682 |
| EDH1 | 2.2 | 2.1 | 0.0 | 0.05 | 0 | 4.4 | 27.41 | 1,243 | 9.9 | 0 | 0 | 0.01 | –0.02 | 1.8 | –17 | 3,318 |
| EDH2 | 2.2 | 0.5 | 0.0 | 0.07 | 0 | 8.5 | 23.64 | 806 | 8.1 | 0 | 0 | 0.01 | –0.06 | 2.2 | –166 | 42,329 |
| EDH3 | 2.1 | 0.5 | 0.0 | 0.04 | 0 | 7.4 | 14.83 | 334 | 6.0 | 0 | 0 | 0.01 | –0.06 | 1.8 | –199 | 60,534 |
| EDH4 | 1.7 | 1.0 | 0.5 | 0.11 | 0 | 4.8 | 37.79 | 6,385 | 6.2 | 0 | 0 | 0.01 | –0.06 | 2.0 | –224 | 70,601 |
| EDH5 | 0.9 | 0.5 | 0.5 | 2.25 | 0 | 28.0 | 798.46 | 643,160 | 1.6 | 0 | 0 | 0.03 | –0.03 | 0.9 | –74 | 60,354 |
| EDH6 | 0.6 | 0.5 | 0.5 | 0.06 | 0 | 1.0 | 83.01 | 18,832 | –1.2 | 0 | 0 | 0.04 | –0.04 | 0.9 | –37 | 133,760 |
| EDH7 | 0.7 | 0.5 | 0.5 | 0.06 | 0 | 2.8 | 67.07 | 6,078 | –1.7 | 0 | 0 | 0.00 | –0.03 | 0.6 | –102 | 39,999 |
| EDH8 | 0.7 | 0.5 | 0.5 | 0.06 | 0 | 1.2 | 134.67 | 41,985 | 0.9 | 0 | 0 | 0.01 | –0.01 | 0.6 | –9 | 2,470 |
| EDH9 | 0.9 | 0.5 | 0.5 | 0.12 | 0 | 1.0 | 152.55 | 68,286 | 3.9 | 0 | 0 | 0.04 | –0.04 | 1.6 | –7 | 51,223 |
| EDM0 | 1.6 | 1.6 | 0.0 | 0.05 | 0 | 1.9 | 51.65 | 6,883 | 10.5 | 0 | 0 | 0.01 | –0.01 | 1.8 | –2 | 1,432 |
| EDM1 | 2.3 | 2.1 | 0.0 | 0.05 | 0 | 4.6 | 26.50 | 1,074 | 9.8 | 0 | 0 | 0.01 | –0.04 | 2.1 | –65 | 12,384 |
| EDM2 | 1.7 | 0.5 | 0.0 | 0.05 | 0 | 5.5 | 20.31 | 690 | 7.6 | 0 | 0 | 0.01 | –0.05 | 2.3 | –124 | 29,914 |
| EDM3 | 2.2 | 1.0 | 0.0 | 0.05 | 0 | 6.4 | 23.39 | 1,083 | 6.3 | 0 | 0 | 0.01 | –0.05 | 1.8 | –143 | 37,740 |
| EDM4 | 1.4 | 0.5 | 0.5 | 0.09 | 0 | 3.4 | 37.07 | 5,656 | 5.0 | 0 | 0 | 0.01 | –0.06 | 1.6 | –235 | 82,308 |
| EDM5 | 0.7 | 0.5 | 0.5 | 0.04 | 0 | 1.8 | 34.65 | 2,570 | 2.2 | 0 | 0 | 0.03 | –0.03 | 0.8 | –33 | 86,622 |
| EDM6 | 0.6 | 0.5 | 0.5 | 0.04 | 0 | 1.2 | 128.12 | 25,789 | 1.5 | 0 | 0 | 0.04 | –0.04 | 0.9 | –51 | 142,750 |
| EDM7 | 0.7 | 0.5 | 0.5 | 0.04 | 0 | 2.7 | 57.02 | 3,837 | 1.5 | 0 | 0 | 0.02 | –0.02 | 0.7 | –37 | 24,936 |
| EDM8 | 1.0 | 0.5 | 0.5 | 0.06 | 0 | 1.7 | 84.40 | 13,153 | 1.5 | 0 | 0 | 0.00 | –0.01 | 0.6 | –9 | 1,794 |
| EDM9 | 1.0 | 0.5 | 0.5 | 0.05 | 0 | 2.0 | 45.00 | 4,634 | 5.7 | 0 | 0 | 0.01 | –0.03 | 1.3 | –49 | 10,723 |
| EDU0 | 1.8 | 1.6 | 0.0 | 0.03 | 0 | 2.5 | 31.97 | 1,890 | 10.4 | 0 | 0 | 0.01 | –0.01 | 2.1 | –4 | 1,523 |
| EDU1 | 2.6 | 2.1 | 0.0 | 0.07 | 0 | 9.0 | 28.83 | 1,342 | 8.9 | 0 | 0 | 0.01 | –0.04 | 2.1 | –88 | 17,866 |
| EDU2 | 1.4 | 0.5 | 0.0 | 0.05 | 0 | 4.5 | 29.47 | 1,520 | 6.5 | 0 | 0 | 2.23 | –2.23 | 88.0 | 0 | 64,403 |
| EDU3 | 2.4 | 1.0 | 0.5 | 0.02 | 0 | 6.7 | 8.89 | 131 | 6.7 | 0 | 0 | 0.01 | –0.06 | 2.0 | –202 | 60,072 |
| EDU4 | 1.1 | 0.5 | 0.5 | 0.06 | 0 | 2.8 | 44.94 | 5,681 | 3.7 | 0 | 0 | 0.01 | –0.05 | 1.2 | –220 | 85,682 |
| EDU5 | 0.6 | 0.5 | 0.5 | 0.03 | 0 | 1.2 | 64.37 | 7,310 | 1.2 | 0 | 0 | 0.04 | –0.04 | 0.8 | –20 | 115,010 |
| EDU6 | 0.6 | 0.5 | 0.5 | 0.04 | 0 | 2.2 | 83.66 | 9,009 | –1.7 | 0 | 0 | 0.04 | –0.04 | 0.9 | –41 | 149,020 |
| EDU7 | 0.6 | 0.5 | 0.5 | 0.05 | 0 | 1.0 | 270.08 | 113,040 | –0.1 | 0 | 0 | 0.01 | –0.02 | 0.6 | –40 | 12,697 |
| EDU8 | 0.7 | 0.5 | 0.5 | 0.04 | 0 | 1.2 | 85.98 | 15,599 | 2.2 | 0 | 0 | 0.01 | –0.01 | 0.7 | –11 | 2,296 |
| EDU9 | 1.1 | 1.1 | 0.5 | 0.05 | 0 | 2.5 | 37.43 | 2,714 | 6.9 | 0 | 0 | 0.01 | –0.03 | 1.4 | –56 | 11,566 |
| EDZ0 | 2.0 | 1.6 | 0.0 | 0.06 | 0 | 3.4 | 1.34 | 2,288 | 12.2 | 0 | 0 | 0.01 | –0.01 | 1.9 | 0 | 1,368 |
| EDZ1 | 2.4 | 1.5 | 0.0 | 0.05 | 0 | 6.7 | 21.53 | 668 | 9.5 | 0 | 0 | 0.01 | –0.06 | 2.5 | –149 | 35,570 |
| EDZ2 | 1.8 | 0.5 | 0.0 | 0.04 | 0 | 7.7 | 18.65 | 425 | 5.9 | 0 | 0 | 0.01 | –0.05 | 1.9 | –164 | 47,837 |
| EDZ3 | 2.2 | 0.5 | 0.5 | 0.06 | 0 | 7.3 | 13.92 | 418 | 6.6 | 0 | 0 | 0.01 | –0.06 | 1.9 | –213 | 65,295 |
| EDZ4 | 0.9 | 0.5 | 0.5 | 0.07 | 0 | 2.6 | 53.35 | 8,458 | 2.0 | 0 | 0 | 0.04 | –0.04 | 1.2 | –77 | 63,379 |
| EDZ5 | 0.6 | 0.5 | 0.5 | 0.05 | 0 | 1.4 | 78.46 | 10,575 | 1.2 | 0 | 0 | 0.04 | –0.04 | 0.8 | –12 | 138,980 |
| EDZ6 | 0.7 | 0.5 | 0.5 | 0.06 | 0 | 3.6 | 78.76 | 7,154 | –1.6 | 0 | 0 | 0.01 | –0.02 | 0.6 | –74 | 29,202 |
| EDZ7 | 0.7 | 0.5 | 0.5 | 0.16 | 0 | 1.2 | 385.68 | 433,280 | 0.5 | 0 | 0 | 0.01 | –0.01 | 0.6 | –6 | 2,095 |
| EDZ8 | 0.8 | 0.5 | 0.5 | 0.03 | 0 | 1.9 | 42.25 | 2,436 | 0.3 | 0 | 0 | 0.01 | –0.01 | 0.9 | –10 | 4,836 |
| EDZ9 | 1.4 | 1.1 | 1.1 | 0.06 | 0 | 3.5 | 30.23 | 1,619 | 9.1 | 0 | 0 | 0.01 | –0.02 | 1.8 | –16 | 2,878 |

**Table 4.3: The descriptive statistics for 40 Eurodollar Futures contracts from 01/08/2007 to 01/01/2014 (BSP = basis point = $1e^{-5}$).**

| ED codes | Mean spread (BPS) | Median spread (BPS) | Mode spread (BPS) | Max. spread | Min. spread | Std spread (BPS) | Skew spread | Kurt. spread | Mean returns ($\frac{1}{1000}$ (BSP)) | Median returns | Mode returns | Max. returns | Min. returns | Std returns (BSP) | Skew returns | Kurt. returns |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EDH0 | 1.0 | 0.5 | 0.5 | 0.16 | 0 | 1.9 | 131.61 | 70,698 | 0.3 | 0 | 0 | 0.01 | –0.06 | 0.6 | –610 | 607,830 |
| EDH1 | 1.3 | 0.5 | 0.5 | 0.18 | 0 | 3.9 | 87.38 | 27,615 | 0.4 | 0 | 0 | 0.01 | –0.05 | 0.6 | –535 | 493,030 |
| EDH2 | 1.3 | 0.5 | 0.5 | 0.18 | 0 | 3.8 | 93.19 | 29,520 | 0.5 | 0 | 0 | 0.00 | –0.04 | 0.5 | –376 | 302,630 |
| EDH3 | 1.7 | 0.5 | 0.5 | 0.06 | 0 | 3.8 | 9.57 | 296 | 0.6 | 0 | 0 | 0.04 | –0.04 | 0.7 | –122 | 205,080 |
| EDH4 | 1.4 | 0.5 | 0.5 | 0.18 | 0 | 6.6 | 38.25 | 4,723 | 4.2 | 0 | 0 | 0.02 | –0.02 | 1.0 | 0 | 4,966 |
| EDH5 | 1.6 | 0.5 | 0.5 | 0.48 | 0 | 1.6 | 136.37 | 30,339 | 4.5 | 0 | 0 | 0.01 | 0.00 | 0.6 | 0 | 297 |
| EDH6 | 1.4 | 0.5 | 0.5 | 0.04 | 0 | 9.2 | 19.51 | 410 | 4.8 | 0 | 0 | 0.01 | –0.01 | 0.9 | 0 | 206 |
| EDH7 | 1.9 | 0.5 | 0.5 | 0.18 | 0 | 12.0 | 30.70 | 2,337 | 6.2 | 0 | 0 | 0.01 | –0.01 | 1.1 | 0 | 257 |
| EDH8 | 1.9 | 0.5 | 0.5 | 0.19 | 0 | 11.0 | 29.64 | 2,167 | 1.6 | 0 | 0 | 0.01 | –0.03 | 0.9 | –45 | 15,031 |
| EDH9 | 1.1 | 0.5 | 0.5 | 0.16 | 0 | 9.8 | 64.45 | 4,384 | 0.4 | 0 | 0 | 0.01 | –0.03 | 0.4 | –206 | 147,460 |
| EDM0 | 1.0 | 0.5 | 0.5 | 0.08 | 0 | 2.2 | 44.83 | 5,650 | 0.4 | 0 | 0 | 0.01 | –0.04 | 0.5 | –437 | 388,800 |
| EDM1 | 1.5 | 0.5 | 0.5 | 0.18 | 0 | 5.1 | 45.22 | 8,734 | 0.4 | 0 | 0 | 0.01 | –0.05 | 0.6 | –546 | 500,910 |
| EDM2 | 1.4 | 0.5 | 0.5 | 0.07 | 0 | 3.4 | 33.23 | 3,024 | 0.5 | 0 | 0 | 0.00 | –0.03 | 0.4 | –219 | 145,100 |
| EDM3 | 1.6 | 0.5 | 0.5 | 0.04 | 0 | 4.2 | 16.02 | 557 | 0.6 | 0 | 0 | 0.00 | –0.05 | 0.7 | –277 | 190,220 |
| EDM4 | 1.5 | 0.5 | 0.5 | 0.09 | 0 | 7.0 | 21.31 | 593 | 4.0 | 0 | 0 | 0.01 | 0.00 | 0.9 | 0 | 153 |
| EDM5 | 1.3 | 0.5 | 0.5 | 0.63 | 0 | 9.6 | 44.04 | 11,161 | 2.0 | 0 | 0 | 0.01 | –0.01 | 0.7 | 1 | 377 |
| EDM6 | 1.4 | 0.5 | 0.5 | 0.04 | 0 | 9.5 | 19.05 | 395 | 5.0 | 0 | 0 | 0.14 | –0.14 | 2.3 | 0 | 301,010 |
| EDM7 | 4.5 | 0.5 | 0.5 | 4.61 | 0 | 34.9 | 132.00 | 17,439 | 6.8 | 0 | 0 | 0.01 | –0.01 | 1.2 | 0 | 200 |
| EDM8 | 1.5 | 0.5 | 0.5 | 2.31 | 0 | 19.0 | 899.14 | 1,072,400 | 1.0 | 0 | 0 | 0.01 | –0.03 | 0.8 | –67 | 27,438 |
| EDM9 | 1.0 | 0.5 | 0.5 | 0.16 | 0 | 2.1 | 110.22 | 49,876 | 0.3 | 0 | 0 | 0.01 | –0.05 | 0.5 | –512 | 492,050 |
| EDU0 | 1.1 | 0.5 | 0.5 | 0.07 | 0 | 1.2 | 27.63 | 2,217 | 0.4 | 0 | 0 | 0.01 | –0.04 | 0.5 | –430 | 377,590 |
| EDU1 | 1.4 | 0.5 | 0.5 | 0.18 | 0 | 6.1 | 40.62 | 5,931 | 0.5 | 0 | 0 | 0.01 | –0.04 | 0.5 | –335 | 261,480 |
| EDU2 | 1.5 | 0.5 | 0.5 | 12.92 | 0 | 27.0 | 4,774.40 | 23,235,000 | 0.5 | 0 | 0 | 0.00 | –0.03 | 0.4 | –271 | 190,970 |
| EDU3 | 1.5 | 0.5 | 0.5 | 0.04 | 0 | 0.6 | 20.72 | 586 | 0.5 | 0 | 0 | 0.02 | –0.05 | 1.0 | –76 | 39,028 |
| EDU4 | 1.6 | 0.5 | 0.5 | 0.19 | 0 | 9.6 | 23.24 | 1,508 | 4.2 | 0 | 0 | 0.00 | 0.00 | 0.8 | 0 | 172 |
| EDU5 | 1.2 | 0.5 | 0.5 | 0.04 | 0 | 7.8 | 22.93 | 566 | 4.6 | 0 | 0 | 0.01 | –0.01 | 0.7 | 0 | 984 |
| EDU6 | 1.6 | 0.5 | 0.5 | 0.07 | 0 | 11.0 | 19.01 | 404 | 5.5 | 0 | 0 | 0.17 | –0.17 | 2.8 | 0 | 311,730 |
| EDU7 | 1.7 | 0.5 | 0.5 | 2.38 | 0 | 15.0 | 1,051.20 | 1,648,500 | 3.9 | 0 | 0 | 0.01 | –0.01 | 1.1 | 0 | 213 |
| EDU8 | 1.3 | 0.5 | 0.5 | 1.82 | 0 | 11.0 | 773.70 | 1,198,900 | 0.7 | 0 | 0 | 0.01 | –0.02 | 0.6 | –67 | 30,066 |
| EDU9 | 1.0 | 0.5 | 0.5 | 0.16 | 0 | 1.7 | 135.07 | 89,802 | 0.3 | 0 | 0 | 0.01 | –0.05 | 0.5 | –556 | 552,290 |
| EDZ0 | 1.3 | 0.5 | 0.5 | 0.18 | 0 | 4.3 | 74.98 | 19,688 | 0.4 | 0 | 0 | 0.01 | –0.05 | 0.6 | –594 | 571,520 |
| EDZ1 | 1.5 | 0.5 | 0.5 | 0.18 | 0 | 6.5 | 32.37 | 3,584 | 0.5 | 0 | 0 | 0.01 | –0.03 | 0.4 | –286 | 211,740 |
| EDZ2 | 1.5 | 0.5 | 0.5 | 0.07 | 0 | 3.4 | 20.63 | 1,921 | 0.5 | 0 | 0 | 0.00 | –0.04 | 0.5 | –348 | 264,280 |
| EDZ3 | 1.3 | 0.5 | 0.5 | 0.05 | 0 | 5.9 | 27.31 | 887 | 0.5 | 0 | 0 | 0.00 | –0.05 | 1.0 | –87 | 43,136 |
| EDZ4 | 1.5 | 0.5 | 0.5 | 0.19 | 0 | 9.4 | 24.15 | 1,569 | 4.3 | 0 | 0 | 0.01 | –0.01 | 0.7 | 0 | 249 |
| EDZ5 | 1.2 | 0.5 | 0.5 | 0.06 | 0 | 8.2 | 21.36 | 493 | 4.4 | 0 | 0 | 0.17 | –0.17 | 2.3 | 0 | 493,600 |
| EDZ6 | 1.8 | 0.5 | 0.5 | 0.13 | 0 | 12.0 | 15.60 | 275 | 5.7 | 0 | 0 | 0.01 | –0.01 | 0.9 | 1 | 276 |
| EDZ7 | 2.9 | 0.5 | 1.0 | 2.38 | 0 | 14.7 | 160.87 | 26,011 | 2.6 | 0 | 0 | 0.53 | –0.53 | 9.3 | 0 | 318,150 |
| EDZ8 | 1.3 | 0.5 | 0.5 | 1.83 | 0 | 12.0 | 473.28 | 613,670 | 0.5 | 0 | 0 | 0.01 | –0.01 | 0.4 | –25 | 10,243 |
| EDZ9 | 0.9 | 0.5 | 0.5 | 0.16 | 0 | 2.1 | 195.95 | 80,831 | 0.3 | 0 | 0 | 0.01 | –0.06 | 0.6 | –624 | 635,860 |

**Table 4.4: Mean value of PIN estimators and PIN for Eurodollar Futures market.**

| | $\alpha$ | $\delta$ | $\mu$ | $\varepsilon$ | **PIN** |
|---|---|---|---|---|---|
| Mean | 0.6444 | 0.3517 | 0.5604 | 0.5414 | 0.2501 |
| Std dev. | (0.0533) | (0.0389) | (0.0087) | (0.0216) | (0.0068) |

than supply.

Thirdly, the estimated mean value of the rate of arrival of the informed trader $\mu$ is 0.5604, which suggests that Eurodollar traders tend to believe that observed orders based on an information-driven probability of 56.04% or just above half of the total orders that trade on the Eurodollar futures market come from informed traders.

Fourthly, the mean estimate $\varepsilon$ is 0.54144, which indicates that the probability of liquidity trading is just above half. On informed days, the parameter $\varepsilon$ represents the market liquidity that comes from uninformed traders, while on uninformed days, it coincides with the total liquidity available in the market because only uninformed traders have arrived in the market.

Having estimated the parameters $\alpha, \delta, \mu, \varepsilon$, all four are used to calculate the PIN, where the probability of information-based trading is $PIN = \frac{\alpha\mu}{\alpha\mu+2\varepsilon}$, the PIN depends on the arrival rate of informed and uninformed traders and the probability of the occurrence of an information event.

The average value of the probability of informed trading or PIN on the Eurodollar Futures market for the whole Eurodollar futures series contracts was also calculated using the average value of $\alpha, \delta, \mu, \varepsilon$ is reported in Table 4.4 as 0.25001.

This suggests that investors who trade in the Eurodollar futures market face a 25.01% probability of trading with a counter-party who is better informed about the direction of the key rate.

The average value of parameters and the PIN for all 40 Eurodollar futures contracts was presented in Table 4.5. The results shown in Table 4.5 illustrate separate descriptive statistics of the PIN value and PIN parameters on 40 Eurodollar futures series contracts from January 1, 1996 to January 1, 2014. These results are consistent with those of Easley et al. [1996], in that the parameters and the PIN measure are between 0 and 1. The interesting result in Table 4.5 is that the highest PIN value appears on EDH5 and EDM6 contracts, which have the same value of 0.7, and EDH2 has the lowest PIN value of 0.677 (see Table 4.5).

#### 4.4.1.2 PIN on the Eurodollar futures market

The next step is to analyse the way in which some events, such as turmoil in the financial market or maturity effects, which have taken place in the Eurodollar futures market may have affected the trading strategies, and also the probability of informed trading. In order to particularly assess the evolution of informed trading over time, the model parameters were estimated on samples using 78 overlapping days. Then, the historical PIN was plotted separately into 10 sub-plots to illustrate the pattern of the PIN value. Each sub-plot contained four different ED contracts with an expiry date within the same years; for instance, the first sub-plot included EDH0, EDM0, EDU0 and EDZ0. These four ED contracts were due to expire in 2000 and 2010. The historical PIN is presented in Figure 4.1. Surprisingly, a similar trend of PIN appears in all 10 sub-plots. Furthermore, some landmarks of the historical PIN are indicated to assess whether

**Table 4.5: Maximum likelihood and PIN estimation.**

| | α | δ | μ | ε | PIN | | α | δ | μ | ε | PIN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| EDH0 | 0.585 | 0.407 | 0.551 | 0.540 | 0.679 | EDM0 | 0.591 | 0.401 | 0.553 | 0.679 | 0.679 |
| | (0.358) | (0.317) | (0.234) | (0.253) | (0.237) | | (0.359) | (0.318) | (0.237) | (0.257) | (0.242) |
| EDH1 | 0.601 | 0.396 | 0.549 | 0.542 | 0.679 | EDM1 | 0.607 | 0.397 | 0.547 | 0.537 | 0.681 |
| | (0.359) | (0.321) | (0.238) | (0.260) | (0.243) | | (0.353) | (0.321) | (0.235) | (0.262) | (0.241) |
| EDH2 | 0.599 | 0.395 | 0.546 | 0.540 | 0.677 | EDM2 | 0.648 | 0.361 | 0.550 | 0.529 | 0.687 |
| | (0.360) | (0.321) | (0.239) | (0.262) | (0.246) | | (0.344) | (0.319) | (0.233) | (0.269) | (0.239) |
| EDH3 | 0.665 | 0.339 | 0.556 | 0.520 | 0.694 | EDM3 | 0.671 | 0.337 | 0.559 | 0.526 | 0.693 |
| | (0.340) | (0.315) | (0.228) | (0.273) | (0.235) | | (0.338) | (0.316) | (0.227) | (0.272) | (0.232) |
| EDH4 | 0.685 | 0.319 | 0.564 | 0.521 | 0.696 | EDM4 | 0.697 | 0.315 | 0.562 | 0.520 | 0.695 |
| | (0.325) | (0.307) | (0.227) | (0.271) | (0.229) | | (0.316) | (0.305) | (0.225) | (0.271) | (0.228) |
| EDH5 | 0.716 | 0.296 | 0.572 | 0.533 | 0.700 | EDM5 | 0.714 | 0.297 | 0.569 | 0.541 | 0.694 |
| | (0.307) | (0.290) | (0.225) | (0.273) | (0.224) | | (0.314) | (0.203) | (0.229) | (0.274) | (0.229) |
| EDH6 | 0.713 | 0.291 | 0.575 | 0.545 | 0.695 | EDM6 | 0.705 | 0.301 | 0.578 | 0.539 | 0.700 |
| | (0.315) | (0.290) | (0.227) | (0.275) | (0.224) | | (0.319) | (0.300) | (0.229) | (0.279) | (0.228) |
| EDH7 | 0.665 | 0.334 | 0.572 | 0.538 | 0.694 | EDM7 | 0.654 | 0.336 | 0.570 | 0.544 | 0.692 |
| | (0.338) | (0.313) | (0.229) | (0.276) | (0.226) | | (0.347) | (0.313) | (0.227) | (0.274) | (0.226) |
| EDH8 | 0.628 | 0.365 | 0.559 | 0.550 | 0.680 | EDM8 | 0.628 | 0.366 | 0.561 | 0.547 | 0.682 |
| | (0.354) | (0.318) | (0.233) | (0.264) | (0.232) | | (0.354) | (0.318) | (0.232) | (0.261) | (0.23) |
| EDH9 | 0.599 | 0.390 | 0.556 | 0.541 | 0.683 | EDM9 | 0.604 | 0.390 | 0.557 | 0.546 | 0.682 |
| | (0.354) | (0.314) | (0.226) | (0.250) | (0.227) | | (0.354) | (0.317) | (0.228) | (0.252) | (0.229) |

| | α | δ | μ | ε | PIN | | α | δ | μ | ε | PIN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| EDU0 | 0.595 | 0.404 | 0.558 | 0.554 | 0.683 | EDZ0 | 0.598 | 0.401 | 0.550 | 0.547 | 0.678 |
| | (0.3653) | (0.3246) | (0.2427) | (0.2603) | (0.2452) | | (0.3643) | (0.3234) | (0.2404) | (0.2613) | (0.2453) |
| EDU1 | 0.605 | 0.393 | 0.552 | 0.536 | 0.686 | EDZ1 | 0.623 | 0.387 | 0.548 | 0.533 | 0.684 |
| | (0.353) | (0.3162) | (0.2358) | (0.2606) | (0.2402) | | (0.3457) | (0.3173) | (0.2362) | (0.2651) | (0.2424) |
| EDU2 | 0.653 | 0.357 | 0.552 | 0.527 | 0.687 | EDZ2 | 0.659 | 0.348 | 0.557 | 0.523 | 0.694 |
| | (0.3438) | (0.3186) | (0.2368) | (0.2724) | (0.2442) | | (0.3435) | (0.3181) | (0.2275) | (0.2717) | (0.233) |
| EDU3 | 0.673 | 0.332 | 0.557 | 0.524 | 0.691 | EDZ3 | 0.673 | 0.333 | 0.559 | 0.523 | 0.693 |
| | (0.3336) | (0.3129) | (0.2268) | (0.2674) | (0.2321) | | (0.3332) | (0.3126) | (0.2279) | (0.2682) | (0.2308) |
| EDU4 | 0.702 | 0.309 | 0.564 | 0.527 | 0.696 | EDZ4 | 0.712 | 0.300 | 0.566 | 0.534 | 0.694 |
| | (0.3155) | (0.3059) | (0.225) | (0.2729) | (0.2281) | | (0.3101) | (0.3029) | (0.2263) | (0.2727) | (0.2276) |
| EDU5 | 0.718 | 0.295 | 0.570 | 0.548 | 0.690 | EDZ5 | 0.722 | 0.292 | 0.576 | 0.546 | 0.697 |
| | (0.3127) | (0.3059) | (0.2305) | (0.2725) | (0.2298) | | (0.3106) | (0.3043) | (0.2287) | (0.2761) | (0.2251) |
| EDU6 | 0.692 | 0.307 | 0.574 | 0.540 | 0.696 | EDZ6 | 0.681 | 0.318 | 0.568 | 0.541 | 0.690 |
| | (0.3289) | (0.3071) | (0.2282) | (0.2769) | (0.2255) | | (0.3295) | (0.3059) | (0.2366) | (0.2766) | (0.2317) |
| EDU7 | 0.452 | 0.348 | 0.569 | 0.543 | 0.691 | EDZ7 | 0.639 | 0.353 | 0.566 | 0.548 | 0.686 |
| | (0.3521) | (0.3174 | (0.2307) | (0.2734) | (0.2304) | | (0.3494) | (0.3151) | (0.2303) | (0.2671) | (0.2299) |
| EDU8 | 0.619 | 0.371 | 0.563 | 0.550 | 0.685 | EDZ8 | 0.615 | 0.380 | 0.557 | 0.548 | 0.681 |
| | (0.3541) | (0.315) | (0.2277) | (0.2582) | (0.2245) | | (0.3482) | (0.3117) | (0.2248) | (0.2509) | (0.2232) |
| EDU9 | 0.595 | 0.399 | 0.552 | 0.542 | 0.679 | EDZ9 | 0.581 | 0.410 | 0.552 | 0.542 | 0.681 |
| | (0.3539) | (0.3146) | (0.2303) | (0.2497) | (0.2323) | | (0.3609) | (0.3167) | (0.2271) | (0.252) | (0.2308) |

Note: The table presents the mean and standard deviations of parameters of the PIN model, and the mean value of PIN for 40 Eurodollar futures series contracts. The parameter $\alpha$ represents the probability of an information event, $\delta$ is the probability of a low signal, $\mu$ is the rate of informed trade arrival, and $\varepsilon$ is the rate of uninformed buy and sell trade arrivals. PIN is the probability of informed trading.

some turning points in the trend of the PIN can be associated with major events that took place in the Eurodollar market, as shown in Figure 4.1.

Overall, the PIN estimates for the majority of Eurodollar futures contracts were high and widely fluctuated between 1996 and 2000; however a decreasing trend can be observed from 2002 until around 2005 for the majority of ED contracts and this was reversed to an upward trend between 2005 and 2006. Interestingly, this increasing trend after 2006 is before the sub-prime crisis. The PIN remained steady at around 0.6 from 2010 but was still below the historical PIN from the first period (1996–2000) (see Figure 4.1).

It can be seen from Figure 4.2 that four particular events in the four periods appear to have exerted some influence on the historical development of the PIN. The first period that influenced the PIN was between 1996 and 2000, when the PIN widely and highly fluctuated. This can possibly be related to the fact that, in this period, the Eurodollar Futures market was not popular or traded less frequently than today. There were very few quotes and trade volumes in that period of time – for example, in the EDM series (Figure 4.3). As a result, the Eurodollar market was uncomplicated and it was easy for informed traders to manipulate it by modifying their trading behaviour to reflect a different degree of PIN. After this period, a decreasing trend of the PIN could be observed after 2002 and one possible explanation for this is that the number of quotes increased rapidly, from less than 100 to more than 10,000 per day, and around 50 trades per day increased to more than 1,000. Therefore, these results of a decreasing PIN could be affected by more frequent trading in futures contracts.

The third period of historical events that affected the value of the PIN was

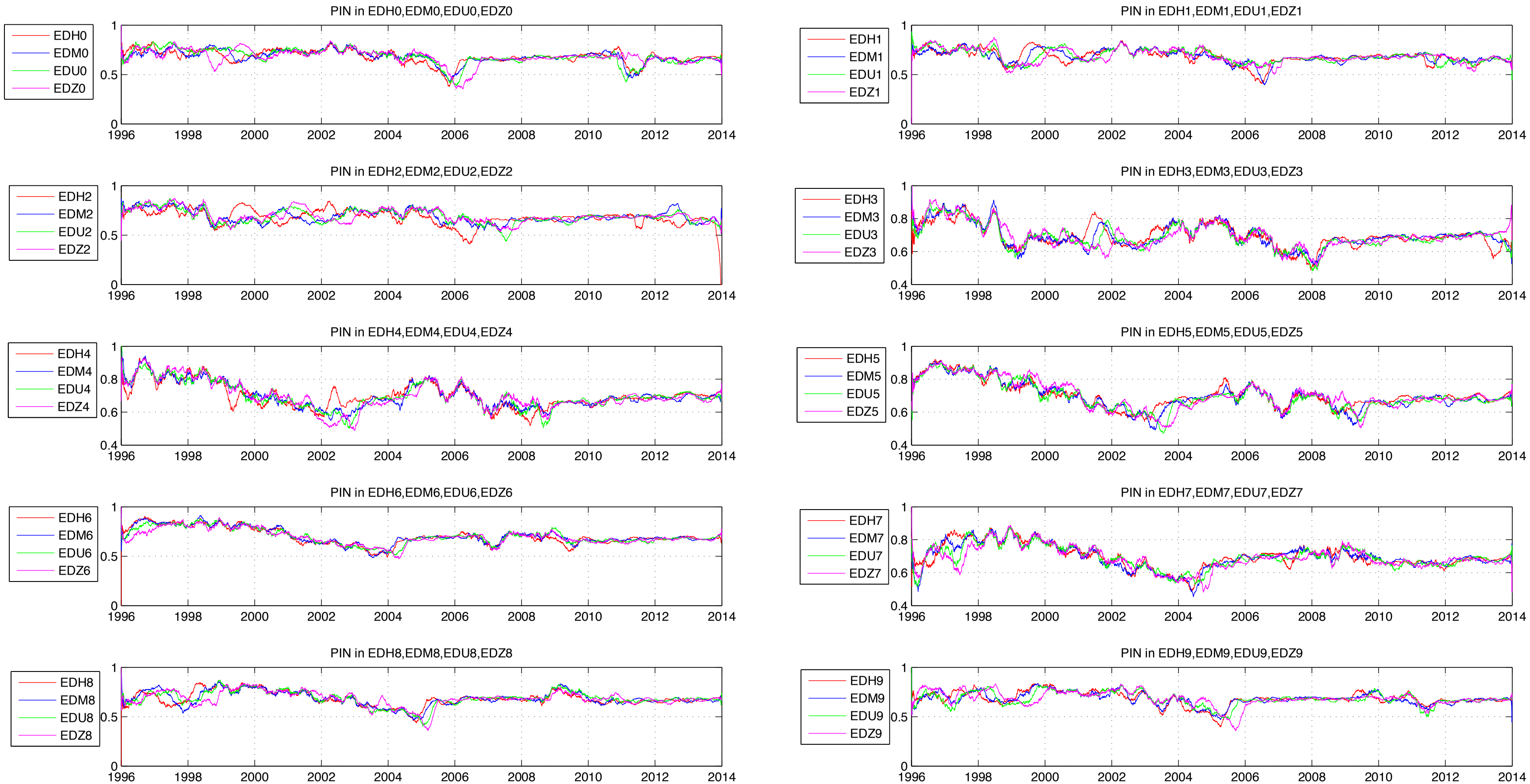



**Figure 4.1: Historical evolution of the PIN for ED?0, ED?1,...,ED?9.**
Note: The historical PIN is separated into 10 groups, each group including four ED contracts that have an expiry date within the same years.

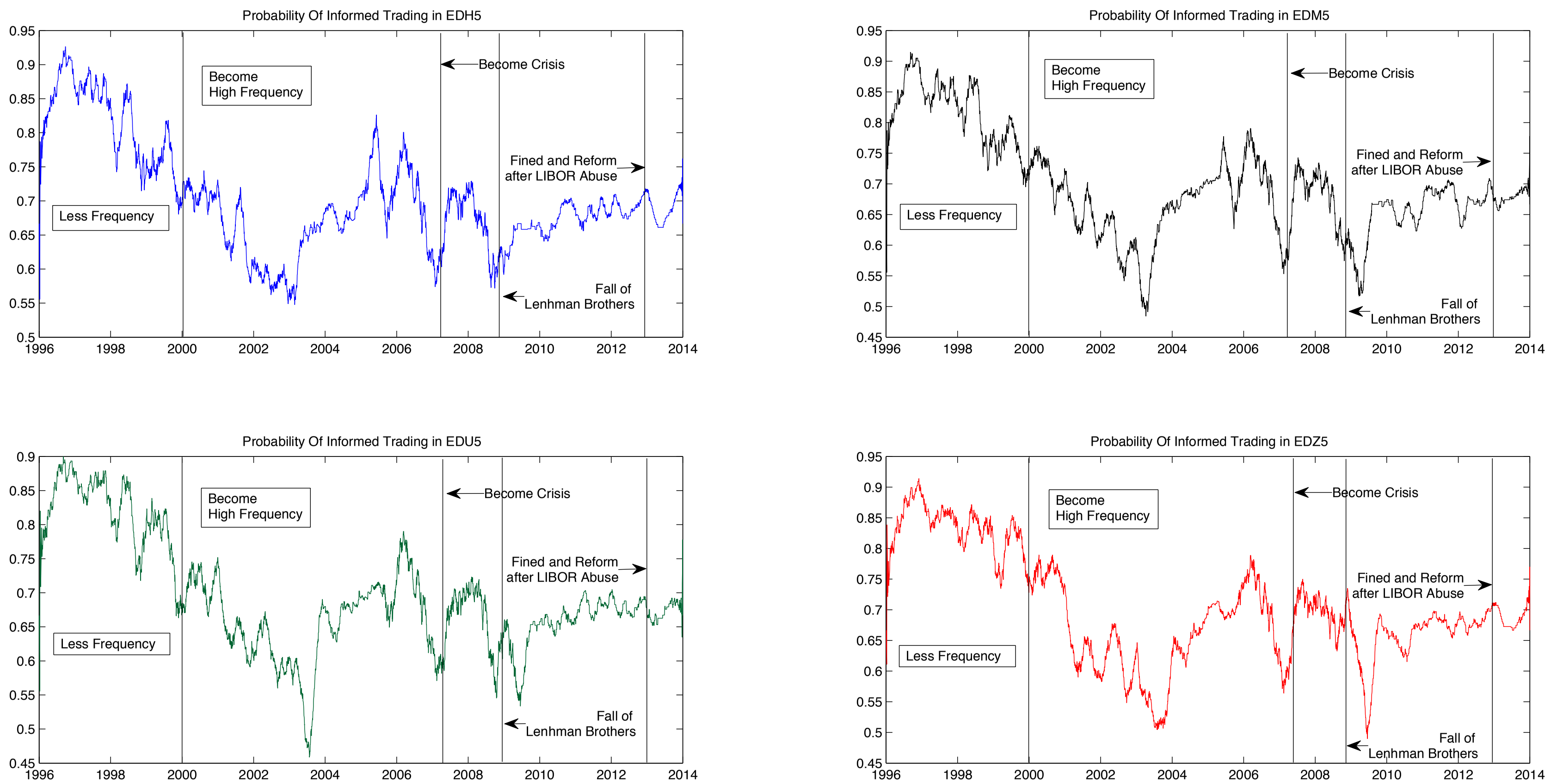



**Figure 4.2: Historical PIN on EDH5 EDM5 EDU5 EDZ5.**

Note: This figure shows the historical value of PIN from 1996 to 2014 for four ED contracts (EDH5, EDM5, EDU5, EDZ5) with a timeline period that affects the change of PIN value. First, the low-frequency trading period. Second, the ED becomes high-frequency trading after the Globex electronic trading platform was introduced. Third, during the financial crisis. Finally, after regulators fined the banks that were involved in LIBOR manipulation.

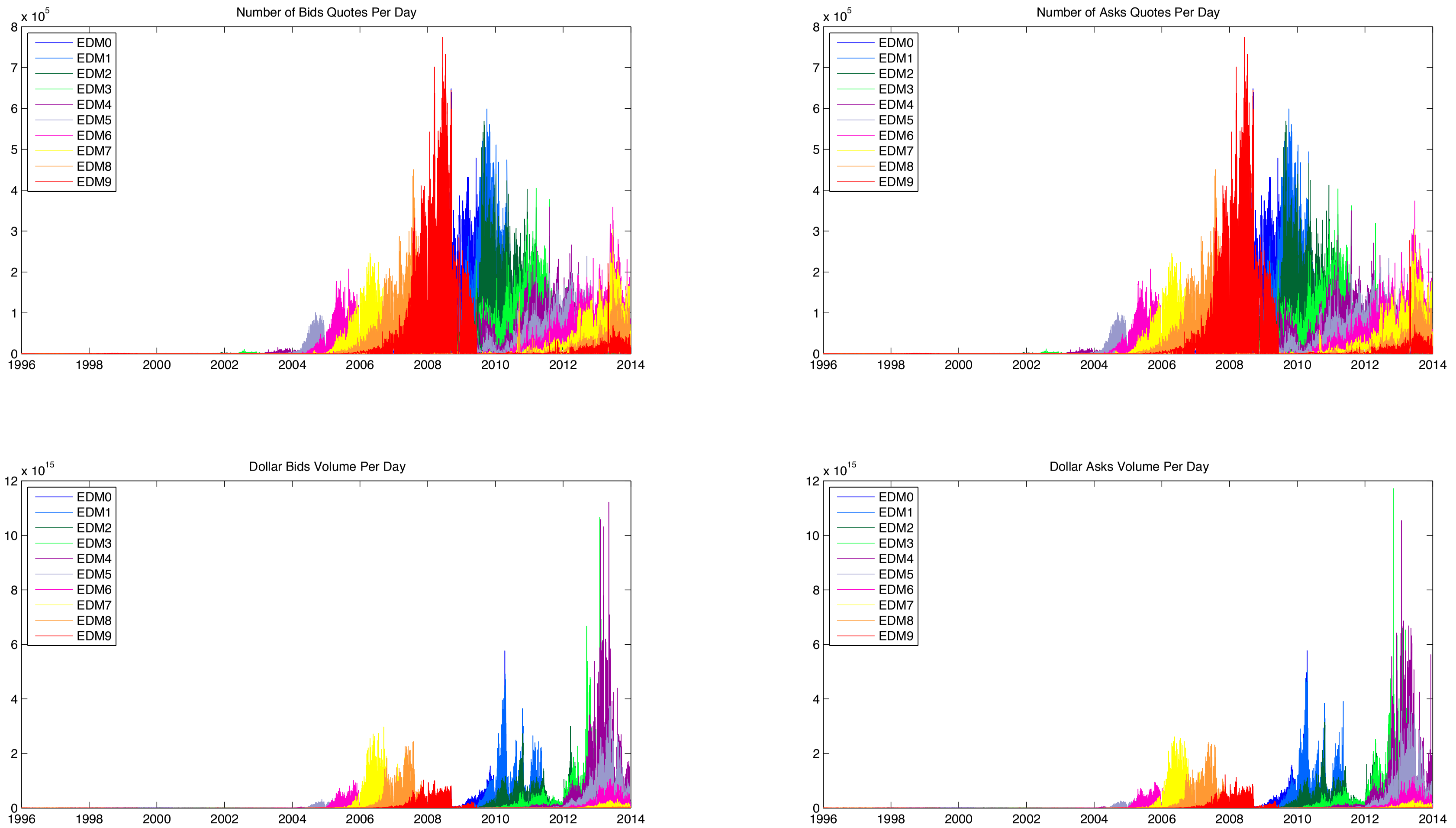



**Figure 4.3: Example of number of quotes and volume of quotes per day for 10 EDM contracts.**
Note: The number of quotes and trade were very low or less frequent from the opening in 1996 until 2000 with a significant increase after 2000, which reached a peak between 2008 and 2010, then a decreasing trend started from 2010. This trend occurred for all 40 Eurodollar Futures.

the market turmoil. The sub-prime crisis mainly affected the money market and used the market to reduce the liquidity in the financial market. The PIN sharply decreased at the beginning of the financial turmoil, approximately in mid-2007, after which the PIN changed rapidly because of some spiking in most Eurodollar contracts. This was possibly affected by the reduced liquidity at the beginning of the crisis, and subsequently the panel bank tried to maintain the key interest rate in the LIBOR submission to protect the reputation of the banks, as explained in the next section.

The final event that affected the historical PIN occurred in 2012. This was when the CFTA and the FSA ordered Barclays Bank, UBS, RBS and Rabobank to pay a penalty for manipulating the LIBOR. The effect of the announcement was visible in the marked decrease of the PIN in 2012 and it became more stabilized in most ED futures contracts after 2013.

## 4.5 Summary Chapter

Overall, the PIN appears to have dropped around the time of the sub-prime crisis (circa 2007) and then bounced back afterwards and sharply dropped again with the collapse of Lehman Brothers; then it bounced back once again indicating that systematic market uncertainty has played a role in information asymmetry. Significantly, the result of the PIN in this thesis is always higher for the Eurodollar futures market than those recorded in the equity market (circa 0.4 to 0.8 versus 0.1 to 0.5; see Abad and Yagüe, 2012; Aslan et al., 2011; Easley et al., 1996). However other PIN studies of interest rate derivatives have found similarly high, if not higher PIN measurements, in particular Kim et al. [2014] find the PIN for the Eurodollar market to be substantially higher on average than that detected in the equity and FX markets.

In summary, in terms of illustrating the presence of informed trading in the Eurodollar futures market, the initial results are somewhat inconclusive over the time period.

# Chapter 5

# Motivating Evidence of LIBOR Manipulation and The PIN Analysis in the LIBOR Reference Rate Derivative Market

Since 2008 a series of allegations emerged in regard to systematic attempts to manipulate the London Interbank Offered Rate (LIBOR). The first controversial study released by the Wall Street Journal suggested that during the financial crisis (2007 - 2008) banks may have understated the interbank borrowing rate to keep them look healthier than the reality. In June 2012 this controversy came to the fore when Barclays was fined £290 million for having rigged the worlds benchmarking borrowing rate. Therefore, this chapter applies a classical test of Pearson Spearman's rank correlation to a series of participating bank s for a LIBOR submission. The results show a very obvious pattern, the Ranking correlation for bank's quotes is effectively identical up to mid 2007 then there is a break in correlation subsequent to the start of the sub-prime crisis. Furthermore, the document catalog of recorded communication between interest rate derivative traders and LIBOR submitters gives a strong support evidence of this market abuse. The evidence is provided in this chapter.

This chapter also implements a version of the probability of informed trading measures first introduced in Easley et al. [1996], for the Eurodollar futures market over the LIBOR manipulation period. The Eurodollar future is one of the most actively traded US Dollar LIBOR benchmarked derivatives in the world with total traded volume in excess of half a quadrillion US dollars in 2011 (Source: CFTC) and is possibly the world's largest financial market. The objective of this chapter is an ex-post review of the effectiveness of the PIN in determining changes in the information structure of the market around documented episodes of recorded manipulation of the benchmark rate, from the various publicly available regulatory reports. In keeping with previous studies on interest rate derivatives, the findings indicate that the average PIN is far higher than for the equity market at or around 2/3 to 3/4. When implementing a rolling measure of the PIN to detect time variation I find a clear pattern of increases and decreases in informed trading relative to the recorded activity in the current regulatory reports. Furthermore, this analysis also finds a strong maturity effect that appears to arise from the strong time variation in trading in these contracts over their life-cycle, from inception to maturity. This work constructs a series of experiments to determine the significance of the trading time for different epochs from 1996 to 2014 , using all inside quotes and trades in these contracts. The results indicate that the PIN could have been used as an early warning of unusual activity in the LIBOR reference rate and anecdotally this chapter demonstrates that on specific dates identified by the CFTC and FSA the recursively estimated PIN reaches a peak over 1,000 basis points higher for certain near-delivery contracts than for others further away in the forward curve.

## 5.1 Introduction

The London InterBank Offered Rate (LIBOR) and the European InterBank Offered Rate (EURIBOR) act as a reference rate for the majority of interest rate derivatives traded in global markets. However, the probity of this rate-setting mechanism has been called into question and several cases of manipulation of the LIBOR fix have been recorded since 2008. This chapter exploits new information released by the various regulatory cases and uses it to conduct a field test of a commonly used method to assess the degree of asymmetric information from the order flows of traded assets, in this case LIBOR referenced Eurodollar futures. The Eurodollar futures market is referenced to the three-month dollar LIBOR rate (henceforth 3M-LIBOR) for maturities out to ten years. The analysis utilizes a large sample of trade and quote data to estimate the probability of informed trading (PIN) using the model of Easley et al. [1996]. The results paint a mixed picture of the effectiveness of the PIN for interest rate futures. In most of the documented cases the PIN varies from the long-run averages in a statistically distinctive manner; however, other effects, such as high-frequency quoting around maturity dates, are found to generate noise that has a similar impact on the PIN.

Evidence is provided for the effectiveness of the PIN model in actually detecting informed behavior in the Eurodollar futures market. To accomplish this, I collect all trades and inside quotes for Eurodollar futures from 1996 to 2014, a data set in excess of two billion observations. I believe this to be the largest study of its type ever conducted. This work will compute the PIN using a variety of approaches and construct test groups to see whether the PIN varies systematically with the events suggested in the regulatory filings, versus a set of controls. In general, the control groups show average PIN variation around "normal" events in the futures contract life-cycle – in this case the maturity of the contract.

The reports on LIBOR manipulation for each bank provide us with a unique field experiment to test the PIN as the number of events is very large and the specificity of the objectives of those involved is very clearly stated. It is worth reviewing the rationale behind this approach. The PIN seeks to detect trades that appear to be the result of informed signals prior to some pay-off event. For equities the future dividend is often somewhat nebulous at higher frequencies, however in futures markets it is not. Each day the outstanding contracts are marked to market using the last one minute of trading or the mid-price of the final outstanding quotes. At maturity the contracts are settled to cash based on the price of 100 minus the 3M-LIBOR on the third Wednesday of the settlement month. It should be stressed that for all "manipulation" is supposed to take place in the reference rate (the 3M-LIBOR) and not directly in the Eurodollar itself.

My foundational assumption is that given that a particular bank is able to influence the reference rate and hence possibly influence the mark to market and certainly influence the final settlement price, then the traders of this bank will have more complete information on the final payoff of this asset over both the trading days and over the overall maturity of the contract. This informational advantage should then be reflected in the PIN, relative to days when no manipulation of the reference rate took place. I will illustrate using examples and summary statistics from the regulatory reports that requests from derivatives traders tended to take place with relatively short notice periods of between one day and one week. The degree of variation in the reference rate that can be created by "high or low balling" the rate is relatively small. However, in this chapter I will also illustrate that the size of the notional positions in the Eurodollar market is so large that there is still a significant incentive to shave the reference rate in your favour.

Overall, the findings are mixed. I find that the PIN reacts strongly to certain types of events and the variation is statistically significant relative to the control group. However, the PIN is always very high for Eurodollar futures market, averaging over 0.5 for the majority of the sample, compared to the equity market where observed PIN, for which IT usually varies between 0.1 and 0.6 from a survey of prior studies. This work also finds that IT is not stable relative to the various estimation assumptions. However, because this is almost entirely driven by the instability of the "trade classification" algorithm used to determine buying pressure, I leave full investigation of this issue to future work.

The contribution of this chapter is threefold. First, it provides a comprehensive statistical analysis of market abuse related to LIBOR manipulation. Second, to implement my analysis I provide, fully catalogued, the manipulation dates of the LIBOR manipulation especially the 3-month LIBOR, from the legal documents proceedings against multinational bank from the U.S. and UK. regulators. Finally, this work provides some evidence of the effectiveness of the PIN as an early warning system for both market manipulation and market uncertainty.

The chapter is organized as follows §(5.2) provides a detailed background to interest rate derivatives including LIBOR and Eurodollar Futures trading on

CME's Globex. §(5.3) explores the PIN and presents the empirical results of PIN on the LIBOR manipulation case. §(5.4) explores the PIN and the maturity effect. Finally, §(5.5) provides some conclusions and outlines directions for future work in this area.

## 5.2 Background of Interest Rate Derivatives Market and LIBOR Manipulation

The purpose of this research is to investigate the comparative microstructure of the Eurodollar Future markets during the LIBOR manipulation event. As a useful natural experiment, the incident of LIBOR manipulation offers a great opportunity for academics to impute and observe parameters of interest in an information asymmetry model and to study the development of those estimators incorporated with the information content.

To understand the Eurodollar futures based on 3-M US Dollar LIBOR, it is deemed useful to introduce the background of the London InterBank Offered Rate or the LIBOR, since this is the benchmark or key interest rate for Eurodollar futures. The original LIBOR was introduced in 1984, when the BBA or British Bankers' Association and other parties, such as the Bank of England, formed a standard settlement rate of contractual terms on interest rate swaps. Two years later, the BBA published the first LIBOR as an official interest rate for a variety of financial instruments transacted across the multi-international financial market, notably syndicated loans, forward rate agreements, interest rate futures, interest rate swaps, and interest rate options [Abrantes-Metz et al., 2012]. The LIBOR submission in now administered by the NYSE: ICE LIBOR unit. At 11.00 am on each weekday, the panel banks submit their rates to an administrator. Then, the highest four and the lowest four of the quoted rates are removed and the remaining middle rates are used to calculate a simple arithmetic mean. Finally, at 11.30am LIBOR is quoted and announced by Thomson Reuters. Since September 2013 the ICE LIBOR unit makes the fixing available to various data vendors (including Thomson Reuters) for distribution to the market and for settlement on maturing interest rate derivatives set relative to this rate. Currently, the daily LIBOR rates are quoted on 10 major currencies, including the Australian Dollar, British Pound, Canadian Dollar, European Euro, Danish Kroner, Japanese Yen, New Zealand Dollar, Swedish Kroner, Swiss Franc and US Dollar.

In London, the LIBOR rate is based on the rates quoted by 16 banks selected by the BBA to provide a daily rate for LIBOR calculation. Currently in the United States, 18 banks are selected by the BBA to form a panel to construct the US Dollar LIBOR. Three US and 16 non-American banks are included in the panel for the US dollar fixing. The US banks are: Bank of America, JP Morgan Chase and Citibank. The non-US Banks are: Bank of Tokyo-Mitsubishi UFJ, Barclays Bank, BNP Paribas, Credit Agricole CIB, Credit Suisse, Deutsche Bank AG, HSBC, Lloyds TSB Bank, Rabobank, Royal Bank of Canada, Société Générale, Sumitomo Mitsui Banking Corporation, The Norinchukin Bank, The Royal Bank of Scotland Group and UBS AG.

As a benchmark of interest rates, the LIBOR is widely used for various financial instruments including standard interbank products such as forward rate agreement, interest rate swaps, interest rate futures, mortgage rates, standard loan rates and Eurodollar Futures trading. Likewise, in the United States around 60 to 80% of prime adjustable rate mortgages and sub-prime mortgages were indexed to the US LIBOR. In addition, the American municipalities borrowed around 75% of their money through products that were linked to the LIBOR.

### 5.2.1 Eurodollar futures and the current LIBOR manipulation cases

This section starts with the definition of Eurodollar, the Eurodollar is US Dollar deposits located outside the legal jurisdiction of the United States. The majority of Eurodollar deposits are located in London, however the Eurodollar deposit account really has very little to do with the futures contract named after it. The Eurodollar future contract made its debut in December 1981 on the Chicago Mercantile Exchange (CME), and this was the first future contract ever settled to cash. Since the transaction market for the Eurodollar futures contract is not so well understood outside the interest rate derivative community, I will explain the specifics in some detail in this section.

The Eurodollar futures themselves are interest rate derivatives with reference to the 3M London interbank offered rate (LIBOR), that have a face value of $1 million with a 3-month maturity and the futures contracts have a maturity of up to ten years. The regular contracts expire either in March, June, September or December, extending outwardly for 40 quarterly expiring contracts in the long term; shorter maturity monthly contracts are traded, however, their relative volume, trading activity and quoting activity are several orders of magnitude lower than the quarterly contracts. The last trading day is the second business day prior to the third Wednesday of the delivery month in both New York and London. At any given time, the exchange lists four of the monthly series, bringing the total number of available Eurodollar future contract maturity types to 44 –

namely, 40 quarterly and 4 monthly serial contracts.

Individually, Eurodollar futures, in effect, parallel forward rate agreements (FRAs), but provide more liquidity by virtue of the clearing house and being marked to market. Of key importance is the cash settlement – Eurodollar traders will receive cash compensation on the same day rather than having to wait until the expiry date. Moreover, large institutions tend to use Eurodollar futures to manage their risk or portfolio to avoid interest rate risk. To hedge against this risk, borrowers will enter into short future contracts to avoid the risk of rising interest rates while lenders will enter into long future contracts to avoid the risk of falling interest rates. However, there are speculators who purely seek to make a profit from Eurodollar trading by betting on the change in interest rates (future directional). Institutional investors will use Eurodollar futures contracts to lock in an interest rate today for their borrowing or lending in the future. Eurodollar futures are the most widely traded short-term interest rate futures in the world, with average daily volume of more than 2 million. Futures contracts are the most active interest rate futures traded in the world, typically in the 7 to 9 million range in the shortest maturity futures.[1]

At this point, after the Eurodollar futures (ED) is introduced, the following section will provide a connection between ED and LIBOR manipulation. I start with a point of the main motivation and a background of my experiment for this thesis, which lies in the run-up to the 2007 financial turmoil and anomalies in the LIBOR rate in the immediate aftermath. The British Bankers' Association's claims that the "BBA LIBOR is the primary benchmark for global's short-term interest rates(Abrantes-Metz et al. [2012]). However, there is a case that the

[1] CME Group, volume data, http://investor.cmegroup.com/investor-relations/downloads.cfm [accessed August 19, 2014].

financial market has once again impaired trust when the media reported that several large international banks were reporting unjustifiably low LIBOR rates. For instance, an analysis conducted by the Wall Street Journal (WSJ) on May 29, 2008 suggested that banks may have reported flawed interest data for 3M-LIBOR[1].

At this juncture, it is useful to provide some background on the events surrounding the post 2008 LIBOR manipulation regulatory actions and how this links to the Eurodollar market both directly and indirectly. In the immediate aftermath of the October 2008 chapter 11 bankruptcy of Lehman Brothers, interbank rates began to exhibit high levels of volatility. Figure 5.1 illustrates the three month US dollar LIBOR submission (henceforth 3M-LIBOR) for submitter banks from August 2008 to January 2009. It can be seen from this figure that prior to the first announcement of Lehman Brothers impending collapse the spread between the highest submitted rate and the lowest rate is less than 10 basis points. As I move towards the declaration of Chapter 11 the spread increases to over one percent or 100 basis points. Subsequent to this period several articles in industry periodicals and newspapers indicated that banks were systematically understating their borrowing costs.

The rationale for this story was that since the interbank market had been effectively frozen, many LIBOR maturities must have been an estimate in order to fill in the blanks. Subsequently, an investigation into the process of fixing the LIBOR rates uncovered systematic manipulation of the rates pre-2007, reporting that the motivation for this activity was to assist the parts of the banks that

[1]The WSJ analysis suggests banks may have reported flawed interest data for LIBOR. (C., Mollenkamp and M., Whitehouse, Study Casts Doubt on Key Rate, http://online.wsj.com/news/articles/SB121200703762027135)[Last accessed August 19, 2015]

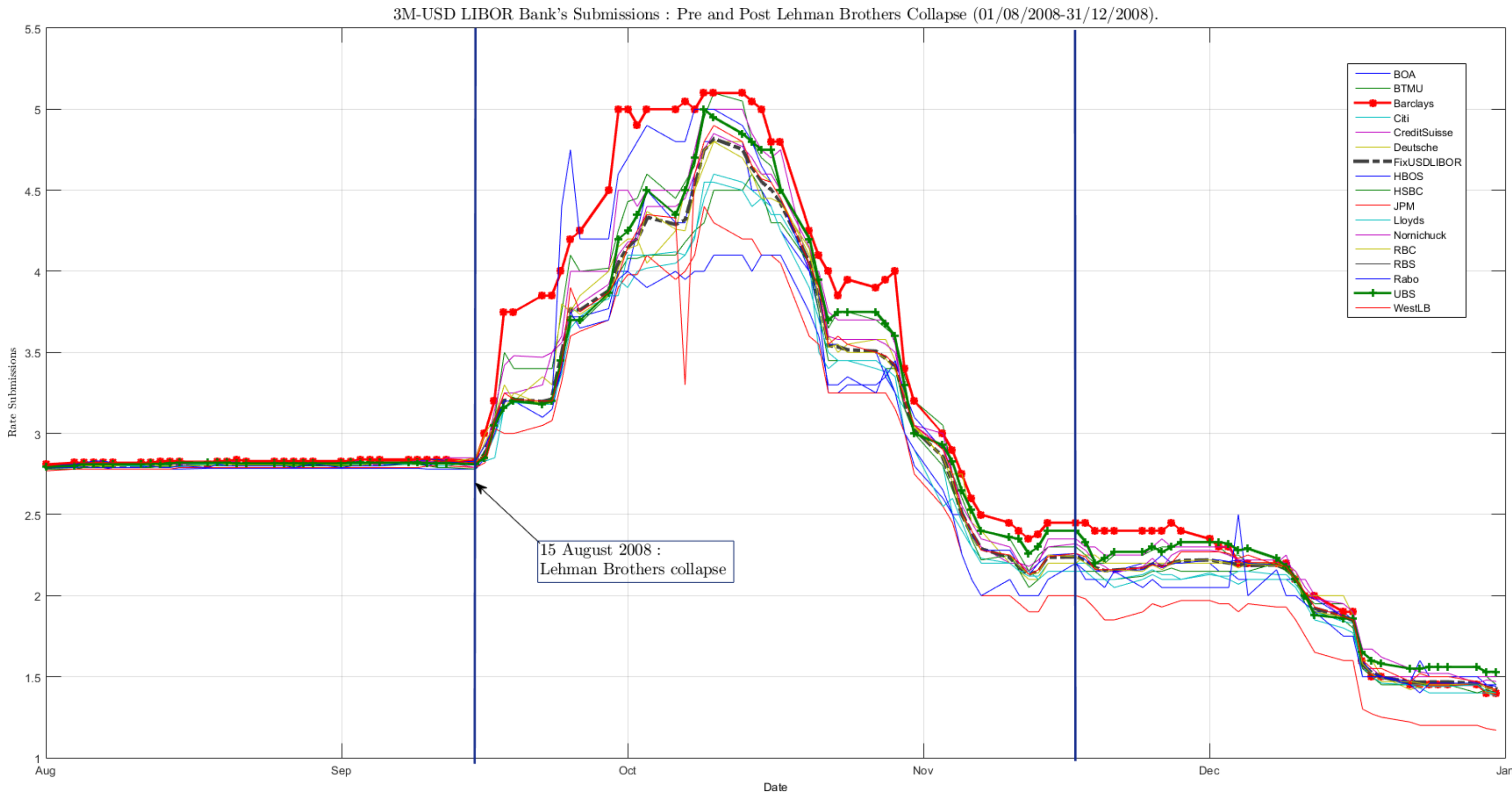



**Figure 5.1: Barclays' 3M US Dollar LIBOR submission around Lehman Brothers collapse.**

Note: This figure presents Barclays' and the other 15 LIBOR panel banks 3M US Dollar LIBOR submissions around the Lehman Brothers collapse on September 15, 2008. It can be seen that during the financial crisis, the banks, to maintain their reputation, submitted LIBOR rates lower than the real rate to avoid looking desperate for cash. Before the collapse, the LIBOR submission spread from the 16 panel banks was narrow and quite stable, whereas after the collapse of Lehman Brothers the submission rate from the member banks widened to reflect the real offer rate to borrow money from other banks.

traded in interest rate derivatives. Post 2008, the systematic mis-reporting of the LIBOR was simply a 'beauty contest' to provide the impression that the banks were being offered better rates (or any rate) than was actually happening in reality. Indeed, several of the reports indicated that bank regulators had pressured institutions to report lower rates as a means of reducing market stress.

Subsequent to the results of the initial sets of investigations into LIBOR manipulation, from September 2013 the responsibility for collecting the LIBOR rates for each bank in the submitter pool has changed from the British Bankers Association in conjunction with Thomson-Reuters to the intercontinental exchange LIBOR unit (ICE-LIBOR) and a new set of guidance has been issued on transaction validity for submitters. This study will focus on the period immediately prior to September 2013.

Table 5.1 illustrates the submission process for a selection of days. First, banks in the submitter pool submit their borrowing costs across maturities and currencies and their submissions are ranked top to bottom. The top and bottom quartiles are eliminated and an average taken of the remaining banks. In Figure 5.2, I plot the submitted 3M LIBOR rates alongside the p-value of the day-to-day Spearman's rank correlation between the banks submitted rates from 1996 to 2015. A very obvious pattern is evident, the rankings for the banks is effectively identical day-on-day up to mid 2007. However, the p-value exhibits substantial spikes subsequent to the start of the sub-prime crisis and then increases markedly after the Lehman Brothers Chapter 11 event mid to late 2008. Indicating substantial changes day-to-day in the ranking of the various banks (note: for banks that do not submit a rate one day, I eliminate their submitted rate and rank both days in their absence). However, my interest is less in the

**Table 5.1: Example of 3M-LIBOR submissions for six days between 2005–2008.**

| Nov 22, 2005 | | May 24, 2007 | | Sep 4, 2007 | | Jun 18, 2008 | | Oct 15, 2008 | |
|---|---|---|---|---|---|---|---|---|---|
| *BTMU* | *4.4* | Norin | *5.37* | *Barclays* | *5.21* | *BTMU* | *2.82* | *Barclays* | *5* |
| *Barclays* | *4.4* | *HBOS* | *5.365* | *Deutsche* | *5.21* | *Norin* | *2.82* | *Cr.Suisse* | *4.75* |
| *BOA* | *4.39* | *BTMU* | *5.36* | *BTMU* | *5.2* | *Barclays* | *2.81* | *RBS* | *4.75* |
| *Citi* | *4.39* | *Barclays* | *5.36* | *BOA* | *5.2* | *HSBC* | *2.81* | *BTMU* | *4.7* |
| **Cr.Suisse** | **4.39** | **BOA** | **5.36** | **Cr.Suisse** | **5.2** | **Rabo** | 2.81 | **HBOS** | **4.65** |
| **Deutsche** | **4.39** | **Citi** | **5.36** | **HBOS** | **5.2** | **WestLB** | **2.81** | **Deutsche** | **4.6** |
| **HBOS** | **4.39** | **Cr.Suisse** | **5.36** | **JPM** | **5.2** | **BOA** | **2.8** | **Norin** | **4.6** |
| **JPM** | **4.39** | **Deutsche** | **5.36** | **Lloyds** | **5.2** | **Citi** | **2.8** | **WestLB** | **4.57** |
| **Lloyds** | **4.39** | **HSBC** | **5.36** | **Norin** | **5.2** | **Cr.Suisse** | **2.8** | **UBS** | **4.53** |
| **Norin** | **4.39** | **JPM** | **5.36** | **Rabo** | **5.2** | **Deutsche** | **2.8** | **BOA** | **4.5** |
| **Rabo** | **4.39** | **Lloyds** | **5.36** | **RBC** | **5.2** | **HBOS** | **2.8** | **HSBC** | **4.5** |
| **UBS** | **4.39** | **Rabo** | **5.36** | **RBS** | **5.2** | **Lloyds** | **2.8** | **Citi** | **4.45** |
| *WestLB* | *4.39* | *RBC* | *5.36* | *UBS* | *5.2* | *RBC* | *2.8* | *Lloyds* | *4.45* |
| *HSBC* | *4.38* | *UBS* | *5.36* | *WestLB* | *5.2* | *UBS* | *2.8* | *RBC* | *4.45* |
| *RBCc* | *4.38* | *WestLB* | *5.36* | *Citi* | *5.14* | *RBS* | *2.795* | *JPM* | *4.1* |
| *RBS* | *4.38* | *RBS* | *5.35* | *HSBC* | *5.14* | *JPM* | *2.78* | *Rabo* | *4.1* |
| **3M-LIBOR** | **4.39** | | **5.36** | | **5.2** | | **2.8025** | | **4.55** |

Note: The rates in bold are those that present the average for the submission. The first three days, November 22, 2005, May 24, 2007 and September 4, 2007 all correspond to fixes that have been investigated and deemed suspicious. September 4, 2007 is a fix that occurs just after the sub-prime crisis, whilst the last two columns are samples of submissions before and after the Lehman Brothers collapse.

period of volatile submissions, but in the preceding period when the ranking was highly stable. There is clear evidence of attempted cartel behaviour in submitting 3M-LIBOR. In Table 5.2 I have summarized the current count for individual requests to submitters to adjust the various LIBOR fixes. I can see that the two most popular were the 3M-LIBOR (303 requests) and the Yen LIBOR (476 requests) across all maturities.

Table 5.3 provides further depth and a timeline of the reported requests for each of the banks that have settled (as of August 2015) with the US Commodities and Futures Trading Commission (CFTC) and the UK Financial Services

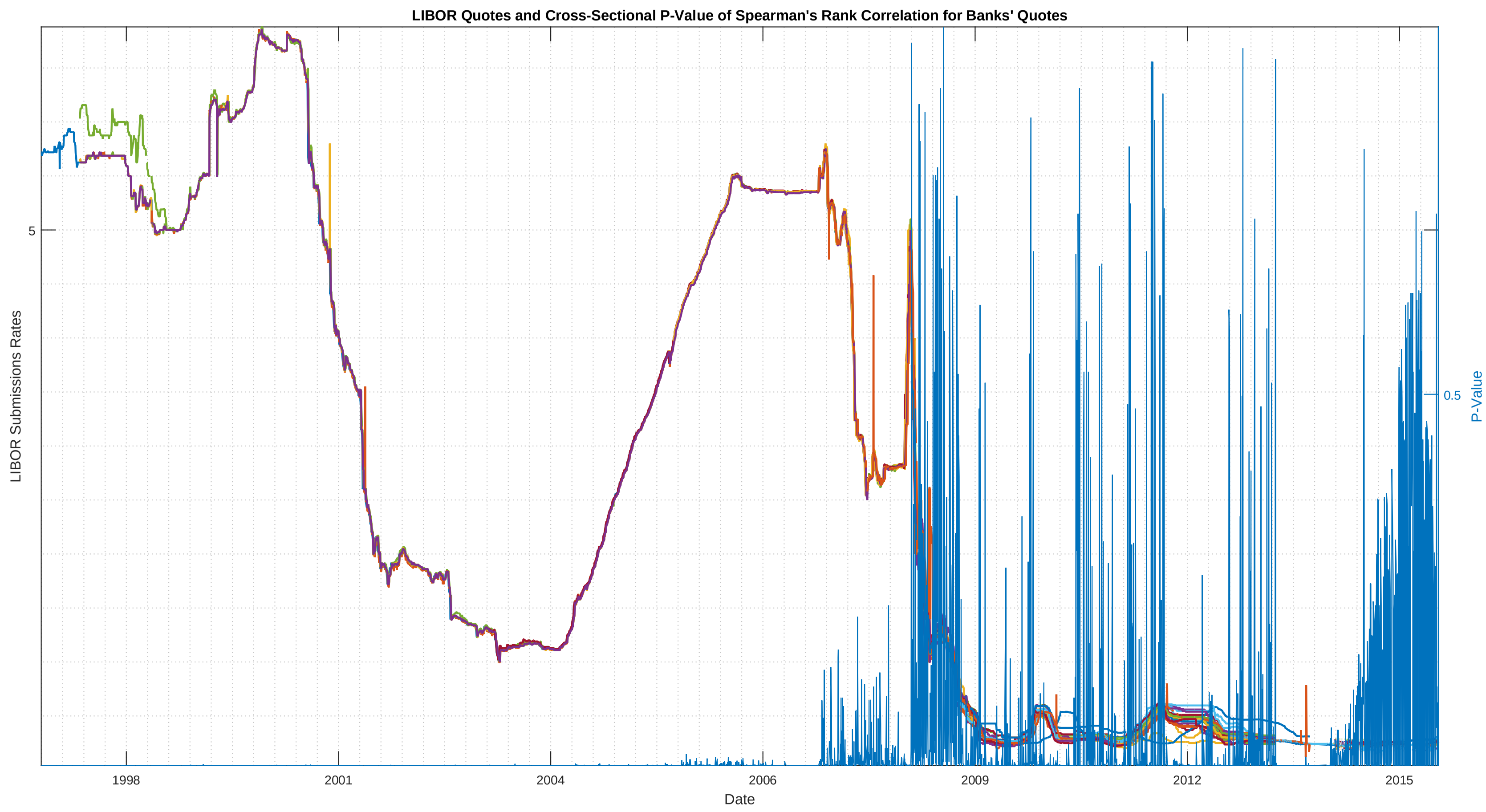



**Figure 5.2: LIBOR Quotes and Cross-Sectional Correlation for Bank‘s Quotes**

Note: This figure shows individual banks’ quotes on USD 3M LIBOR(left axis) and p-value of cross-sectional correlation from the panels banks(right axis). It can be seen in this figure that the individual bank quotes are very tightly clustered from 1996 until before the collapse of Lehman Brothers. In this period the ranking correlation of banks’ quotes are notably high as the p-vale of Spearman’s rank correlation test are near to zero which means the ranking of the 16 banks’ quotes mostly stayed at the same level. However, after 2008 there are significant changes of ranking correlation of banks’ quotes which demonstrates that they did not remain at the same level as during the previous period.

**Table 5.2: Number of communication requested on LIBOR manipulation.**

| LIBOR Currencies | Barclays | Lloyds | RBS | UBS | Rabobank | Total |
|---|---|---|---|---|---|---|
| 1M-USD LIBOR | 111 | 1 | 7 | 22 | 19 | 160 |
| **3M-USD LIBOR** | **127** | **36** | **34** | **57** | **49** | **303** |
| Sterling LIBOR | 22 | 60 | 21 | 36 | 63 | 202 |
| Euribor | 5 | 1 | 46 | 54 | 105 | 211 |
| YEN LIBOR | 160 | 53 | 3 | 81 | 179 | 476 |
| EuroYen | - | - | - | 86 | - | 86 |
| **Total** | **425** | **151** | **111** | **336** | **415** | **1,438** |

Note: This table presents the number of requests by derivatives traders to their LIBOR fixing desks for the 3M-USD LIBOR and other LIBOR or similar rates. The comparison of the number of requests between 3M-USD LIBOR and other LIBOR currencies mentioned in CFTC documents in the matter of LIBOR manipulation for Barclays, RBS, UBS, Rabobank and Lloyds Bank.

Authority (FSA, now called the Financial Conduct Authority) for the 3M LIBOR. The table also provides a summary of the relevant communications. In several cases, see for instance Barclays, the requests came directly from the interest rate derivatives desks, providing prima facie evidence of private information on the reference 3M-LIBOR.

There is evidence from the summaries for several of the banks in question, see CFTC, 2012 for an example, that the primary object of interest for the derivatives traders requesting adjustments to the 3M-LIBOR and others were interest rate swaps (IRS) and Eurodollar futures. For further analysis, I have chosen the Eurodollar futures market for two reasons. First, data availability. The CME provides tapes for every inside quote and trade in Quarterly and nearest month Eurodollar futures from 1996 to 2015. From 2008 the CME provides the complete limit order book to ten levels (in practice only five are populated) and this is in effect, a population dataset of activity in this market. The IRS for my sample was either completely or nearly completely an over-the-counter market, so the ability to collect a complete picture of market activity for these transactions is virtually impossible (indeed it is very difficult for those with actual regulatory oversight). Second, and arguably more importantly, the in-house traders for each of the submitter banks were generally assigned to positions needed to offset positions the bank has taken on when facilitating its IRS activity. The most common method for offsetting fixed or floating positions in IRSs is to use Eurodollar strips. CME groups own analysis of the Eurodollar futures market by Sturm and Barker [2011], illustrates the concomitant rise in Eurodollar futures activity with the total volume of IRS activity. Hence, the provision of private information on futures fixings of the 3M-LIBOR would be available to certain traders within the Eurodollar futures market.

As discussed previously, Eurodollar futures as a basic construct, serves as a primitive component in a great number of other US dollar denoted interest rate derivatives (and indeed for many cross currency derivatives with the US dollar). Eurodollar strips and term spreads can be used to synthesize a variety of points on the dollar forward curve. Furthermore, since the advent of the Dodd-Frank act in 2010 the trade in dollar interest rate swaps has now increased margin requirements and the cost of the synthetic provision through the Eurodollar market has led to direct structuring rather than simple offsetting by originating banks, see Labuszewski [2011].

To illustrate the trading connection between the 3M-LIBOR and the Eurodollar it is useful to describe the crucial events in a days trading. Eurodollar futures are traded around the clock on Central Standard Time (CST; Chicago time) from Sunday to Friday, 23 hours a day. On a typical trading day, the CME is divided into two stages: the per-opening trading stage and the bilateral stage. The Eurodollar futures are exchanged exclusively in the open outcry on regular CME floor trading between 7:20 am to 2:00 pm. In addition, trading on Globex solely begins at 5:00 pm CST on Sunday, which represents the beginning of Monday's trading, and continues to operate overnight until 4:00 pm CST on the following day. However, Globex closes for regularly scheduled maintenance between 4:00 pm and 5.00 pm CST every day. For instance, 5:00 pm on Sunday represents the beginning of the trading session for Monday's trading day and it continues until 4:00 pm on Friday, when Globex closes for the weekend and re-opens again at 5:00 pm on Sunday to begin the next week's trading.

At 11.00 am UK time (05:00) on each weekday, the panel banks submit their rates to an administrator. Then, the highest four and the lowest four of the quoted rates are removed and the remaining middle rates are used to calculate a simple arithmetic mean. Finally, at 11.30am UK time (05:30 CST) LIBOR is quoted and announced by Thomson Reuters. Since September 2013 the ICE LIBOR unit makes the fixing available to various data vendors (including Thomson Reuters) for distribution to the market and for settlement on maturing interest rate derivatives set relative to this rate. Currently, the daily LIBOR rates are quoted on 10 major currencies, including the Australian Dollar, British Pound, Canadian Dollar, European Euro, Danish Kroner, Japanese Yen, New Zealand Dollar, Swedish Kroner, Swiss Franc and US Dollar.

Hence, each day the reference short term yield curve (out to one year) from the LIBOR is updated. The futures themselves are exceptions of 100 minus the 3M-LIBOR at the date of maturity of the future. For instance, if the implied interest rate is 0.250%, the IMM index is quoted as 99.750 (IMM Index = 100.000 – 0.250% = 99.750), and this is also the price of a futures contract. The gain or loss of these contracts is described by one basis point (0.01%), which equates to a 25.00 US dollar movement in contract value, derived as follows: $1 million notional loan $\times$(0.01%) $\times$ (90/360) = $25.00. However, the minimum price fluctuation is set at half a basis point or 0.005%, equal to $12.50, based on $1 million face value 90-day instruments. For example, if the price was rising from 96.99 at the beginning of the day and closed at 97.00 (implying a LIBOR decrease from 3.01 to 3.00%), $25 is paid from the seller's margin account into that of the buyer so that buying the contract is equivalent to lending money, and selling the contract short is equivalent to borrowing it.

Indeed, the Eurodollar does not deliver an actual time account, but merely the present value cash equivalent of that account. The regular contracts expire either in March, June, September or December, extending outwardly for 40 quarterly expiring contracts in the long term; shorter maturity monthly contracts are traded; however, their relative volume, trading activity and quoting activity are several orders of magnitude lower than the quarterly contracts. The last trading day is the second business day prior to the third Wednesday of the delivery month in both New York and London. At any given time, the exchange lists four of the monthly series, bringing the total number of available Eurodollar future contract maturity types to 44 – namely, 40 quarterly and 4 monthly serial contracts. I have data for all contracts, but for brevity I focus on the regular quarterly contracts that expire at the on 40 IMM dates out to ten years. Currently, more than 70% of Eurodollar futures are traded simultaneously on the electronic trading platform; CME Globex platform.[1]

The core source for the regulatory violations in relation to the 3M-LIBOR

[1] CME Group, Understanding Eurodollar Futures, http://www.cmegroup.com/trading/interest-rates/understanding-eurodollar-futures.html [accessed August 18, 2015].

will be the CFTC reports relating to the current group of banks and brokers that have settled fines with various national regulatory bodies. As of August 1, 2015 this group consists of Barclays, UBS, RBS and Rabobank. Each bank has been investigated and fined by the Commodity Futures Trading Commission (CFTC) and the Financial Services Authority (FSA). Furthermore, on April 23, 2015 the Deutsche Bank is the latest bank that has been fined by CFTC for manipulation, attempted manipulation, and false reporting of LIBOR and Euribor. The documentation of manipulation of the LIBOR reference rate will be based on each bank's report. Overall, the reports cover a period of seven years from 2005 to 2012. I also focus on the reports primarily relating to the 3M US LIBOR submissions as this is the reference for Eurodollar futures.

The FSA [2012] report documents the earliest LIBOR manipulation event in 2005 in the dataset when Barclays' derivatives employees made a total of 257 requests to Barclays LIBOR submitter to set the submission rate for LIBOR and Euribor. This was a routine request related to their derivatives trading positions and their own bonuses. Figure 5.3 shows an example graph showing that Barclays' submission was consistent with certain requests. For example, on November 22, 2005, Barclays' swap trader made a request in relation to Barclays' 3M US dollar LIBOR submission: "We have to get kicked out of the fixings tomorrow; we need a 4.41 fix in 3M (high fix)". The submitter sent a positive response to this request. Figure 5.3 shows the change in Barclays' submission compared to other panel banks' submission and the final 3M rate.

Figure 5.3 shows that Barclays' 3M US dollar LIBOR submission, which had been at the final benchmark rate, increased to a level above the benchmark rate on the day the swap trader requested a higher submission. It then remained

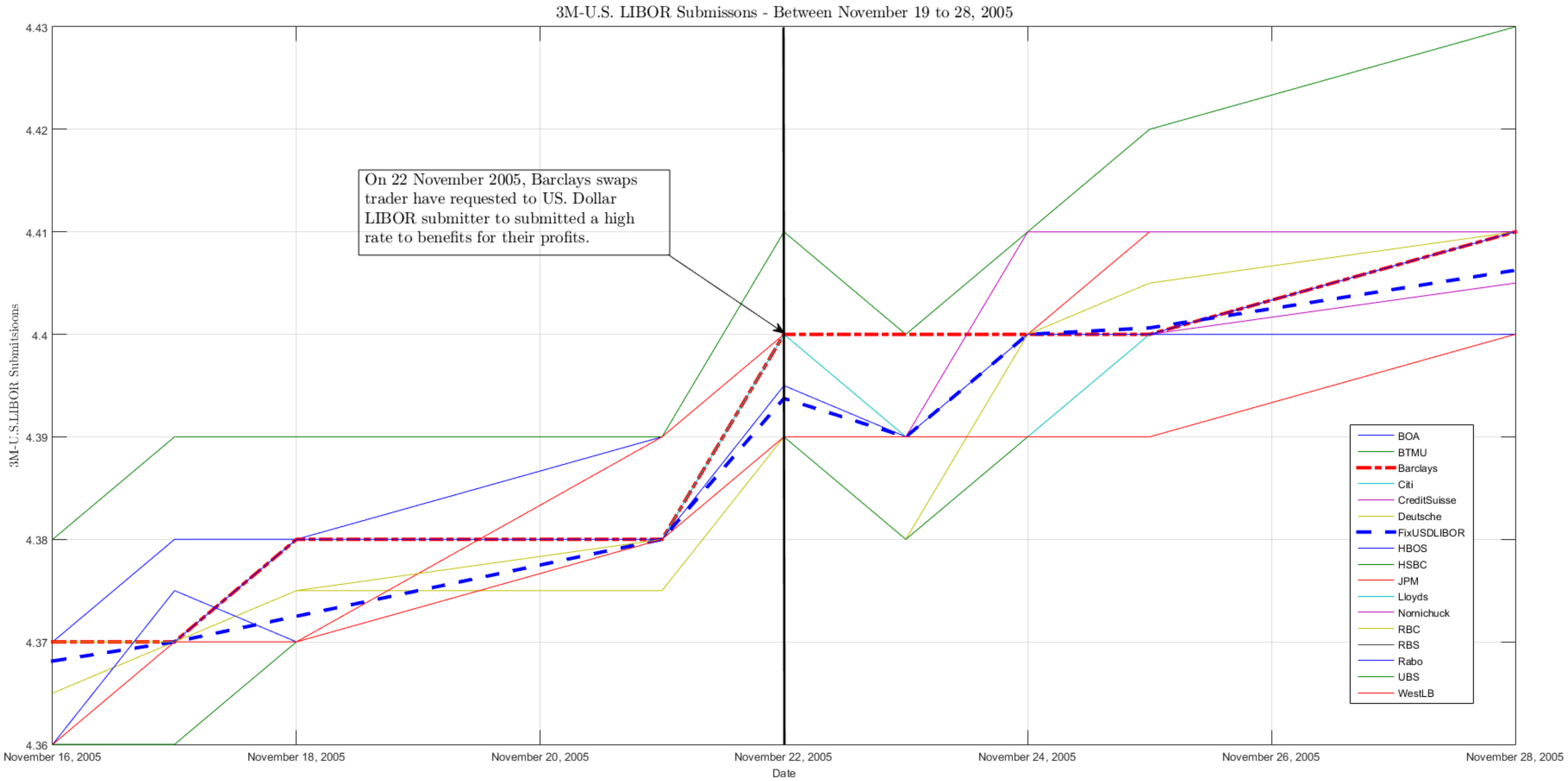



**Figure 5.3: Barclays' 3M US Dollar LIBOR submission around November 22, 2005.**

Note: This figure presents how Barclays' LIBOR submitter manipulated LIBOR when he dishonestly fixed the submission rate on November 22, 2005. It can be seen in this figure that during a normal day Barclays' LIBOR submitter regularly fixed his LIBOR between one of eight which is in the middle of the submission rate (U.S.LIBOR has 16 bank members on the panel). However, on November 22, 2005 which is identified as the date of a LIBOR manipulation event, the interest rate derivatives trader sent a request to the LIBOR submitter to submit a high fix, so the traders could maintain their trading position. In response, the LIBOR submitter fixed the LIBOR at a higher rate than the normal level in order to increase daily LIBOR to the rate the traders had requested.

high at the same level as the benchmark rate and finally increased again until the benchmark rate reached 4.41 as the swap trader had requested. Therefore, Barclays' submissions, on November 22, 2005 and later that week, were consistent with the request for a high 3M US Dollar LIBOR. The LIBOR investigation also showed that the artificial rate had been tight in both high and low fixing, for instance on May 23, 2007 the Barclays swaps trader requested the submitter to submit a low LIBOR at 5.36 for the next day(May 24, 2007). He sent an email to submitter "Pls., go for 5.36 again tomorrow, very long and would be hurt by higher setting...Thank"; then on May 23, 2007 Barclays submitted the 3M-USD LIBOR at 5.36 and the LIBOR was announced at 5.36 as their swap trader request. Another example was on September 04, 2007 regarding CFTC report when the UBS swap traders explained to their manager why they requested a low LIBOR submission, so their submitter submitted a low rate at 5.2(see Table 5.1 for Barclay's LIBOR submission).

There is substantial evidence that Barclays' LIBOR fix was consistently manipulated during the volatile global conditions of the sub-prime crisis (circa 2007–2008), in order to manage what it believed were inaccurate and negative public and media perceptions that it had a liquidity problem, (by lowering its submissions – which were believed to be too low (see p. 3, CFTC, 2012)). This activity continued after BNP Paribas House's hedge funds had been frozen, followed by the collapse of Northern Rock and Lehman Brothers. Finally, after Barclays raised7.3 billion of capital from its Qatari shareholders, the CFTC began an investigation in the United States into the suspected "low-balling" of LIBOR submissions by global banks.[1] In early December 2007, a Barclays' employee alerted the U.K.

[1]P. Aldrick, Barclays: how the LIBOR manipulation unfolded, http://www.telegraph.co.uk/finance/ newsbysector/banksandfinance/9360469/Barclays-how-the-LIBOR-scandal-unfolded.html, [accessed August 19, 2014].

Financial Services Authority (FSA) in which LIBOR rate setting at a dishonesty level was discussed (see p. 22, CFTC, 2012).

The analysis is continued by comparing the LIBOR submission rates with the benchmark rate before and after the collapse of Lehman Brothers. Figure 5.1 and Table 5.1 are examples of graphs and tables showing that panel banks submitted rates that were believed to be abnormally low before the collapse of Lehman Brothers (June 18, 2008). When the end of Lehman Brothers made it clear to all that the crisis would last, the LIBOR rose sharply. It had dropped heavily during the previous period to what would be an almost "normal" level for an almost normal situation, but was a totally abnormal level for a critical situation (October 15, 2008).

Figure 5.2 shows individual banks' quotes on USD 3M LIBOR (left axis) and p-value of cross-sectional correlation from this panels' banks (right axis). It can be seen in this figure that the individual bank quotes are very tightly clustered from 1996 until before the collapse of Lehman Brothers (on August 15, 2008) where a clear structural break begins as shown in figure 5.1. Before the banks collapse, the ranking correlation of banks' quotes are notably high as the p-vale of Spearman's rank correlation test, are nearly zero. It can be assumed that banking submissions for the 16 individual banks mostly remained at the same level or at a constant rank throughout this period. After the collapse, the submission quotes are highly volatile and farther from the LIBOR fixing. It also appears that the p-value of the banks rank submission is highly volatile. Since 2008, the significant spikes of p-value show that the ranking of bank quotes is continually changing over the period . Furthermore, the correlation breaks completely with the collapse of Lehman Brothers. Therefore, although it is not clear whether the banks manipulated their submission rate before and during the crisis, the significant change of ranking correlation of the banks' quotes suggest that something different was happening post crisis.

In order to track the LIBOR manipulation, Timothy Geithner, President of the Federal Reserve Bank of New York, sent a document containing a proposal for consultation to tackle the problem to Sir Mervyn King, the governor of the Bank of England.[1] After receiving this document from the New York Fed, the FSA in London officially joined the international investigation, as a result of which the CFTC broadened its investigation. It decided to cooperate with the Securities and Exchange Commission (SEC) and began to communicate with multinational regulators, such as the European Commission (EC), the Swiss Competition Commission (ComCo), Canada's Competition Bureau and the Japanese Regulator. A major effect of this cooperation was that the Swiss bank UBS agreed to end a tax dispute over US citizens and hand over 4,450 account details to assist the LIBOR investigation.

The FSA launched an investigation into Barclays to respond to allegations of LIBOR manipulation in June 2010, and in 2011, the FTC Capital and Charles Schwab Corporation filed a lawsuit against 12 major banks, including Barclays, RBS, HSBC and Lloyds, claiming that they had conspired to depress the LIBOR artificially.[2] After the investigation, some of the major banks that were found to have manipulated the LIBOR rate, including Barclays and UBS, were fined by

[1] Timothy Geithner made recommendations on LIBOR to Sir Mervyn King (published online June 1, 2008), http://data.newyorkfed.org/newsevents/news/markets/2012/LIBOR/June-1-2008LIBOR-recommendations.pdf, [accessed August 19, 2014].

[2] M. Gilbert, G Finch and A. Worrachate, London banks seen rigging rates losing credibility with markets, Bloomberg Markets Magazine (published online Nov 22, 2011) http://www.bloomberg.com/news/print/2011-11-23/london-banks-seen-rigging-rates-for-decades-losing-credibility-in-markets.html, [accessed August 19, 2014].

Commodity Futures Trading, the US Department of Justice, and the Financial Services Authority in the United Kingdom. Subsequently, RBS and ICAP were fined by US and UK regulators in 2013, and later the European Commission announced fines of six major banks for manipulating the LIBOR rate for the Japanese Yen in the period 2007–2010. After the US Department of Justice completed a criminal investigation into the LIBOR manipulation on June 27, 2012, Barclays was the first bank to be fined – $200 million by the Commodity Futures Trading Commission [CFTC, 2012], $160 million by the US Department of Justice and £144.5 million by the FSA [FSA, 2012] – for attempting to manipulate the world's benchmarking borrowing rate, LIBOR.

From extensive textual analysis of the current settlements, it is evident that the 3M-LIBOR is the most commonly mentioned US Dollar rate (second to the JPY rate) , followed by the 1M-LIBOR. A summary of the evidence of the abuse of the 3M-LIBOR and 1M-LIBOR is provided in Table 5.3, with an example of the communication content between traders and LIBOR submitters during the LIBOR manipulation.

## 5.3 PIN and LIBOR Manipulation

The following reviews the anecdotal trends in the estimated PIN coefficients for the various contracts relative to the specific events catalogued in Table 5.3 versus the longer run patterns observed in the data over the 1996 to 2014 sample period. This section begins by comparing the historical PIN with specific events related to the LIBOR manipulation records. Since the LIBOR abuse was revealed by the US Department of Justice, FSA and CFTC investigations, I make the conjecture

**Table 5.3: Time-line of the reports of LIBOR manipulation.**

| Year | Date | Maturity date | Bank | Inappropriate submissions following request by | Summary of email and conversations about LIBOR abuse action | Action |
|---|---|---|---|---|---|---|
| 2005 | March 22 | 13/06/05 | Deutsche Bank | Deutsche Bank US Dollar LIBOR submitter communicated with US Dollar trader. | "if you have an interest in a high or a low fix let me know" then the trader reply "Thanks-our CP guys have been looking for it a bit higher". | High setting |
| | May 27 | 13/06/05 | Barclays | Barclays' derivatives traders requested to US Dollar LIBOR submitter. | "Who's going to put my low fixing in?" The US Dollar LIBOR submitter replied "will be here if you have any requests for the fixing". | Low setting |
| | Aug 2 | 19/09/05 | Rabobank | The Rabobank US Dollar Desk Manager email to Senior Manager. | "Can you please set 6mL as high as possible?" Senior Manager reply on the next day: "Ok, leave it with us, will do our best." | High setting |
| | Sep 21 | 19/12/05 | Deutsche Bank | The London MMD manager communicate with US Dollar LIBOR submitter | " Lower mate lower", the submitter reply "will see what I can do...", then the manager reply "Is doin it on purpose because they have the exact opposition-on which they lost 25Mio so far-let take them on", and finally submiter reply "ok let see if we can hurt them a little bit mor the" | Low setting |
| | Nov 22 | 19/12/05 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "We have to get kicked out of the fixings tomorrow, we need a 4.17 fix in 1m (low fix), We need a 4.41 fix in 3m (high fix)." | Low and High setting. |
| 2006 | Feb 1 | 13/03/06 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "We need 3m to stay low for the next 3 sets". | Low setting |

*Continued on next page*

| Year | Date | Maturity date | Bank | Inappropriate submissions following request by | Summary of email and conversations about LIBOR abuse action | Action |
|---|---|---|---|---|---|---|
| | Feb 3 | 13/03/06 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "Would love to get a high 1m, also a low 3m if possible, thanks" | Low and High setting. |
| | Feb 7 | 13/03/06 | Barclays | Barclays' derivatives traders requested to US Dollar LIBOR submitter. | "High 1m and high 3m if it's possible please. Have v. large 3m coming up for the next 10 day or so." | High setting |
| | Feb 15 | 13/03/06 | Barclays | Barclays' derivatives traders request to US Dollar LIBOR submitter. | Trader requested "Please go for [unchanged], or lower if poss". | Low setting |
| | Mar 10 | 19/06/06 | Barclays | Barclays' derivatives traders request to US Dollar LIBOR submitter. | Two US Derivatives trades made email requests for a low three month US Dollar LIBOR submission for the coming Monday 13th. | Low setting |
| | Mar 16 | 19/06/06 | Barclays | Barclays' derivatives traders request to US Dollar LIBOR submitter. | Traders requested a high 1m and low 3m US Dollar LIBOR submission. Submitter responded "For you... anything. I am going to go 78 and 92.5." | Low and High setting. |
| | Mar 27 | 19/06/06 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "We need low 1m and 3m LIBOR." | Low setting |
| | Mar 31 | 19/06/06 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "We have another bix fixing tomorrow and with the market move I was hoping we could set the 1m and 3m LIBORs as high as possible." | High setting |
| | Apr 7 | 19/06/06 | Barclays | Barclays' derivatives traders request to US Dollar LIBOR submitter. | Trader requested low 1m and 3m US Dollar LIBOR: "If it's not to low 1m and 3m would be nice, but please fell free to say 'no'. Coffees will be coming your way either way, just to say thank for your help in the past few weeks." A submitter responded "Done...for you big boy". | Low setting |

*Continued on next page*

| Year | Date | Maturity date | Bank | Inappropriate submissions following request by | Summary of email and conversations about LIBOR abuse action | Action |
|---|---|---|---|---|---|---|
| | Jun 30 | 18/09/06 | Rabobank | The Rabobank US Dollar trader email to US Dollar trader submitter. | "Morning mate, hope u got through yesterday ok. Usual request. Low 1s, high 3s if you can" | Low and High Setting |
| | Aug 6 | 18/09/06 | Barclays | Barclays' derivatives traders requested to US Dollar LIBOR submitter. | Trader requested "Pls set 3m LIBOR as high as possible today?" Submitter responded "Sure 5.37 okay?" | High setting |
| | Sep 1 | 18/09/06 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "High 3mth LIBOR for ur old mucker pls chief". | High setting |
| | Sep 13 | 18/09/06 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "Can we please keep the LIBOR fixing at 5.39 for the next few days. We do not want it to fix any higher than that." | Low setting |
| | Sep 15 | 18/12/06 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "Usual favours, can you keep 3s LIBOR at 39 for the next few days pls mate, cheers". | High setting. |
| | Oct 13 | 18/12/06 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "If you would be so kind and keep 3s low for monday's fix, Think it will come in at 37.5, would be a bit painful if was much higher." | Low setting. |
| | Dec 14 | 18/12/06 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "For Monday we are very long 3m cash here in NY and would like the setting to be as low as possible..' Thanks." | Low setting |
| 2007 | Mar 16 | 18/06/07 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "High 3mth LIBOR on Monday pls." | High setting. |

*Continued on next page*

| Year | Date | Maturity date | Bank | Inappropriate submissions following request by | Summary of email and conversations about LIBOR abuse action | Action |
|---|---|---|---|---|---|---|
| | March 14 | 18/06/07 | Deutsche Bank | London pool trading manager communicate with US Dollar LIBOR submitter. | "These markets falling in is not good for us personally. We need good old fasioned boom time...", submitter reply " rekon 3s libor only 34.75 fyg even with edh where it is now which is b11x", the manager "Get it lower, we need it..", and submitter "just spoke to him. now thinking 34.5, I think should be lower still...' | Low setting |
| | May 23 | 18/06/07 | Barclays | Barclays swaps traders request to US Dollar LIBOR submitters. | "Pls., go for 5.36 LIBOR again tomorrow, very long and would be hurt by a higher setting....thanks." | Low setting |
| | Aug 3 | 17/09/07 | Barclays | Barclays US Dollar traders request to US Dollar LIBOR submitters. | During the financial crisis, Barclays directed its US Dollar LIBOR submitters to lower their daily US Dollar LIBOR submissions in order to protect Barclays'reputation against market perceptions that Barclays had a liquidity problem based in part on its high LIBOR submissions. | Low setting |
| | Aug 10 | 17/09/07 | UBS | The UBS US Dollar trader have requested to US Dollar LIBOR submitters. | Trader requested lower rate on 1 week and 2 week LIBOR US Dollar submission. | Low setting. |
| | Aug 13 | 17/09/07 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "Gonna need a frickin high 6 mth fix tomorrow if ok with u, 5.42? I'll send a mail, write a post it not as well, it's 750 million 6mth fix! Could be risky!" | High setting. |
| | Aug 14 | 17/09/07 | UBS | The UBS US Dollar trader request to US Dollar LIBOR submitters. | "my indications are deliberately on the low side ..." | Low setting. |

*Continued on next page*

| Year | Date | Maturity date | Bank | Inappropriate submissions following request by | Summary of email and conversations about LIBOR abuse action | Action |
|---|---|---|---|---|---|---|
| | Sep 3 to 10 | 17/09/07 | UBS | Conversation between the UBS US Dollar traders and US Dollar LIBOR submission manager (ALM Manager). | Traders explained to ALM Manager why UBS wanted to submit low LIBOR submission in most currencies. | Low setting. |
| | Sep 7 | 17/09/07 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "If you can keep em up there that would save my arse a bit (til sep rolls off), cheers matey." | High setting. |
| | Oct 17 | 17/12/07 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "A nice low 1 month for LIBOR for the rest of the week please matey. Cheers." | Low setting. |
| | Nov 5 | 17/12/07 | UBS | Conversation between the UBS US Dollar traders and US Dollar LIBOR submission manager (ALM Manager). | Trader requested lower rate on LIBOR US Dollar submission. | Low setting. |
| | Nov 23 | 17/12/07 | UBS | Conversation between the UBS US Dollar traders and US Dollar LIBOR submission manager (ALM Manager). | Trader requested lower rate on LIBOR US Dollar submission. | Low setting. |
| | Nov 28 | 17/12/07 | Barclays | Barclays US Dollar traders have requested to US Dollar LIBOR submitters. | "LIBORs are not reflecting the true cost of money. I am going to set..., probably at the top of the range of rates set by LIBOR contributors...The true cost of money is anything from 5-15 basic points higher." | High setting |

*Continued on next page*

| Year | Date | Maturity date | Bank | Inappropriate submissions following request by | Summary of email and conversations about LIBOR abuse action | Action |
|---|---|---|---|---|---|---|
| | Nov 29 | 17/12/07 | Barclays | Barclays US Dollar LIBOR submitters convened a telephone discussion with the senior Barclays Treasury managers and the US Dollar LIBOR submitters. | The supervisor said "if the submitters submitted the rate for a particular tenor at 5.50, which was the rate they believed to be the appropriate submission", Barclays would be twenty basic points above the pack and it's going to causes shit storm. | Low setting |
| | Dec 4 | 17/12/07 | Barclays | The Barclays senior US Dollar LIBOR submitter emails his supervisor. | [SIC] He stating that he submitted Barclays' 1m LIBOR at 5.30%, which was 4 basic points over the next highest submission and almost five basic points over the LIBOR FIXING. | High setting |
| | Dec 13 | 17/12/07 | Deutsche Bank | Frankfurt Non-Euro desk manager request to London pool trading manager. | I need your help..it it suits you can we put in a high LIBOR till next tuesday in the 3MTS? then he reply "ok". | High setting |
| 2008 | Feb 5 | 17/03/08 | Barclays | The Barclays US Dollar LIBOR trader stated in a call to his manager. | Manager instructed trader to tell Barclays' submitter to keep LIBOR low. | Low setting |
| | Mar 3 | 17/03/08 | Barclays | The Barclays senior US Dollar LIBOR submitter called to his colleague. | The manager told him that he should not be quite as aggressive in his submissions, meaning his submissions were too high. | Low setting |

*Continued on next page*

| Year | Date | Maturity date | Bank | Inappropriate submissions following request by | Summary of email and conversations about LIBOR abuse action | Action |
|---|---|---|---|---|---|---|
| | Mar 17 | 16/06/08 | Rabobank | The Rabobank US Dollar trader email to US Dollar Trader submitter. | "I know LIBORs are gonna come off today, I hv [SIC] quite a lot of fixings today, so the higher the better for me please." | High setting. |
| | Jun 17 | 16/06/08 | UBS | Conversation between the UBS US Dollar traders and US Dollar LIBOR submission manager (ALM Manager). | US Dollar Trader-Submitter state that "we will start lowering over the next few days to get to more or less middle of the pack until future notice did you need me for any thing else?" ALM Manager replied "nope that was it, thx we should bring it down fast so we are in line by friday with the pack". | Low setting. |

Note: This is a synthesis of reports from the CFTC, FSA and FCA on the inappropriate US Dollar LIBOR submissions by employees of Barclays, Rabobank and UBS from 2005 to 2008. It focuses on the US dollar submission, particularly anything relating to 3-month submissions. The majority of reports on manipulation refer to the 1-month and 3-month LIBOR rates, predominantly the 3-month. This can be attributed to the fact that the Eurodollar Future is primarily a 3-month forward LIBOR rate, while the 1-month rates are commonly used for the floating rate leg of fixed to floating swaps. Since these are the products that have the highest notional valuations outstanding, they have the most potential for profitable informed trades.

that individual events had an impact on the PIN in the Eurodollar futures market as the ED trading is based on this key rate. A closer look at the way in which the LIBOR investigation caused changes in the PIN begins with a document analysis using the Commodity Futures Trading Commission (CFTC) document when the CFTC investigated some of the major banks, such as Barclays and RBS, regarding dishonest or manipulative submissions of the LIBOR rate. As previously noted, two epochs appear to be important in this context, first the pre-crisis attempts to generate competitive advantage directly in the banks proprietary interest rate trading positions and second after the crisis as a systematic attempt to reduce the perception of increased borrowing costs as part of a beauty contest designed to 'reassure' investors and regulators.

According to the CFTC settlement document, Barclays' traders had attempted to manipulate the US Dollar LIBOR from at least mid-2005 to the autumn of 2007, and sporadically thereafter until 2009. For instance, on November 28, 2007, Barclays's employees, including the bank's senior Treasury managers, submitted a US Dollar LIBOR rate that was higher than the actual rate and LIBOR abuse occurred on the following day and into early December 2007.[1]

This activity is investigated in this study from the PIN value of the EDH8 contract, which had the closest expiry date series to this date. The PIN increased from November 24, 2007, when its value was 0.54 and reached a peak of 0.715 on December 15, 2007. Figure 5.4 illustrates the manipulation that began on November 28, 2007. This increasing value of the PIN is consistent with the CFTC document; for example, on November 28, 29 and 30, 2007 the PIN values

[1] CFTC, Order instituting proceedings pursuant to sections 6(c) and 6(d) of the Commodity Exchange Act, as amended, making findings and imposing remedial sanctions; In the Matter of Barclays plc, pp. 20–22.

were 0.64, 0.65 and 0.66 respectively, and reached a peak of around 0.715 on December 15, 2007 as Barclays' US Dollar LIBOR traders asked the submitter to keep the rate high.

There were many other instances of LIBOR abuse that can be illustrated by the movement of the PIN. The final notice from the Financial Services Authority (FSA) issued to Barclays Bank plc stated that at least 14 of the bank's interest rate derivatives traders had requested the LIBOR submitter to fix the LIBOR rate. However, this was not always done only to improve their trading position, but also sometimes to protect the bank's reputation. For instance, on February 5, 2008, a Barclays' US Dollar Derivatives manager instructed traders to keep the LIBOR rate lower than the real rate to make the bank look healthier than usual during the financial turmoil.[1] It is interesting to see how this activity is reflected in the ED PIN to determine if the level of informed trading that occurred in the Eurodollar market during this period changed substantially. Prima facie evidence indicates that the LIBOR abuse appears to have possibly exerted some influence on the historical development of the PIN. The PINs of EDM8, EDU8, EDZ8 and EDH9 are used to indicate the LIBOR setting from February 5 to 11, 2008. Of course, the limitation of this study is that that we cannot identify the specific trades undertaken by the specific banks as this information is not available publicly.

Figure 5.5 illustrates the historical PIN during this period. The PIN indicator shows an increasing trend from around 0.68 on February 3 2008 to just around 0.70 on February 5, 2008 reaching a peak of around 0.76–0.78 on February 11, 2008. This figure illustrates that the PIN was still in an upward trend until

[1]Financial Services Authority, Final Notice To Barclays Bank plc, p. 11.

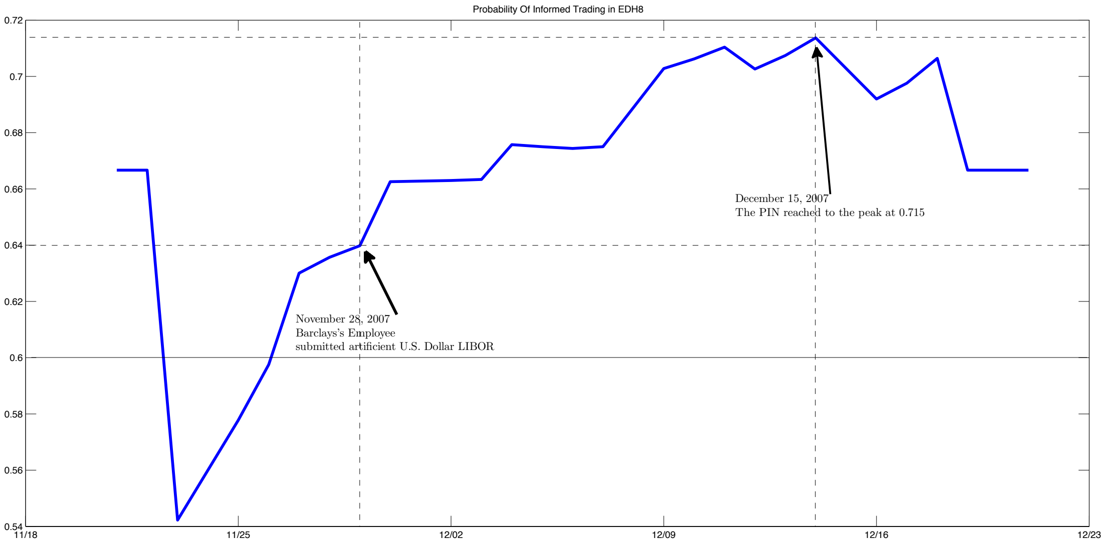



**Figure 5.4: Probability of informed trading between November 28 and December 15, 2007.**
Note: This figure shows the development of the PIN during the LIBOR abuse between November 28 and December 15, 2007. According to the CFTC document, Barclays' traders attempted to manipulate the US Dollar LIBOR on November 28, 2007, when Barclays' employees, including senior Barclays Treasury managers, submitted a higher US Dollar LIBOR rate than the real rate.

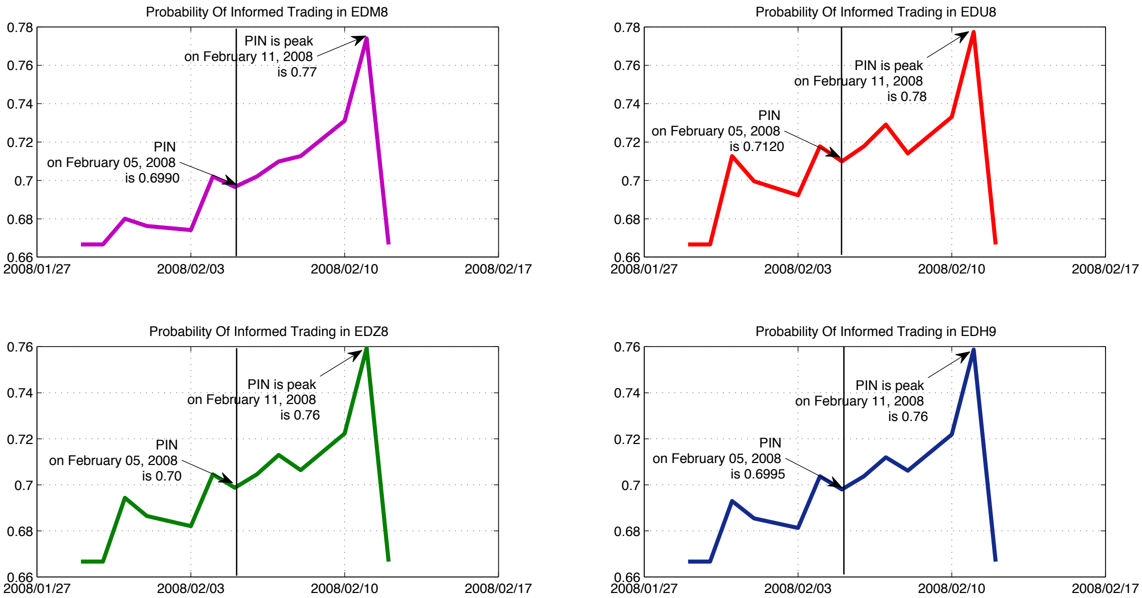



**Figure 5.5: Probability of informed trading between February 5 and 11, 2008.**

Note: This figure illustrates the development of the PINs of EDM8, EDU8, EDZ8 and EDH9 between February 3 and 10, 2008. On February 5, 2008 a Barclays' Derivatives manager instructed traders to keep the LIBOR rate low. There appears to be an increasing trend of the PINs during this period and a drop after the event date.

February 5, 2008, and the increase shown in this figure is possibly evidence of the abuse of Eurodollar Futures.

Barclays was not the only major high street bank involved in manipulating the LIBOR. Rabobank (Coöperative Central Raiffeisen Boerenleenbank B.A.), the Dutch multinational banking and financial services company headquartered in Utrecht was also guilty of this 'misdemeanour' and fined accordingly. According to the CFTC settlement report (CFTC [2013]), the Rabobank LIBOR submitter submitted preferential rates in an attempt to manipulate the US Dollar LIBOR to benefit its trading position. In fact, Rabobank employees frequently attempted to manipulate the US Dollar LIBOR over a period from at least mid-2005 to at least late 2008. There are many examples of submitters being requested by traders to fix the LIBOR; for example, on September 07, 2007, Rabobank's senior US Dollar trader emailed the US Dollar submitter to keep the 3M US Dollar LIBOR high for the rest of the week.[1] The fact that this period was characterized by an increasing trend of the historical PIN is evidence that the Eurodollar futures market was affected by the setting of the LIBOR. This manipulation influence on the EDU7 contract, which expired on September 19, 2007 or just about two weeks after the event, is used to investigate this incident by matching it with the US Dollar LIBOR. The results indicate that, on September 07, 2007 the PIN had increased from 0.68 on September 6 to 0.70 , and it was kept higher than 0.70 for the remainder of the week (see Figure 5.6). The PIN reached a peak of 0.80 on September 17 and dropped to 0.695 on the following day. The historical PIN for this period could provide some evidence that the Eurodollar market had been manipulated by informed traders; moreover, another result is that the value of

[1] CFTC, Order instituting proceedings pursuant to detection(c) and 6(d) of the commodity exchange act, as amended, making findings and imposing remedial sanctions in the matter of Coöperative Central Raiffeisen Boerenleenbank B.A., p. 10.

the PIN during this period was higher than average for EDU7 (0.6855).

Most of the LIBOR abuse activity that affected the Eurodollar futures trading occurred before and during the sub-prime crisis; however, one major event that had an impact on the LIBOR and Eurodollar Futures trading occurred much later, on February 28, 2012 when the US Department of Justice was conducting a criminal investigation into the LIBOR manipulation. Later that year, Barclays was the first bank to be fined $200 million by the CFTC, $160 million by the US Department of Justice, and 59.5 million by the UK's FSA for attempting to manipulate the key rate. Therefore, the Eurodollar Futures trading during this period is also investigated in this paper by simulating the PINs of EDH2, EDM2, EDU2 and EDZ2 contracts to compare them before, during and after 2012, when the LIBOR was manipulated. The action taken by the CFTC and FSA was a warning to traders to be wary of submitting an honest LIBOR rate and it should have reduced the instances of informed trading in the Eurodollar market. According to the investigation in this study , there was a significant drop in the PINs of these four contracts in mid-2012 (see Figure 5.7), especially EDM2, which expired in June 2012 with a maturity date around three months after the regulators announcement.

The EDM2 PIN dropped by 24% from around 0.84 to 0.60 from the beginning of 2012 to the end of that year. The decrease of the PIN would have been affected by the investigation, lawsuit and fines for LIBOR manipulation by the CTFC and the FSA, which could have influenced some informed traders. This influence is illustrated by the significant change in the historical PIN in 2012 with the downward trend after the LIBOR investigation (see Figure 5.7). Furthermore, after the LIBOR Regulation reform in July 2013, the PIN of every contract significantly

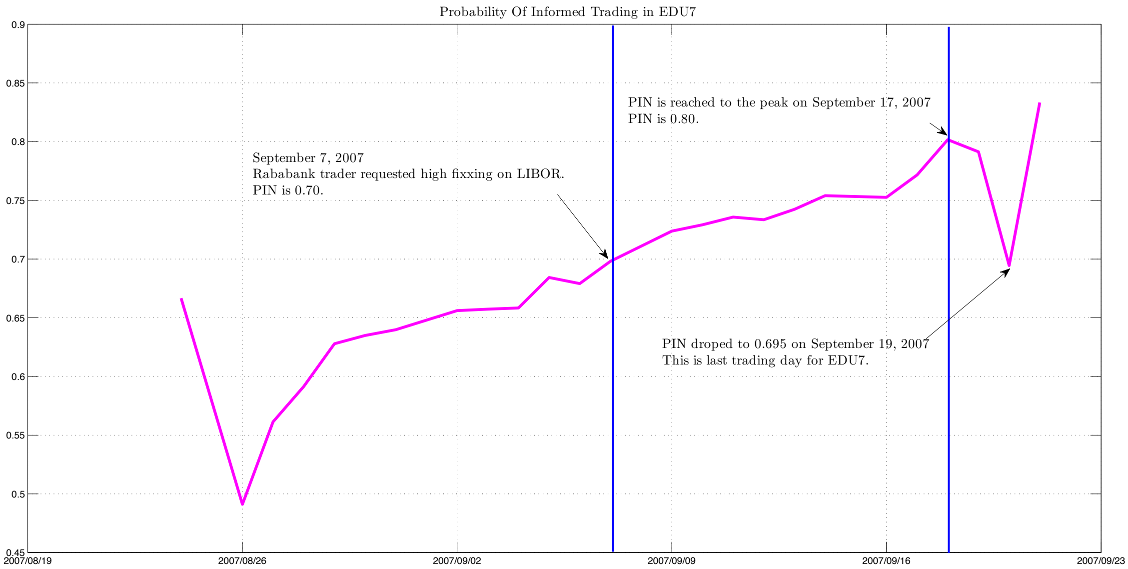



**Figure 5.6: Probability of informed trading between September 7 and September 22, 2007.**

Note: This figure presents the value of the PIN of EDU7, which reacted to the LIBOR abuse between September 7, 2007 and September 22, 2007 when Rabobank's US Dollar LIBOR traders asked the submitter to keep the LIBOR high for the rest of the week. As shown in Figures 5 and 6, the trend of the PIN also increased and remained high for the rest of that week. Since this trend could have been affected by the LIBOR abuse by Rabobank's employees, it is probably evidence to prove that the LIBOR was manipulated during this period.

dropped from around 0.68 to around 0.55 in the second half that year.

An interesting observation is that after the reform of the LIBOR fixing system and the exposure of bad practice, the PIN computed for the various Eurodollar contracts exhibits a steady decline with lower than average value of 0.68 after 2013. However, given the degree of fluctuation in the PIN through time and across contracts, further analysis is needed to determine the persistence of this trend, see Figure 5.7).

## 5.4 PIN and the Maturity Date Effect

This section contains an illustration of the PIN before and after expiration to investigate the relationship between the PIN and the maturity date. Its evolution is plotted with the expiration date of all 40 Eurodollar contracts to illustrate how it reacted before and after the maturity date. The most striking result to emerge from this figure is that, after the spiking of the PIN before the expiry date, there was an aggressive decline until the last trading day, and finally, another increase appears when the Eurodollar futures begin trading again on the day after the expiry date (see Figure 5.8). This pattern of the PIN appears in all 40 Eurodollar series contracts.

The PINs for individual Eurodollar futures contacts are divided into two periods, the first of which is between January 1, 1996 and December 31, 2007, while the second is between January 1, 2008 and December 31, 2013. The average value of the daily PINs was constructed for 60, 30, 20 and 10 days for fixed time windows around categories of identified events. The first identified sub-event was

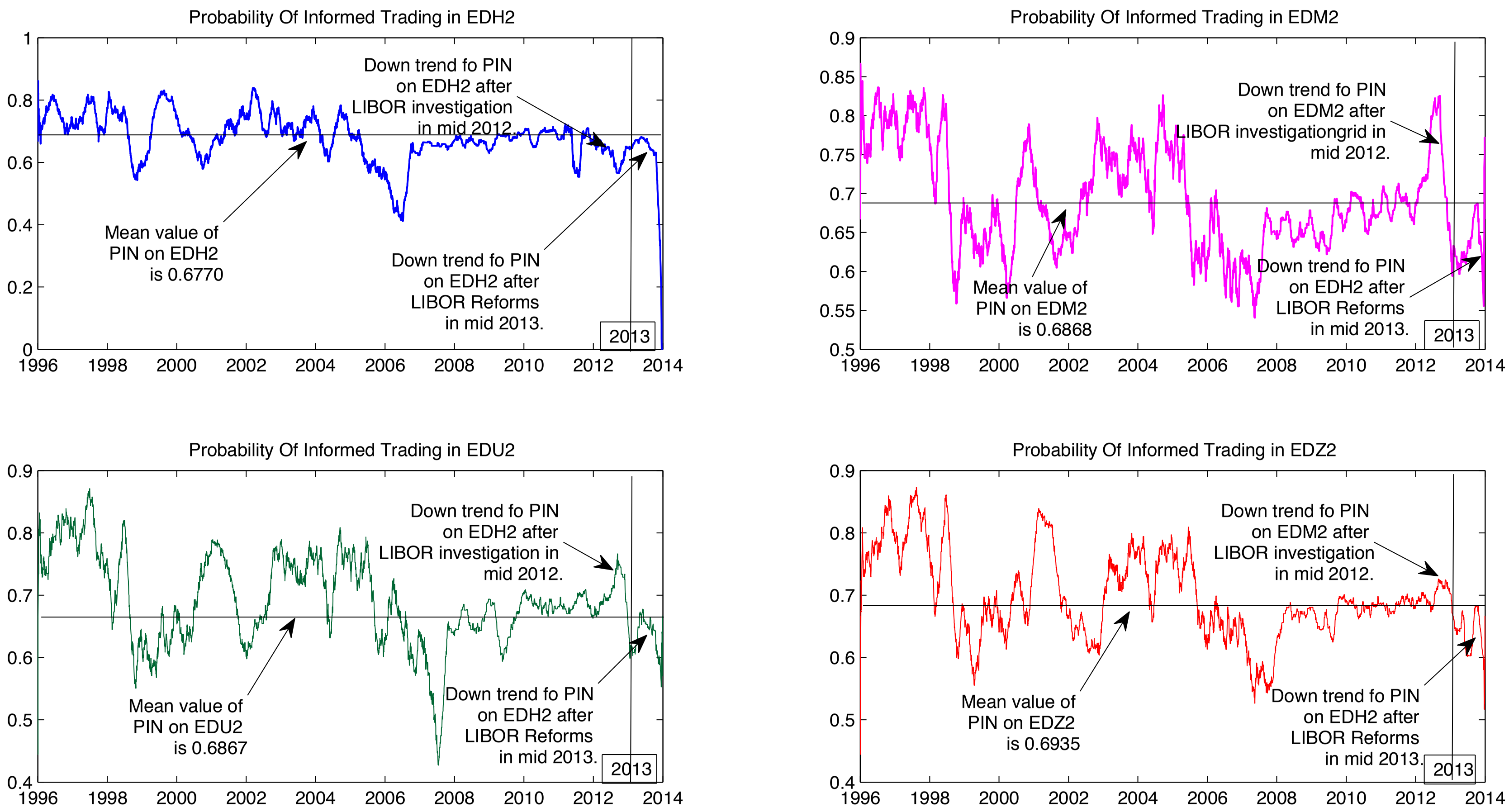



**Figure 5.7: The probability of informed trading after LIBOR criminal investigation and reform.**
Note: This figure demonstrates the modification of the PIN in two periods, the first of which was after the PIN investigation and the announcement of fines imposed on Barclays, RBS, UBS and Rabo bank by the CFTC. There was a decreasing trend of the PIN in the first quarter of 2012, after the investigation of all ED contracts, followed by an upward trend from the last quarter of 2012 until mid-2013. The second period was after the LIBOR reform in mid-2013, when another downward trend of PIN appeared, probably as a result of the new law to protect the LIBOR from manipulation.

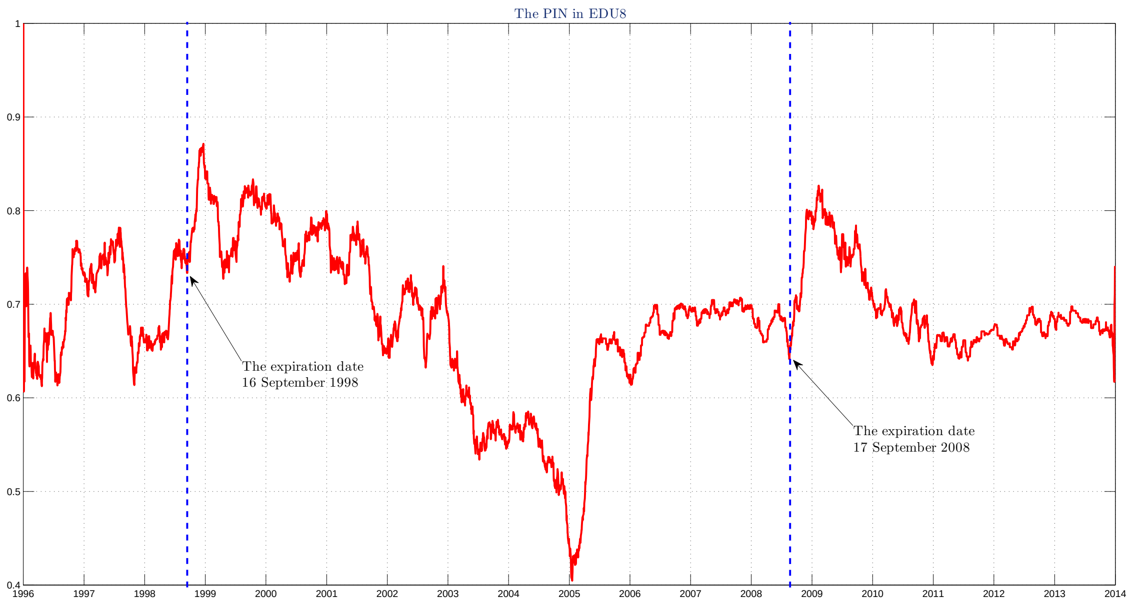



**Figure 5.8: The historical PIN of EDU8 with the expiry date.**
Note: This figure presents the historical PIN of EDU8 around the expiry date. The PIN for this event was dropping rapidly before the last trading day, but increased when EDU8 began trading again on the following day. This pattern is repeated for all the Eurodollar futures contracts.

based on the regulatory activity around the current catalogue documenting LIBOR manipulation from the CFTC and the FSA, from Table 5.3 while the second was based on contract maturity dates by year (see Table 5.4).

In terms of the first sub-events, the average value of the PINs was not significantly different between 60, 30, 20 and 10 days for fixed time windows around regulatory based events. The smallest PINs appeared 10 days before the identified event for both the pre-2008 and post-2008 sample, whereas the highest PIN appeared 10 days after the event. Furthermore, in the pre-2008 period, the PINs slightly increased from the event date and reached a peak 60 days after the event with an average value of 0.7069.

As for the second sub-event, which was around the maturity date, the average PIN slightly increased from 60 days before the maturity date, and then the PIN dropped on the last trading day and increased again after the expiry date. Interestingly, the average PIN was identical on the maturity date, before and after 2008. Although there were slight differences in the evolution of the PIN for both regulatory-based events and around the maturity date, these differences were not found to be statistically significant (see Table 5.4).

## 5.5 Summary Chapter

I have undertaken an empirical microstructure analysis of the Eurodollar futures market, based around a rolling window adaptation of the model of Easley and O'Hara [1992] and Easley et al. [1996]. This is probably the first attempt to apply a simple sequential trade model to analyse the reaction between the probability of

**Table 5.4: Comparison of the results of the mean value and standard deviation.**

| Day before the event | Mean value of PIN for LIBOR abuse post-2008 | Mean value of PIN for LIBOR abuse pre-2008 | PIN before/after maturity date post-2008 | PIN before/after maturity date pre-2008 |
|---|---|---|---|---|
| **–60** | **0.6752** | **0.6938** | **0.6933** | **0.6926** |
| | (0.0089) | (0.0196) | (0.0324) | (0.0223) |
| **–30** | **0.6724** | **0.6946** | **0.6975** | **0.6933** |
| | (0.0089) | (0.0237) | (0.0400) | (0.0278) |
| **–20** | **0.6717** | **0.6957** | **0.6999** | **0.6926** |
| | (0.0088) | (0.0260) | (0.0433) | (0.0313) |
| **–10** | **0.6731** | **0.6967** | **0.7006** | **0.6930** |
| | (0.0081) | (0.0287) | (0.0459) | (0.0354) |
| **0** | **0.6732** | **0.6968** | **0.6552** | **0.6695** |
| | (0.0117) | (0.0325) | (0.4656) | (0.0368) |
| **10** | **0.6738** | **0.6995** | **0.7066** | **0.6978** |
| | (0.0053) | (0.0324) | (0.0448) | (0.0420) |
| **20** | **0.6686** | **0.7020** | **0.7097** | **0.6979** |
| | (0.0053) | (0.0318) | (0.0477) | (0.0466) |
| **30** | **0.6671** | **0.7043** | **0.7116** | **0.6970** |
| | (0.0043) | (0.0322) | (0.0501) | (0.0488) |
| **60** | **0.6663** | **0.7069** | **0.7157** | **0.6891** |
| | (0.0017) | (0.0350) | (0.0550) | (0.0530) |

Note: This table presents the comparison of the results of mean value and standard deviation of the PIN 60, 30, 20 and 10 days before and after the events of LIBOR manipulation according to the CFTC and FSA documents, and also before and after the maturity date in two periods: from 1996 to the end of 2007 and from 2008 to the end of 2013.

informed trading and market abuse based on manipulating the LIBOR for highly active derivatives markets, such as Eurodollar futures, or LIBOR-based derivatives traded on the CME's Globex platform. However, the difference between equity and derivatives markets should be considered.

Overall, the evolution of the PIN appears to vary systematically with maturity , since it dropped around the last trading day prior to maturity and then recovered when the Eurodollar futures resumed trading the following day. The findings show that in the lead up to some of the more blatant attempts at manipulating the LIBOR fix the PIN provides a good early warning signal. However, the average adjustments to the PIN relative to the dates recorded in the current catalogue of documentation from the CFTC and the FSA on LIBOR manipulation was not statistically significantly different relative to both the persistent variation relative to maturity and the general long-run variation in the PIN. This may be because the PIN approach was unsuitable or it was possibly used incorrectly for this type of data, or the LIBOR manipulation only had a minor effect. However, I leave full investigation of this issue to future work as more granularity emerges from the current round of court cases provides more direct evidence of channels of informed trading (thus allowing us to better the cases).

# Chapter 6

# Order Flow Toxicity and Informed Trading Around Known Market Manipulation Events: Evidence from Interest Rate Futures

The objective of this paper is an ex-post review of the effectiveness of PIN and VPIN in determining changes in the information structure and order flow of a futures market around documented episodes of recorded manipulation of the reference rate, from the various publicly available regulatory reports. In keeping with previous studies on interest rate derivatives, this analysis finds that the average PIN is far higher for futures than for the equity market at or above 60%. Furthermore, I find a very strong connection between PIN, VPIN and time to maturity of the contract that is not fully explained by the time variation in activity in the market. However, an event study using both a new bootstrap approach and asymptotic standard error on the VPIN and PIN respectively around documented LIBOR manipulation cases has mixed results. For certain events I see a substantial change in the average detected levels of PIN and VPIN, however a cross sectional analysis of all reported cases up to mid-2015, indicates no significant change in the PIN and VPIN for the contracts in this Eurodollar sample.

## 6.1 Introduction

There are currently two key debates ongoing in the financial community related to issues around market manipulation. First, in the policy and practitioner community, there is evidence of manipulation of key benchmarks by financial institutions, and there is some disagreement as to whether this occurred primarily to improve their trading opportunities or to reduce the perception of balance sheet weakness. Second, is an academic debate as to whether the current high profile measures of the probability-of-informed trading (PIN) and 'market-toxicity' (measured by a volume adjusted PIN denoted VPIN) actually provide substantive measurement of the phenomena in question.

This study takes a comprehensive dataset from Chicago Mercantile Exchange (CME) tape data, which covers every inside quote and trade from 1996 to 2015 for the 40 LIBOR referenced quarterly dated Eurodollar futures contracts. First, I apply different flavours of the PIN and VPIN metrics over a variety of estimation windows. A variety of tests is constructed to see if the pattern of the PIN and VPIN exhibit structural changes around documented cases of manipulation of the LIBOR reference rate. Finally, I compare this to systematic fluctuations in these measures relative to the futures term structure.

Unsurprisingly, given the scale of the task, this study finds very mixed results. Both PIN and VPIN vary systematically and in a statistically significant pattern in respect to the term structure of the futures contracts. PIN varies in a v-shaped pattern, with long (2000 to 3500 days) and short maturity (0 to 500 days) contracts having significantly higher PIN and VPIN measurements than intermediate contracts (which are actually the most heavily traded). VPIN measurements of informed trading are substantially lower, over the entire gamut of calculation measures. However, when I move to documented cases of market manipulation in the reference rate, the results are ambiguous. There are definitive cases when the PIN (and to an extent certain flavours of the VPIN) shift systematically around a relevant documented LIBOR manipulation event. However, when I build cross sectional averages across events, I find no significant evidence of systematic shifts in either the PIN or VPIN metric. It should be noted that whilst I have included every documented case of manipulation directly linked to the relevant reference rates, my list is necessarily incomplete as the regulatory actions have tended to focus on sample charges to the firms involved, rather than documenting every occurrence and its motivation.

According to the studies of PIN in chapter 4, 5, and the VPIN in chapter 6, the results are very mixed; however, my conclusion lends support to the Andersen and Bondarenko [2014a,b,c] results. The VPIN is very sensitive to the choice on how to discretize the volume bucket and the number of volume in each bucket. In addition, the long run correlation between these VPINs is low with the average correlation coefficient of 0.32. Another remarkable result is the correlation between PIN and VPINs which is surprisingly low and extremely dismilar in value with the average correlation coefficient of 0.11. The PIN is constantly high with the average of 0.688 whereas the VPIN is constantly low with the average of 0.126. Form these results, I suggest that the VPIN cannot be an approximation of the PIN.

The contribution of this chapter is threefold. First, this is the only paper to provide a comprehensive analysis of informed trading in Eurodollar futures from 1996 to 2015. Second, to implement my analysis I have introduced several innovations in the estimation process, including an exact asymptotic representation of the measurement error of the PIN and a new bootstrap-in-bootstrap method for estimation of the VPIN. Finally, my methodological approach to decomposing informed trading by both event and the term-structure is a new contribution to this literature and my algorithms are available for other researchers to implement such studies in different market settings.

The remainder of this paper is organized as follows. First, §.6.2 outlines the debate on informed trading, in the academic literature in regard to the efficacy of the PIN and more recent VPIN algorithms. To this end, §.6.3 I look more deeply at the adaptation of the Glosten and Milgrom [1985] model of informed trading and demonstrate how a version of the classic version of the PIN could be used to empirically fit this model, in the spirit of Glosten and Milgrom [1985] and Easley et al. [1996]. In this section, I document the various empirical strategies to estimate the PIN and VPIN including several new features specific to this study relating to the distribution of these critical statistics. §.6.4 provides a detailed set of descriptive statistics of the comprehensive dataset of Eurodollar futures. Subsequently §.6.5 presents the analysis of my dataset. For each contract maturity, the PIN and VPIN is estimated and I construct experiments by aggregation across the terms structure, in addition to using the LIBOR manipulation events discussed previously. Finally, in §.6.6 presents the summaries of the key findings and some brief concluding comments and directions for future research.

## 6.2 The VPIN Debate and Further Background

The PIN model was first introduced in Easley and O'Hara [1992] and Easley et al. [1996] and estimated over relatively high frequency data for US equities. Subsequent work focused on the PIN as an asset pricing factor or related the analysis directly to corporate governance. The theoretical model underlying the PIN evolved from the information based trading literature that started with Grossman and Stiglitz [1980] and Hellwig [1980] then continued with Admati [1985], Hasbrouck [1991] and Kyle [1985]. However, the most direct theoretical antecedent to the PIN model is Glosten and Milgrom [1985] and this is the starting point for interpreting the impact of LIBOR manipulation on the Eurodollar futures market. In general the theoretical models of informed trading have traveled down the standard roots, Glosten and Milgrom [1985] and Kyle [1985] have used Bayesian games to analyse simple information flows in single assets, whereas Hellwig [1980], Admati [1985] and Admati and Pfleiderer [1988] have utilized a rational expectations framework for single and multiple assets.

However, regardless of the solution device applied, either rational expectations or a Bayes-Nash equilibrium, the general set-up is very similar. First, certain traders are provided with a forward looking information endowment (usually a noisy forecast of direction) and then there is a trading game against another group of traders who either lack this information or whose order flow is constrained. For instance, in the case of Eurodollar futures, we know that floating legs of IRSs are offset at or around international money market (IMM) dates, hence one particular institution may not have any choice in the timing and net quantity of futures demanded whether long or short. Hence a trader with these constraints is a net supplier of random order-flow to the market regardless of how the term structure is updated by the new fixings.

One of the unique selling points of the PIN model is the ease of implementation. The Glosten and Milgrom [1985] model is easily to re-cast in terms of net order flow and hence one merely needs to sign each trade as a buy or a sell to determine the net order-flow imbalance. The PIN is then computed via maximum-likelihood estimation. However, signing trades as a buy or a sell can be a difficult process. The standard Lee and Ready [1991] algorithm matches trades based on the distance between the transacted price and the nearest quotes. Several prior studies have identified difficulties in estimating the PIN attributing most of the problems to trade-classification, see Ke [2014];Yan and Zhang [2012];Boehmer et al. [2007]. It is evident that when switching to very high frequency data (such as that generated by futures trading) classifying trades becomes ever more problematic.

As a solution to the PIN classification problem, Easley et al. [2012] proposed a volume based approach known as VPIN, and explicitly introduced the notion of order-flow-toxicity. Order-flow-toxicity is defined as a systematic imbalance in the buy and sell side volume of the order-book that leads to substantial adjustments in spread. This is in opposition to the normal state of affairs where the liquidity would remain relatively constant. However, the proposed application to the May 2010 'flash crash' has proved somewhat controversial. In many respects the ease of implementing the VPIN and the unambiguous nature of its measurement of market data, forms part of the issue. Andersen and Bondarenko [2013] re-computed the VPIN metric for S&P 500 E-mini futures and refuted the applicability of the method to forecast extreme liquidity events. Andersen and Bondarenko [2013] concluded that the VPIN was highly sensitive to the assumptions placed on discretizing the volume of orders, reducing its efficacy as a regulatory instrument for predicting order-flow-toxicity exceed events.

In a series of follow up papers Andersen and Bondarenko [2014c], Easley et al. [2014] and again Andersen and Bondarenko [2014b] debated the relative merits of the VPIN approach. In another follow-up, Andersen and Bondarenko [2014a] proposed an alternative approach to specifically capture order-flow-toxicity. My experience with estimating VPIN falls somewhere between Andersen and Bondarenko [2014c] and Easley et al. [2014]. VPIN is indeed sensitive to the choice on how to discretize volume. However, I do find the change in VPIN, and I denote this $\Delta$VPIN which is consistent across a variety of choices. I do not find VPIN and the original PIN to be correlated under my trade-classification approach. Furthermore, I do not find that the VPIN lies within a 95% confidence bound computed from the maximum likelihood estimates for the underlying parameters. This indicates that for the purposes of identification of informed trading in Eurodollar futures markets, (which uses the same platform as the S&P 500 E-mini), the PIN and VPIN are, unfortunately, incongruent across a wide variety of specifications (including a bootstrapped version).

## 6.3 The model

### 6.3.1 The Market Algorithm

This section explains how the LIBOR manipulation affects the Eurodollar trading by deriving the fundamentals of sequential trading with superior information proposed by Glosten and Milgrom [1985] . This model focuses on stylized leadership market in which a Eurodollar price will reach at time $T$ with value $S_T = \{\underline{S}, \overline{S}\}$, where $\underline{S} \leq \overline{S}$. Uninformed traders $(\varepsilon)$ know the Eurodollar price will be $\underline{S} \leq S_t \leq \overline{S}$,however, they do not know either $S_t = \underline{S}$ or $S_t = \overline{S}$ will occur, so these traders are equally likely to buy or sell. Unlikely uninformed traders, informed traders $(\mu)$ are said to be informed of knowing the terminal value of the asset $S_T$. This sequential trade can be explained in terms of probability($\mathbb{P}$) which is the one that makes most sense. With theoretical explanations incorporated with ED trading, on each trading period traders arrive on the market sequentially in which knowing the value of ED futures equals $S_T$, suppose that $\mathbb{P}(S_T = \underline{S}) = \delta$ and $\mathbb{P}(S_T = \overline{S} = 1-\delta)$, which trade occurs at time $t \in [0, T)$. The probability pay off for each possible scenario of $S_T$ is, if the ED expected price is equal $S_T = \underline{S} \leq S_t$ informed traders will wait on the sell side with probability $\delta\mu$, as uninformed Traders is risk neutral, so they will locate in both buy and sell side with rate $\frac{1}{2}(1-\mu)\delta$. If the expected price equals $S_t \leq \underline{S} = S_T$ informed Traders will arrive at buy side with probability $(1-\delta)\mu$ and uninformed traders at rate $\frac{1}{2}(1-\mu)(1-\delta)$.

Whilst forward knowledge of the fix has many applications it is worth postulating a specific example of a trading strategy in ED futures around LIBOR announcements. Consider the period 2005 to 2007 from Figure 5.2. Here, the spread of submissions for the LIBOR fix are very tight. For a set of banks typically sitting in the middle eight submitters, a collaborative submission (as indicated by the cross bank emails noted previously) could allow for reasonable fore-knowledge of the fix in the second decimal, for instance an unexpected increase in the submission by one basis point or \$25 for each contract. The per day quote volume (in terms of numbers of contracts quoted at the inside best bid best ask) for the Eurodollar market is denoted in Figure 2.2 and it can be seen that contract volumes can exceed $7^{15}$ for a single day. Therefore a \$25 dollar a contract advantage has potential for substantial profits, if a suitable shock can be engineered against the current market expectations. It is worth noting that as a forward rate contract, the shock can be engineered from any LIBOR fixing, however, it is apparent that the size of the impact needs to be reasonably close to maturity contracts.

As the update on the reference rate is at 11:00.00 GMT time we can think of this as being the realization of a signal that will update the information set of all traders. Prior to the update the request to the submitter pool for a 'high' or a 'low' fixing equates to '$\delta$' and '$1-\delta$' respectively.

Shortly after 11:00:00 GMT , around 11:30:00 GMT or 06:30:00 central time (recalling that the ED future is traded around the clock) the LIBOR rate is announced, then information set for the futures are updated for all participants as a realization. However, in a smaller scale study on ED futures Kim et al. [2014] reports that the highest fraction of informed trading is located in the ED market around 7.00 to 10.00 CT which is after the announcement rather than in the lead up. I will show later that this is consistent with my macro-view argument that maturity and activity play a much more important role in the degree of informed trading and order toxicity within the market. After receiving the signal, informed derivatives traders wait to 'Buy' or 'Sell' regarding which type of information they have received from the submitter. If they arrive on buying side then $\mathbb{P}(Buy) = 1 + \mu(1-2\delta)/2$, while on the sell side $\mathbb{P}(Sell) = (1-\mu)(1-2\delta)/2$, hence $\mathbb{P}(Buy) + \mathbb{P}(Sell) = 1$. Subsequently, in each trading period the nature of the event occurs with, $\mathbb{P}(Buy) = \mathbb{P}(sell) = 0.5$ then $\delta_t = 0.5, \mathbb{P}(S_T = \underline{S})$ and $\mathbb{P}(S_T = \overline{S}) = 1 - \delta$ at $t = 0$.

In the period after the LIBOR is published, informed traders arrive with a degree of positive adverse selection that can be observed by increased spread. At this point, a market maker recognizes a price level at which he intends to enter, whether long position (Bid, B) or short position (Ask, A).Although the market maker doesn't know that other market participants at period $t$ are informed or not, he can update his belief about the value of $S_T$ as trades are revealed, so $\mathbb{P}(S_T = \underline{S}|Buy) = \mathbb{P}(S_T = \underline{S}, Buy)/\mathbb{P}(Buy) = \delta_t(1-\mu)/1 + \mu(1-2\delta)$. If $(S_T) = \underline{S} \leq S_t$ informed traders lower their expectation, since $E[S_T|Buy] = \underline{S}(1-\mu)\delta_t + \overline{S}(1+\mu)(1-\delta_t)/1 - \mu(1-2\delta_t)$ and increase expectation on $S_T$, since $E[S_T|Sell] = \underline{S}(1+\mu)\delta_t + \overline{S}(1-\mu)(1-\delta_t)/1 + \mu(1-2\delta_t)$. Therefore the spread is $A - B = 4(1-\delta_t)\delta_t\mu(\overline{S} - \underline{S})/1 - (1-2\delta_t)^2\mu^2$.

Conversely, in the case of $(S_T) = \underline{S} \leq S_t$ informed trading $(\mu)$ occurs when $A = E[S_T|Buy]$ with a profit is gained by direct wealth transfer from the loss incurred by uninformed trades $(\varepsilon)$. I can therefore write $(A - E[S_T|\mu, Sell])\mathbb{P}(\mu|Sell) = (A - E[S_T|\varepsilon, Sell])\mathbb{P}(\varepsilon|Sell)$, since by construction $\mathbb{P}(\mu|Sell) + \mathbb{P}(\varepsilon|Sell) = 1$. This argument can be made for B when $(S_T) = S_t \leq S_t$. Recall that the original interpretation of the Glosten and Milgrom [1985] model operates in a zero sum outcome. Definitively, a futures market can be seen in this sense as the settlement at marking to market, from the price at 13:59:00 and 14:00:00 CT (19.59.00

to 20:00.00 GMT), provides a definitive result on the days activities. Therefore, one explanation for the post fixing informed trading is that the update of the LIBOR curve itself generates new differentiated sets of information on $E(S_T)$ between informed traders and uninformed traders prior to the final settlement at 14:00:00.

### 6.3.2 The Asymptotic Standard Error of The PIN

The previous discussion indicates why a binomial high versus low fix model is very appropriate for this setting, indeed one could argue the relatively clean mechanism for discerning the days winners and losers makes this far more appropriate than the many other applications of this approach. The original PIN is a measure by Easley et al. [1996] of asymmetric information fitted via maximum likelihood estimation directly from measurements of the order flow. The tree diagram in Figure 6.1 presents the dynamics of the game and frames the original PIN type model in terms of the information structure of the ED futures market.

I restrict my interest and the estimation of the model to each contract individually. Whilst attempting to build a simultaneous equation model across all 40 contracts seems attractive there are substantial drawbacks in terms of numerical tractability. This analysis index discrete trading time by $i = 1, 2, \ldots, I$ and time is considered to be continuous within discrete trading block and is denoted by $t \in [0, T)$. Market participants buy or sell the asset at bid and ask prices posted on the limit order-book during the trading period. Information events are independently distributed at the beginning of each trading block, and occur with probability $\alpha$. However, informed traders will only trade when they perceive that their LIBOR fixers will be able to generate an information event; they will buy

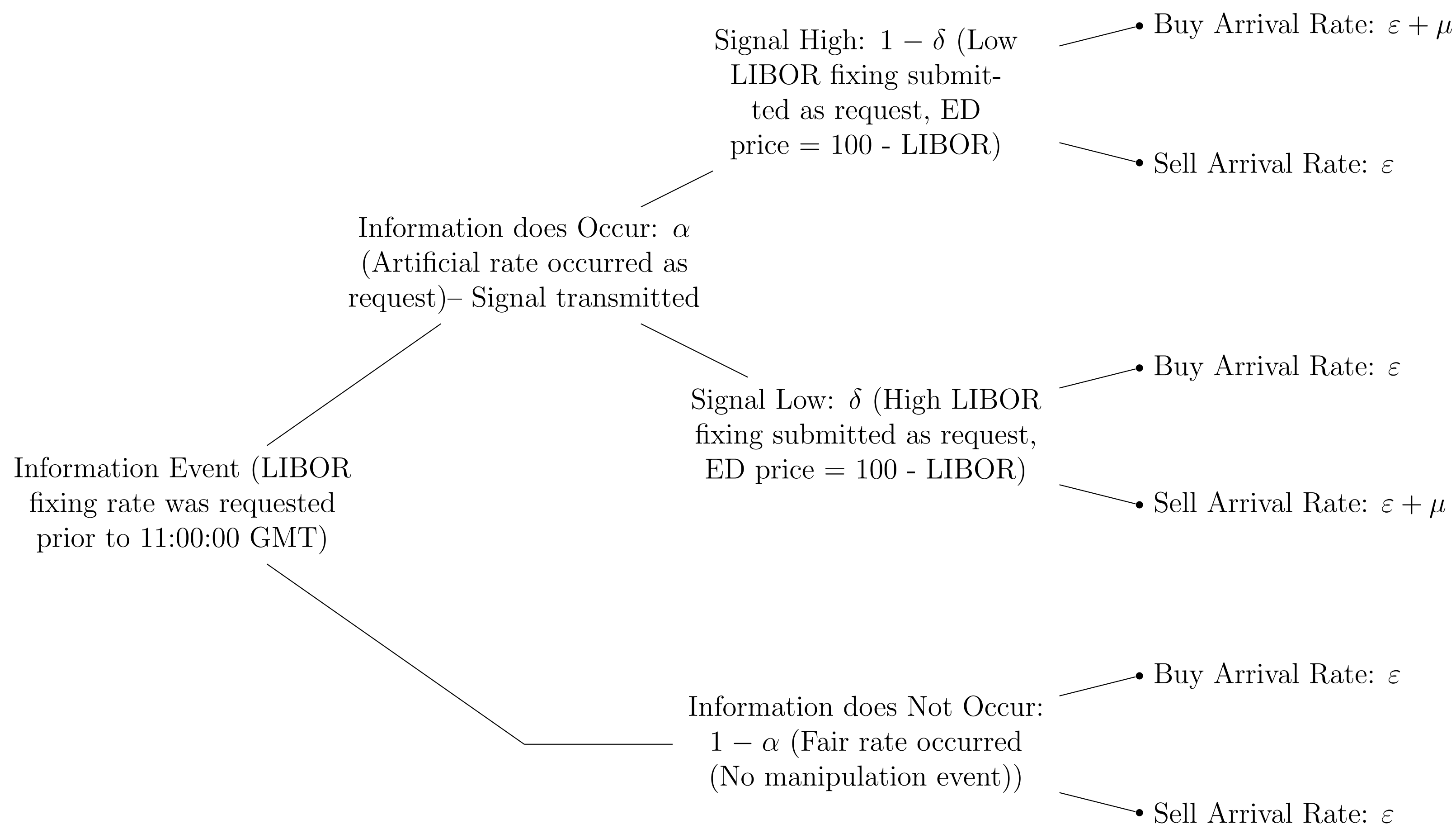



**Figure 6.1: The Eurodollar price and a tree diagram of the sequential trading progress**
Note: that after 11.00 GMT the LIBOR FIX is published, if the rate that occurred has been requested by derivatives traders, then the published data is similar to a signal with rate $\alpha$. When the fix is high, the price of the contract will decrease (ED price = 100 - LIBOR) and when the fix is low the price will be higher, if the bank can guarantee that its fix will impact on the LIBOR average determination.

an asset if they receive a signal of good news with probability $1-\delta$ and sell if they receive a signal of bad news with probability $\delta$. As $\varepsilon_b$ and $\varepsilon_s$ are the selling and buying rate of uninformed traders, who are supposed to have the same intensity (0.5), therefore the probability of uninformed traders is $\varepsilon_s = \varepsilon_b = \varepsilon$. In this way, the set of parameters is reduced to $\alpha, \delta, \mu$ and $\varepsilon$.

Within each trading block, order-flow arrives from both uninformed traders and informed traders subject to the receipt of their signal. Both uninformed buyers and sellers arrive in the market at rate $\varepsilon$. Competitive informed traders who are risk neutral will arrive when information events have occurred. If they receive a definitive signal of an up tick, they will arrive to buy orders; conversely, they will submit sell orders if they receive a bad signal or bad news. The arrival rate for this process is $\mu$. Following convention, the arrival processes are assumed to be independent. On good event days, the arrival rate for buy orders is $\varepsilon + \mu$ and for sell orders is $\varepsilon$. On the other hand, if there is any bad news or bad signal on any trading day, the arrival rate of sell orders is $\varepsilon + \mu$ and buy orders arrive at a rate of $\mu$. Finally, if there is no news or no signal on that day, only uninformed traders arrive for both buy and sell orders at arrival rate $\varepsilon$.

In each block, the news arrival contains one of three types of information. On the CME Globex market makers provide liquidity via the 'mass-quote' system whereby a single market marker can control a very large number of standing quotes within the various levels of the limit order book (effectively mimicking a larger group of individual traders). Following Easley et al. [1996] it can be presumed that the market maker knows that there is some probability attached to each branch and has some knowledge of the order arrival process for each branch. However, the market maker does not know which of the three branches has been selected. Since he cannot directly observe which type of branch has been selected he uses Bayesian updating from the observed order-flow to adjust his beliefs about the nature of the information events during a single trading block. Let $P(t) = (P_n(t), P_b(t), P_g(t))$ be a liquidity provider's belief about the occurrence of information event "no news" (n), "bad news" (b), and "good news" (g) at time $t$. Then, his prior beliefs at time 0 is $P(t_0) = (1 - \alpha, \alpha\delta, \alpha(1 - \delta))$ .

To determine quotes at time $t$, the market maker updates his prior belief on the condition of an arrival order of the relevant type. For instance, the bid at time $t$, $B(t)$, is the expected value of the asset conditional both on the history of the process prior to the arrival of order at time $t$ and on the fact that someone wants to sell the asset. Let $(s_t)$ denote the event that a sell order arrives at time $t$. Let $P_n(t|s_t)$ be the market maker's updated belief vector conditional on the history prior to time $t$ and on the event that a sell order arrives at time $t$.

Following Bayes' rule, the market maker's posterior probability on no news at time $t$, if an order to sell arrives at $t$, is $\mathbb{P}_n(t|s_t) = \mathbb{P}_n(t) \cdot \varepsilon/\varepsilon + \mathbb{P}_b(t)\mu$, the posterior probability on bad news is $P_b(t|s_t) = P_b(t) \cdot (\varepsilon + \mu)/\varepsilon + P_b(t)\mu$ and the posterior probability on good news is $\mathbb{P}_g(t|s_t) = \mathbb{P}_g(t) \cdot \varepsilon/\varepsilon + \mathbb{P}_b(t)\mu$. Comparable to the case of buy orders $(b_t)$ the posterior probability on good news at time $t$ is $\mathbb{P}_g(t|b_t) = \mathbb{P}_g(t) \cdot (\varepsilon + \mu)/\varepsilon + \mathbb{P}_g(t)\mu$ the posterior probability on bad news is $\mathbb{P}_b(t|b_t) = \mathbb{P}_b(t) \cdot \varepsilon/\varepsilon + \mathbb{P}_g(t)\mu$ and the posterior probability on no news is $\mathbb{P}_n(t|b_t) = \mathbb{P}_n(t) \cdot \varepsilon/\varepsilon + \mathbb{P}_g(t)\mu$. At the end of the trading on any day, the full information value of the asset is realized. If it is good news on trading day $(i)$ the informed trader knows that the value of the asset at the end of the day is worth $\bar{S}_i$, similarly it is $\underline{S}_i$ if it is bad news on day $i$ and the asset on day $i$ is worth $S_i^* = \delta\underline{S}_i + (1 + \delta)\bar{S}_i$, (so, $\bar{S}_i > S_i^* > \underline{S}_i$) if there is no news at all.

To complete the bid or ask at time $t$, the liquidity provider updates his position on the condition of arrival order according to the information type. At time $t$ the expected value of the asset, conditional on the history of trade prior to time $t$, is $\mathbb{E}[S_i|t] = \mathbb{P}_n(t)S^* + \mathbb{P}_b(t)\underline{S}_i + \mathbb{P}_g(t)\bar{S}_i$ where $S_i^* = \delta\underline{S}_i + (1+\delta)\bar{S}_i$ is the prior expected value of the asset. How might this work in practice on Globex? Well the market maker uses the mass quote system to place a large number of resting quotes at various levels within the order-book. If one side of the order-book receives a disproportionate number of trades then resting quotes on the opposite side (at level one of the limit order book) update and are canceled, and new quotes are instigated to restore balance and maintain the magnitude of the spread. New quotes deeper in the order book on the side with the draw down are instigated automatically. Whilst, in principle, this should ensure that price shifts are driven by supply and demand, the 'Flash Crash' of May 2010, originating on the Globex platform, indicates that this mechanism is not always perfectly functional. From this adjustment process the bid is the expected value of the asset conditional on someone wanting to sell the asset to a liquidity provider. Thus, the bid–ask spread at time $t$ is denoted by $\Sigma(t) = A(t) - B(t)$. This spread is $\Sigma(t) = \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}(\bar{S}_i - \mathbb{E}[S_i|t]) + \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}(\mathbb{E}[S_i|t] - \underline{S}_i)$. The first term in the bid–ask spread equation is a probability of buy order based on information and the second is the term for sells. The spread for the initial quotes in the period, $\Sigma$, has a particular simple form in the natural case in which good and bad events are equally likely. That is, if $\delta = 1 - \delta$, then $\Sigma = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}(\bar{S}_i - \underline{S}_i)$. The key component of this model is the probability that an order is from an informed trader, which is called the PIN $= \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}$, where $\alpha\mu + 2\varepsilon$ is the arrival rate for all orders and $\alpha\mu$ is the arrival rate for information-based orders. Therefore, the PIN is a measure of the fraction of orders that arise from informed traders relative to the overall order flow, and the spread equation shows that it is the key determinant of spreads. These equations clarify that liquidity providers need to correctly estimate the PIN in order to identify the optimal levels at which to enter the market. An unanticipated increase in PIN will result in losses to those liquidity providers who do not adjust their prices.

It is difficult to estimate the parameter vector $\theta = (\alpha, \delta, \varepsilon, \mu)$ because it cannot be directly observed in either the occurrence of information events or the associated arrival of uninformed and informed traders. In fact, in terms of measuring the daily arrival rate of sell $(s_t)$ and buy $(b_t)$ it is possible to infer these values using maximum likelihood and assuming that the trading process follows a Poisson process [Karyampas and Paiardini, 2011].

The Easley et al. [1996] PIN estimator considers the likelihood of order arrivals during a discrete trading block. In this model, the likelihood of observing sell $S$ and buy $B$ orders on each type of information occurs on a no event block with probability $1-\alpha$, a bad event block with probability $\alpha\delta$ and a good event block with probability $\alpha(1-\delta)$. Therefore, similar to the Eq. 3.0.4, the likelihood is

$$\begin{aligned} L_i(\Theta) =& (1-\alpha)e^{-\varepsilon_b}\frac{(\varepsilon_b)^{B_i}}{B_i!}e^{-\varepsilon_s}\frac{(\varepsilon_s)^{S_i}}{S_i!} \\ &+ \alpha\delta e^{-\varepsilon_b}\frac{(\varepsilon_b)^{B_i}}{B_i!}e^{-(\varepsilon_s+\mu)}\frac{(\varepsilon_s+\mu)^{S_i}}{S_i!} \\ &+ \alpha(1-\delta)e^{-(\varepsilon_b+\mu)}\frac{(\varepsilon_b+\mu)^{B_i}}{B_i!}e^{-\varepsilon_s}\frac{(\varepsilon_s)^{S_i}}{S_i!}\,. \end{aligned} \tag{6.3.1}$$

Since only one type of information occurs in each trading block, the maximum likelihood estimator of the information event parameters $\alpha$ and $\delta$ will be either 0 or 1. However, these parameters can be estimated from each block of buy and sells, assuming that the blocks are approximately independent. The likelihood of observing the data $M = (B_i, S_i)_{i=1}^{I}$ across $I$ trading block is just the product of the daily likelihood function: $L(M|\Theta) = \prod_{i=1}^{I} L(\Theta|B_i, S_i)$ , where $M = ((B_1, S_1), ..., (B_I, S_I))$ represents the data set for every 1, 5, and 10 minutes[1].

The PIN estimates are computed by maximizing the parameter vector $\theta$ from any data set $M$, which is normally taken to be from the daily trades and quotes. As a quantity estimated by maximum likelihood I can make use the 'delta method' to compute analytical confidence bounds. First, let the log-likelihood function of the PIN be given by:

$$\begin{aligned}\mathscr{L}[M|\Theta] = & \sum_{i=1}^{I} \log(L_i(\Theta)) \\ = & \sum_{i=1}^{I} \log\left(\frac{-\alpha(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i}+\alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i}-(\alpha-1)\varepsilon^{B_i+S_i}}{B_i!S_i!}\right) - (2\varepsilon+\mu). \\ . & \end{aligned} \tag{6.3.2}$$

The analytical gradient, $\nabla\mathscr{L}[M|\Theta] = \frac{\partial}{\partial\theta}\sum_{i=1}^{l} L_i(B_i, S_i|\Theta)$, of $\mathscr{L}[M|\Theta]$ is given by:

$$\nabla\mathscr{L}[M|\Theta] =$$
$$\sum_{i=1}^{l}\left(\begin{array}{c} \left(\alpha - \frac{\varepsilon^{B_i+S_i}}{(\delta-1)\varepsilon_i^S(\mu+\varepsilon)_i^B-\delta\varepsilon_i^B(\mu+\varepsilon)_i^S+\varepsilon^{B_i+S_i}}\right)^{-1} \\ \left(\frac{\varepsilon_i^S\left((\alpha-1)\varepsilon_i^B-\alpha(\mu+\varepsilon)_i^B\right)}{\alpha\left(\varepsilon_i^S(\mu+\varepsilon)_i^B-\varepsilon_i^B(\mu+\varepsilon)_i^S\right)}+\delta\right)^{-1} \\ \frac{\varepsilon_i^S\left(\alpha(\delta-1)(\mu+\varepsilon)_i^B(B_i\varepsilon+S_i(\mu+\varepsilon))+(\alpha-1)(B_i+S_i)\varepsilon_i^B(\mu+\varepsilon)\right)-\alpha\delta\varepsilon_i^B(\mu+\varepsilon)_i^S(B_i(\mu+\varepsilon)+S_i\varepsilon)}{\varepsilon(\mu+\varepsilon)\left(\alpha(\delta-1)\varepsilon_i^S(\mu+\varepsilon)_i^B-\alpha\delta\varepsilon_i^B(\mu+\varepsilon)_i^S+(\alpha-1)\varepsilon^{B_i+S_i}\right)} - 2 \\ \frac{\alpha\left(B_i(\delta-1)\varepsilon_i^S(\mu+\varepsilon)_i^B-\delta S_i\varepsilon_i^B(\mu+\varepsilon)_i^S\right)}{(\mu+\varepsilon)\left(\alpha(\delta-1)\varepsilon_i^S(\mu+\varepsilon)_i^B-\alpha\delta\varepsilon_i^B(\mu+\varepsilon)_i^S+(\alpha-1)\varepsilon^{B_i+S_i}\right)} \end{array}\right) \tag{6.3.3}$$

Let $\nabla^2\mathscr{L}[M|\Theta]$ be the Hessian matrix of second order derivatives such that

[1]The PIN's parameters is calculated by the maximum likelihood from the data set $M$ where the $M = ((B_1, S_1), ..., (B_I, S_I)$ is the data set for every 1, 5, and 10 minutes. Then the daily PIN's parameters $\alpha, \delta, \mu, \varepsilon$ is calculated by using the mean average.

$\nabla^2 \mathscr{L}[M|\Theta] = \frac{\partial^2}{\partial\theta\partial\Theta'} \sum_{i=1}^{I} L_i(B_i, S_i|\Theta)$. For the maximized likelihood, with parameter vector $\hat{\Theta}$ the Hessian (Fisher information matrix) is denoted by $\nabla^2 \mathscr{L}[M|\hat{\Theta}]$. Let $\nabla \mathrm{P}[\Theta] = \frac{\partial}{\partial\Theta} \mathrm{P}[\Theta]$ and is derived analytically by:

$$\nabla \mathrm{P}[\Theta] = \begin{pmatrix} \frac{\mu}{\alpha\mu+2\varepsilon} - \frac{\alpha\mu^2}{(\alpha\mu+2\varepsilon)^2} \\ 0 \\ -\frac{2\alpha\mu}{(\alpha\mu+2\varepsilon)^2} \\ \frac{\alpha}{\alpha\mu+2\varepsilon} - \frac{\alpha^2\mu}{(\alpha\mu+2\varepsilon)^2} \end{pmatrix} \tag{6.3.4}$$

at the maxima, asymptotically, the PIN estimator is derived from a standard binomial tree and as such the likelihood estimator will obey the standard central limit theorems. Subsequently, let $\bar{\Theta}$ be the true parameter estimates, from the standard Gaussian limit theorem the Cramér-Rao bound is given by $\sqrt{I}(\bar{\Theta} - \hat{\Theta}) \sim \mathcal{N}(0, \nabla^2 \mathscr{L}[M|\hat{\Theta}]^{-1})$. From the standard delta method, the variance of the PIN estimates will be given by:

$$\sqrt{I}(\mathrm{P}[\bar{\Theta}] - \mathrm{P}[\hat{\Theta}]) \to^{\mathrm{d}} \mathcal{N}(0, \nabla \mathrm{P}[\hat{\Theta}]' \nabla^2 \mathscr{L}[\mathrm{M}|\hat{\Theta}]^{-1} \nabla \mathrm{P}[\hat{\Theta}])$$

It is relatively straightforward to observe that the variance of the PIN estimates $\nabla \mathrm{P}[\hat{\Theta}]' \nabla^2 \mathscr{L}[\mathrm{M}|\hat{\Theta}]^{-1} \nabla \mathrm{P}[\hat{\Theta}]$ is easily specified as follows. Let $I\hat{\Omega} = \nabla^2 \mathscr{L}[M|\hat{\Theta}]^{-1}$, with elements $[\hat{\omega}_{ij}]$ where $i, j \in \{\alpha, \delta, \varepsilon, \mu\}$ be the covariance matrix of the physically estimated parameters, $\{\alpha, \delta, \varepsilon, \mu\}$ [1]. To test the significance of shifts in the PIN, I compute cumulative sums of the PIN estimates and use the standard central limit theorem to compute confidence intervals. I will now move on to look at the VPIN, which is computed directly from the moments of the order-flow so

[1] Whilst the point estimates of the PIN are commonly used across academic and practitioner implementations the exact identification of the asymptotic standard deviation has not been written down and it is useful for determining the significance of adjustments in the level of the PIN around specific events. The exact derivation was computed using the Mathematica mathematics processing tool.

---

I compute the error variance of VPIN via iid. bootstrap resampling.

**Proposition 1.** ***The standard error of the estimated PIN:*** *the asymptotic standard error of the PIN is given by*

$$\widehat{st.dev}(\mathrm{P}[\bar{\Theta}] - \mathrm{P}[\hat{\Theta}]) \rightarrow^{\mathrm{d}} \sqrt{\frac{4\hat{\alpha}\left(\hat{\mu}\left(2\hat{\varepsilon}(\hat{\varepsilon}-\hat{\alpha})\hat{\omega}_{\hat{\varepsilon}\hat{\mu}} + \hat{\alpha}\hat{\mu}\hat{\omega}_{\hat{\varepsilon}\hat{\varepsilon}} - 2\hat{\mu}\hat{\varepsilon}\hat{\omega}_{\hat{\delta}\hat{\varepsilon}}\right) + \hat{\alpha}\hat{\varepsilon}^2\hat{\omega}_{\hat{\mu}\hat{\mu}}\right) + 4\hat{\mu}^2\hat{\varepsilon}^2\hat{\omega}_{\hat{\alpha}\hat{\alpha}}}{(\hat{\alpha}\hat{\mu}+2\hat{\varepsilon})^4}} \tag{6.3.5}$$

*Proof.* Follows from the statements above and a detailed derivation is given in the following section. □

Traditional implementations of PIN (see for instance the Matlab implementation by Paolo Zagaglia on the Mathworks webpage and the SAS standard calculation) invoke an unconstrained optimization to compute the PIN. However, I take advantage of the analytical nature of the gradient function and make full use of a standard Newtonian approach, this is both faster and less prone to an erroneous exit. The matrix of second order derivatives allows us to compute an asymptotic standard error, which I can then use to construct my rolling z-tests to detect the significance of changes in PIN over time.

To compute the asymptotic variance of the PIN from the maximum likelihood estimate, I take advantage of the implicit asymptotic normality of the variance covariance matrix from the estimator and proceed using the standard Delta method approach.

Let $B_i$ and $S_i$ be the observed counts of buys and sells from the order-flow, using my trade classification mechanism. Unfortunately, I do not have a good way of measuring the error covariance matrix on $B_i$ and $S_i$ therefore I have to treat the estimates as non-stochastic in this instance. The log-likelihood of the $i$ observation of $B_i$ and $S_i$ from $i \in \{1, \ldots, I\}$, given my model setting is identical to the standard PIN formulation, with parameter vector $\Theta = [\alpha, \delta, \varepsilon, \mu]'$, is denoted by

$$\mathfrak{L}_i(\Theta) = \log\left(\frac{-\alpha(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} + \alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} - (\alpha-1)\varepsilon^{B_i+S_i}}{B_i!S_i!}\right) - 2\varepsilon \tag{6.3.6}$$

summing over the $I$ observations $\mathfrak{L}(\theta) = \sum_i \mathfrak{L}_i(\theta)$ and taking the first derivative yields the analytic gradient:

$$\nabla\mathfrak{L}(\Theta) = \sum_i \begin{bmatrix} \left(\alpha - \frac{\varepsilon^{B_i+S_i}}{(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} + \varepsilon^{B_i+S_i}}\right)^{-1} \\ \left(\frac{\varepsilon^{S_i}\left((\alpha-1)\varepsilon^{B_i} - \alpha(\mu+\varepsilon)^{B_i}\right)}{\alpha\left(\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \varepsilon^{B_i}(\mu+\varepsilon)^{S_i}\right)} + \delta\right)^{-1} \\ \frac{\varepsilon^{S_i}\left(\alpha(\delta-1)(\mu+\varepsilon)^{B_i}(B_i\varepsilon + S_i(\mu+\varepsilon)) + (\alpha-1)(B_i+S_i)\varepsilon^{B_i}(\mu+\varepsilon)\right) - \alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i}(B_i(\mu+\varepsilon) + S_i\varepsilon)}{\varepsilon(\mu+\varepsilon)\left(\alpha(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} + (\alpha-1)\varepsilon^{B_i+S_i}\right)} - 2 \\ \frac{\alpha\left(B_i(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \delta S_i\varepsilon^{B_i}(\mu+\varepsilon)^{S_i}\right)}{(\mu+\varepsilon)\left(\alpha(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} + (\alpha-1)\varepsilon^{B_i+S_i}\right)} \end{bmatrix} \tag{6.3.7}$$

with analytical second derivative $H = \nabla^2\mathbb{L}[M|\Theta]$, with elements $[h_{ij}]$ where

$i, j \in \{\alpha, \delta, \varepsilon, \mu\}$ the unique (upper triangular) elements are given as follows:

$$h_{\alpha\alpha} = \sum_i -\frac{\left((\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} + \varepsilon^{B_i+S_i}\right)^2}{\mathscr{A}^2} \tag{6.3.8}$$

$$h_{\alpha\delta} = \sum_i \frac{\varepsilon^{B_i+S_i}\left(\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} - \varepsilon^{S_i}(\mu+\varepsilon)^{B_i}\right)}{\mathscr{A}^2} \tag{6.3.9}$$

$$h_{\delta\varepsilon} = \sum_i \frac{\mu\mathscr{D}\varepsilon^{B_i+S_i-1}}{\mathscr{A}^2(\mu+\varepsilon)} \tag{6.3.10}$$

$$h_{\varepsilon\mu} = \sum_i \frac{\mathscr{D}\varepsilon^{B_i+S_i}}{\mathscr{A}^2(\mu+\varepsilon)} \tag{6.3.11}$$

$$h_{\delta\delta} = \sum_i -\frac{\alpha^2\left(\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \varepsilon^{B_i}(\mu+\varepsilon)^{S_i}\right)^2}{\mathscr{A}^2} \tag{6.3.12}$$

$$h_{\delta\varepsilon} = \sum_i \frac{\alpha\mu(\mathscr{E}\mathscr{G})\varepsilon^{B_i+S_i-1}}{\mathscr{A}^2(\mu+\varepsilon)} \tag{6.3.13}$$

$$h_{\delta\mu} = \sum_i -\frac{(\mathscr{E}\mathscr{G})\varepsilon^{B_i+S_i}}{\mathscr{A}^2(\mu+\varepsilon)} \tag{6.3.14}$$

$$h_{\varepsilon\varepsilon} = \sum_i \frac{\mathscr{A}\mathscr{B} - \frac{\mathscr{F}^2}{\varepsilon^2(\mu+\varepsilon)^2}}{\mathscr{A}^2} \tag{6.3.15}$$

$$h_{\varepsilon\mu} = \sum_i \frac{\mathscr{C}}{\mathscr{A}^2\varepsilon(\mu+\varepsilon)^2} \tag{6.3.16}$$

$$h_{\mu\mu} = \sum_i \frac{\frac{(-\alpha)\mathscr{A}\mathscr{D}}{(\mu+\varepsilon)^2} - \left(\alpha B_i(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i-1} - \alpha\delta S_i\varepsilon^{B_i}(\mu+\varepsilon)^{S_i-1}\right)^2}{\mathscr{A}^2} \tag{6.3.17}$$

where the functions $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, $\mathscr{D}$, $\mathscr{E}$, $\mathscr{F}$ and $\mathscr{G}$ are defined as follows:

$$\mathscr{A} = \alpha(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} + (\alpha-1)\varepsilon^{B_i+S_i} \tag{6.3.18}$$

$$\begin{aligned}\mathscr{B} =& \alpha(B_i-1)B_i\delta\varepsilon^{B_i-2}(\mu+\varepsilon)^{S_i} + 2\alpha B_i\delta S_i\varepsilon^{B_i-1}(\mu+\varepsilon)^{S_i-1} \\ &- \alpha(\delta-1)(S_i-1)S_i\varepsilon^{S_i-2}(\mu+\varepsilon)^{B_i} - 2\alpha B_i(\delta-1)S_i\varepsilon^{S_i-1}(\mu+\varepsilon)^{B_i-1} \\ &- \alpha(B_i-1)B_i(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i-2} \\ &+ \alpha\delta(S_i-1)S_i\varepsilon^{B_i}(\mu+\varepsilon)^{S_i-2} - (\alpha-1)(B_i+S_i-1)(B_i+S_i)\varepsilon^{B_i+S_i-2}\end{aligned} \tag{6.3.19}$$

$$\mathscr{C} = \alpha(\mathscr{C}_a + \mathscr{C}_b + \mathscr{C}_c) \tag{6.3.20}$$

$$\mathscr{C}_a = \delta S_i\varepsilon^{B_i}(\mu+\varepsilon)^{S_i}\left(\varepsilon^{S_i}(\mu S_i+\varepsilon)\left(\alpha(\delta-1)(\mu+\varepsilon)^{B_i} + (\alpha-1)\varepsilon^{B_i}\right) - \alpha\delta\varepsilon^{B_i+1}(\mu+\varepsilon)^{S_i}\right) \tag{6.3.21}$$

$$\mathscr{C}_b = B_i(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i}\left(\alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i}(\varepsilon-2\mu S_i) + \varepsilon^{S_i+1}\left(-\alpha(\delta-1)(\mu+\varepsilon)^{B_i} - (\alpha-1)\varepsilon^{B_i}\right)\right) \tag{6.3.22}$$

$$\mathscr{C}_c = -{B_i}^2(\delta-1)\mu\varepsilon^{B_i+S_i}(\mu+\varepsilon)^{B_i}\left((\alpha-1)\varepsilon^{S_i} - \alpha\delta(\mu+\varepsilon)^{S_i}\right) \tag{6.3.23}$$

$$\mathscr{D} = B_i(\delta-1)\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} - \delta S_i\varepsilon^{B_i}(\mu+\varepsilon)^{S_i} \tag{6.3.24}$$

$$\mathscr{E} = (\alpha-1)S_i\varepsilon^{B_i} + \alpha(B_i-S_i)(\mu+\varepsilon)^{B_i} \tag{6.3.25}$$

$$\begin{aligned}\mathscr{F} =& \alpha\delta\varepsilon^{B_i}(\mu+\varepsilon)^{S_i}(B_i(\mu+\varepsilon)+S_i\varepsilon) \\ &- \varepsilon^{S_i}\left(\alpha(\delta-1)(\mu+\varepsilon)^{B_i}(B_i\varepsilon+S_i(\mu+\varepsilon)) + (\alpha-1)(B_i+S_i)\varepsilon^{B_i}(\mu+\varepsilon)\right)\end{aligned} \tag{6.3.26}$$

$$\mathscr{G} = (\mu+\varepsilon)^{S_i} - (\alpha-1)B_i\varepsilon^{S_i}(\mu+\varepsilon)^{B_i} \tag{6.3.27}$$

setting $\hat{\Theta} = [\hat{\alpha}, \hat{\delta}, \hat{\varepsilon}, \hat{\mu}]'$, to the be the numerically evaluated parameters such that $\nabla\mathfrak{L}(\hat{\Theta}) = \mathbf{0}_4$. Where $\mathbf{0}_4$ is a $4 \times 1$ vector of zeros. Under mild regularity conditions the maximum of the function $\mathfrak{L}(\Theta)$ can be quickly computed via the Newton-Raphson approach.

The estimate of the asymptotic covariance matrix $\hat{\Omega} = \text{cov}[\bar{\Theta} - \hat{\Theta}]$, where $\bar{\Theta}$ is the true parameter vector is computed from the Hessian matrix following Fisher's theorem by $I\hat{\Omega} = H^{-1}$.

#### 6.3.2.1 The Asymptotic Estimate Variance of PIN

From the theoretical composition the formulation of the estimated PIN is $\text{P}[\Theta] = \frac{\alpha\mu}{\alpha\mu+2\varepsilon}$. Inserting my estimated values yields $\text{P}[\hat{\Theta}] = \frac{\hat{\alpha}\hat{\mu}}{\hat{\alpha}\hat{\mu}+2\hat{\varepsilon}}$. My presumption is that $\sqrt{I}(\text{P}[\bar{\Theta}] - \text{P}[\hat{\Theta}]) \rightarrow^{\text{d}} \mathcal{N}(0, \varsigma^2)$, where $\varsigma^2$ is an asymptotic variance. I can confirm this presumption relatively easily, first by confirming that the gradient $\nabla \mathrm{P}[\Theta]$, is a smooth function of the underlying parameters:

$$\nabla \mathrm{P}[\Theta] = \begin{pmatrix} \frac{\mu}{\alpha\mu+2\varepsilon} - \frac{\alpha\mu^2}{(\alpha\mu+2\varepsilon)^2} \\ 0 \\ -\frac{2\alpha\mu}{(\alpha\mu+2\varepsilon)^2} \\ \frac{\alpha}{\alpha\mu+2\varepsilon} - \frac{\alpha^2\mu}{(\alpha\mu+2\varepsilon)^2} \end{pmatrix} \tag{6.3.28}$$

and second I can derive the exact formulation via the Delta method: $\varsigma^2 = \nabla \mathrm{P}[\hat{\Theta}]'\hat{\Omega}\nabla \mathrm{P}[\hat{\Theta}]$. Substitution and rearrangement yields $\sqrt{I}(\mathrm{P}[\bar{\Theta}] - \mathrm{P}[\hat{\Theta}]) \to^{\mathrm{d}} \mathcal{N}(0, \nabla \mathrm{P}[\hat{\Theta}]'\nabla^2 \mathbb{L}[\mathrm{M}|\hat{\Theta}]^{-1}\nabla \mathrm{P}[\hat{\Theta}])$. Evaluating the term $\nabla \mathrm{P}[\hat{\Theta}]'\nabla^2 \mathbb{L}[\mathrm{M}|\hat{\Theta}]^{-1}\nabla \mathrm{P}[\hat{\Theta}]$, dividing by $I$ and taking the square root, yields the standard error in the proposition 1

### 6.3.3 The implementation of Volume-Synchronized Probability of Informed Trading (VPIN)

This work uses a variety of choices for implementing VPIN with arrival rate of informed and uninformed traders in Realized Volatility(RV) style. For original work Easley et al. [2011], the fixed number of volume buckets VPIN(n) is fixed at 50, however I consider $n = 20, 50, 100, 200$ and compute the bucket size $(V_i)$ as a fraction of daily trading volume to avoid the bias of activity or inactivity of trading period. So, the VPIN is calculated with $V_i = 1/20, 1/50, 1/100$, and $1/200$. My choices for complimenting VPIN not only calculate with different $n$ and $V_i$ but also using tick-by-tick data instead of one-minute interval as in the original work. I work with tick-by-tick data because I want to implement this approach with the real high frequency world, trading at the speed of light or in a fraction of millisecond, to investigate how the VPIN performs when it is applied with HFTs data.

Prior implementations of VPIN have utilized it as a separate pricing factor within a standard Fama and Macbeth style cross sectional asset pricing model. Given that the purpose of this chapter is to look specifically at the time series variation in the PIN and VPIN measures around well understood events in the life-cycle of the contract then I need to deal specifically with identification of the variation in the measures. For the PIN I can appeal to the asymptotic properties of the maximum likelihood estimator to build a standard error; however the suggested calculation of the VPIN does not lend itself to direct derivation of the input parameters Cramér-Rao bound and hence a standard error. I therefore choose a double block bootstrap to compute the confidence intervals. It should be noted that consistency of a bootstrap in this case is effectively a tautology, in that I impose consistency on the estimator as each individual step is known to be consistent, but the actual order-flow imbalance does not have a structural meaning from deep model parameters (although one could reconstruct a model if necessary).

It can be seen that from the delta method approach used to compute the error variance for the PIN it is quite possible that the likely data generating process (from the original PIN type model) results in a VPIN that is not a pivotal statistic (i.e. the distribution of the VPIN statistic is dependent on the true parameters of the model); as such I iid the bootstrap in both the major steps in the calculation, the construction of the volume buckets and the calculation of the resulting VPIN, to compute the confidence intervals for the point estimates.

My suggested two stage nested bootstrap requires 9,801 resampled proceeds as follows:

**Step 1:** Bootstrap across buckets. This is the outer loop. Let $V$, $P$, $P^A$ and $P^B$ be the matrices of volumes, transacted price and nearest preceding bid and ask quote prices over a day with $V^\dagger$, $P^\dagger$, $P^{A\dagger}$ and $P^{B\dagger}$ being a draw from their equivalent i.i.d. bootstrap with replacement counterparts. For my purposes I assume that the rows are jointly bootstrapped such that the resorted rows for $V^{**}$, $P^\dagger$, $P^{A\dagger}$ and $P^{B\dagger}$ are all identical. As the trade classifications for matched sets of the bootstrapped matrices is identical. The calculation in Equation 3.0.25 proceeds exactly, with $V_i^\dagger$ and $P_i^\dagger$ replacing $V_i$ and $P_i$.

**Step 2:** For each $V_i^\dagger$ and $P_i^\dagger$ compute the bootstrapped $V_\tau^{B\dagger}$ and $V_\tau^{S\dagger}$ and construct the collection of bootstrap order imbalances $OI^\dagger = [V_\tau^{S**} - V_\tau^{B\dagger}]_{\tau \in \{1,\ldots,N\}}$.

**Step 3:** I now add an interior loop. For each $OI^\dagger = [V_\tau^{S\dagger} - V_\tau^{B\dagger}]_{\tau \in \{1,\ldots,N\}}$ iid resample with replacement 99 times $OI^\dagger$, with each bootstrap resample labelled $OI^\ddagger$ and compute $VPIN^\ddagger$. Retain the median $VPIN^\ddagger$ and return this as $VPIN^\dagger$.

**Step 4:** Sort the collection of 99 outer loop resampled median $VPIN^\ddagger$ and then collect the desired confidence intervals (in my case I choose a 95% confidence interval).

These confidence bounds are applied in the subsequent plots for VPIN presented in subsequent sections and the cumulative error is computed for the confidence bounds in the tabulated VPIN adjustments.

## 6.4 Data Sources and Data Pre-processing

The data used in this research are introduced in this section, together with a preliminary processing of Eurodollar Futures data. As previously stated, Eurodollar futures contracts are traded on the CME's Globex platform and CME pit (open-outcry) trades. Data for both electronic and open-outcry are directly obtained from the CME tapes for the period January 1, 1996 to January 1, 2014. Pit trades are quotes from CME by the ED code and GE for the Globex code; the amalgamated tapes are classified under the ED moniker. The volume ratio between the Globex and open-outcry is between three and four orders of magnitude over the sample, so separating the pit trades from the electronic trades currently appears to be less interesting. The CME tapes data for the 44 Eurodollar futures contracts is available from the Thomson Reuter database via the Tick History system. I have conducted an analysis on the 4 monthly contracts, however the data is very sparse and the volumes are very small (up to 5 orders of magnitude for busy versus busy days) compared to the 40 quarterly contracts, so for brevity this is excluded from the results.

### 6.4.1 Eurodollar futures descriptive statistics

The data records for 40 quarterly Eurodollar future contracts is separated into two different time-frames, the first of which was from January 1, 1996 to July 31, 2007, and the second, from August 1, 2007 to January 1, 2014. The descriptive statistics for each ED futures contract in the two periods are shown in Tables 4.2 and 4.3.

In terms of the first period, the highest mean bid/ask spread belongs to EDU1, and EDZ0 has the highest mean trades returns. The smallest mean bid–ask spread belongs to EDH6, and EDU6 has the lowest mean trade returns. As for the second period, EDM7 presents both the largest mean spread and mean returns, while EDZ9 shows the lowest mean spread, EDZ9 the lowest mean spread, and EDM9 the lowest mean return (see Table 4.2 and 4.3).

#### 6.4.1.1 The descriptive of the term-structure of PIN and PIN's parameters

Table 6.1 presents an estimated PIN and Tables 6.2, 6.3, 6.4, 6.5 present PIN's parameters with its standard deviation for ten to one year to maturity on the period of 1996 until 2013. The results in this table show that the probability of information event between 1996 and 2001 is high, then begins to drop around 2002; the highest of $\alpha$ is in 2000 at 0.968 and 0.933 for ten and one year to maturity, respectively. This trend may be consistent with the PIN as both parameters have a similar trend, however, the highest PIN appears in 1997 with 0.844 and 0.830 for ten and one year to maturity, respectively. During the same period the arrival rate of informed traders and uninformed traders $\mu$, $\varepsilon$ fluctuates mostly between 0.4 and 0.6, except for 2000 when $\varepsilon$ fluctuates around 0.2. After an upward trend of information based trading for Eurodollar trading in the first period, the downward trend appears between 2002 and 2003. This trend can be investigated from the decreases of $\alpha$ and PIN. After 2003, the PIN and the $\alpha$ are increasing, then turning to a downward trend around 2007. The highest $\alpha$ appears in 2007 at 0.945 for ten year to maturity, however, the PIN is just about 0.631, which is lower than 2004, 2005 and 2006 for the same ten year to maturity. From these results, focusing on PIN and $\alpha$, this analysis finds that the PIN is not purely driven by $\alpha$ or the probability of information that occurs on the trading day. The

**Table 6.1: The Term Structure of PIN Coefficient**
**Time To Maturity**

| Year | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1996** | 0.8098 | 0.7647 | 0.6914 | 0.7150 | 0.7153 | 0.6993 | 0.7623 | 0.7713 | 0.8092 | 0.8170 |
| | (0.2788) | (0.2848) | (0.2887) | (0.3087) | (0.3015) | (0.3263) | (0.3099) | (0.3043) | (0.2780) | (0.2772) |
| **1997** | 0.8444 | 0.8103 | 0.7483 | 0.6866 | 0.6975 | 0.7456 | 0.7374 | 0.7836 | 0.8056 | 0.8300 |
| | (0.2517) | (0.2577) | (0.2728) | (0.2718) | (0.2525) | (0.2917) | (0.3073) | (0.2879) | (0.2716) | (0.2646) |
| **1998** | 0.8065 | 0.8068 | 0.7802 | 0.7366 | 0.7154 | 0.6922 | 0.7360 | 0.6876 | 0.7278 | 0.8073 |
| | (0.2724) | (0.2699) | (0.2813) | (0.2694) | (0.2653) | (0.2787) | (0.2425) | (0.3036) | (0.2999) | (0.2717) |
| **1999** | 0.7276 | 0.7817 | 0.8105 | 0.7850 | 0.7607 | 0.7632 | 0.6647 | 0.7087 | 0.6933 | 0.7109 |
| | (0.3015) | (0.3071) | (0.2893) | (0.2922) | (0.3121) | (0.2721) | (0.1925) | (0.2421) | (0.2554) | (0.2823) |
| **2000** | 0.6997 | 0.7331 | 0.7671 | 0.7560 | 0.7698 | 0.7546 | 0.7360 | 0.6667 | 0.6778 | 0.6757 |
| | (0.2434) | (0.2447) | (0.2464) | (0.2546) | (0.2731) | (0.2654) | (0.2770) | (0.2073) | (0.2194) | (0.2310) |
| **2001** | 0.6388 | 0.6522 | 0.6796 | 0.7133 | 0.7203 | 0.7458 | 0.7090 | 0.7106 | 0.6832 | 0.6962 |
| | (0.2132) | (0.2229) | (0.2537) | (0.2747) | (0.2759) | (0.2829) | (0.2502) | (0.2578) | (0.2259) | (0.2019) |
| **2002** | 0.6381 | 0.6073 | 0.6251 | 0.6698 | 0.6834 | 0.7152 | 0.7634 | 0.7364 | 0.7050 | 0.6215 |
| | (0.1838) | (0.2027) | (0.2351) | (0.2567) | (0.2875) | (0.2731) | (0.2852) | (0.2538) | (0.2418) | (0.1799) |
| **2003** | 0.6167 | 0.6161 | 0.5717 | 0.5974 | 0.5957 | 0.6568 | 0.6989 | 0.7195 | 0.7078 | 0.6927 |
| | (0.1817) | (0.1629) | (0.1830) | (0.2061) | (0.2002) | (0.3044) | (0.2989) | (0.2949) | (0.2601) | (0.1959) |
| **2004** | 0.7098 | 0.6468 | 0.6279 | 0.5763 | 0.5526 | 0.5789 | 0.6672 | 0.7043 | 0.7282 | 0.7128 |
| | (0.1951) | (0.1662) | (0.1625) | (0.2106) | (0.2160) | (0.2618) | (0.3165) | (0.3137) | (0.3069) | (0.2752) |
| **2005** | 0.7122 | 0.7101 | 0.6849 | 0.6645 | 0.6748 | 0.5695 | 0.5816 | 0.6060 | 0.6807 | 0.7503 |
| | (0.2551) | (0.2464) | (0.1237) | (0.1463) | (0.1658) | (0.2214) | (0.2352) | (0.2856) | (0.2944) | (0.2958) |
| **2006** | 0.7046 | 0.6888 | 0.6922 | 0.6922 | 0.6852 | 0.6705 | 0.6431 | 0.5489 | 0.5885 | 0.6941 |
| | (0.3113) | (0.2662) | (0.2590) | (0.1300) | (0.1345) | (0.1247) | (0.1588) | (0.2764) | (0.2989) | (0.3212) |
| **2007** | 0.6314 | 0.6687 | 0.6797 | 0.7076 | 0.6874 | 0.6798 | 0.6627 | 0.6588 | 0.6383 | 0.5902 |
| | (0.2956) | (0.3046) | (0.3022) | (0.2549) | (0.1736) | (0.1300) | (0.1297) | (0.1001) | (0.1703) | (0.2684) |
| **2008** | 0.6338 | 0.6585 | 0.6612 | 0.7146 | 0.7142 | 0.6843 | 0.6794 | 0.6721 | 0.6572 | 0.6289 |
| | (0.1935) | (0.2315) | (0.2903) | (0.2832) | (0.2702) | (0.2283) | (0.1043) | (0.1254) | (0.1352) | (0.1870) |
| **2009** | 0.6692 | 0.6450 | 0.6496 | 0.6510 | 0.6975 | 0.6924 | 0.6960 | 0.6840 | 0.6696 | 0.6584 |
| | (0.1006) | (0.1502) | (0.1805) | (0.2376) | (0.2376) | (0.2342) | (0.2073) | (0.1017) | (0.1210) | (0.1311) |
| **2010** | 0.6782 | 0.6685 | 0.6728 | 0.6480 | 0.6599 | 0.6653 | 0.6960 | 0.7054 | 0.6845 | 0.6869 |
| | (0.1149) | (0.1226) | (0.1025) | (0.1327) | (0.1275) | (0.1631) | (0.2048) | (0.1713) | (0.1281) | (0.1163) |
| **2011** | 0.6884 | 0.6828 | 0.6839 | 0.6806 | 0.6634 | 0.6340 | 0.6245 | 0.6146 | 0.6731 | 0.6814 |
| | (0.1048) | (0.1214) | (0.1164) | (0.0979) | (0.1212) | (0.1998) | (0.2392) | (0.2971) | (0.2308) | (0.1194) |
| **2012** | 0.6976 | 0.6961 | 0.6776 | 0.6845 | 0.6676 | 0.6721 | 0.6552 | 0.6520 | 0.6293 | 0.7058 |
| | (0.1190) | (0.1153) | (0.1101) | (0.0987) | (0.1206) | (0.1282) | (0.1356) | (0.1747) | (0.2129) | (0.1735) |
| **2013** | 0.6866 | 0.6876 | 0.6854 | 0.6784 | 0.6825 | 0.6705 | 0.6811 | 0.6593 | 0.6537 | 0.6341 |
| | (0.1401) | (0.1379) | (0.1041) | (0.0853) | (0.1086) | (0.1200) | (0.0984) | (0.1177) | (0.1356) | (0.1788) |

Note: This table present the term structure of PIN between 1996 until 2013.

**Table 6.2: The Term Structure of $\alpha$**
**Time To Maturity**

| Year | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1996** | 0.6862 | 0.4397 | 0.3694 | 0.6110 | 0.5798 | 0.8135 | 0.7971 | 0.8427 | 0.8444 | 0.8241 |
| | (0.4172) | (0.4599) | (0.4425) | (0.4523) | (0.4530) | (0.3238) | (0.3365) | (0.3084) | (0.3012) | (0.3215) |
| **1997** | 0.8397 | 0.6715 | 0.4508 | 0.3725 | 0.3328 | 0.5229 | 0.7255 | 0.7540 | 0.8000 | 0.8392 |
| | (0.3043) | (0.4132) | (0.4367) | (0.4275) | (0.4167) | (0.4702) | (0.3800) | (0.3791) | (0.3468) | (0.3130) |
| **1998** | 0.8959 | 0.8844 | 0.7302 | 0.4949 | 0.5258 | 0.5335 | 0.3227 | 0.8212 | 0.8005 | 0.8366 |
| | (0.2522) | (0.2619) | (0.3890) | (0.4274) | (0.4232) | (0.3774) | (0.4406) | (0.3408) | (0.3570) | (0.3305) |
| **1999** | 0.9298 | 0.9261 | 0.9233 | 0.8853 | 0.8506 | 0.6524 | 0.4157 | 0.3710 | 0.5723 | 0.9254 |
| | (0.2191) | (0.2156) | (0.2249) | (0.2477) | (0.3224) | (0.4398) | (0.3204) | (0.4701) | (0.4867) | (0.2356) |
| **2000** | 0.9541 | 0.9620 | 0.9612 | 0.9594 | 0.9556 | 0.8109 | 0.6672 | 0.3857 | 0.4373 | 0.9333 |
| | (0.1925) | (0.1614) | (0.1583) | (0.1481) | (0.1587) | (0.3394) | (0.4331) | (0.3629) | (0.4262) | (0.2304) |
| **2001** | 0.9683 | 0.9851 | 0.9643 | 0.9645 | 0.9612 | 0.9740 | 0.8149 | 0.6860 | 0.6992 | 0.5722 |
| | (0.1575) | (0.0972) | (0.1633) | (0.1554) | (0.1622) | (0.1090) | (0.2865) | (0.4060) | (0.3771) | (0.4806) |
| **2002** | 0.6755 | 0.9828 | 0.9677 | 0.9571 | 0.9569 | 0.9419 | 0.9209 | 0.8182 | 0.7052 | 0.5992 |
| | (0.3670) | (0.1190) | (0.1621) | (0.1746) | (0.1713) | (0.1761) | (0.2088) | (0.2702) | (0.3065) | (0.3175) |
| **2003** | 0.5988 | 0.7065 | 0.9769 | 0.9909 | 0.9904 | 0.9624 | 0.9572 | 0.9390 | 0.7771 | 0.6109 |
| | (0.3169) | (0.2925) | (0.1379) | (0.0848) | (0.0767) | (0.1627) | (0.1533) | (0.1779) | (0.2984) | (0.2919) |
| **2004** | 0.5980 | 0.5023 | 0.5978 | 0.7392 | 0.8601 | 0.9028 | 0.9586 | 0.9374 | 0.9291 | 0.7428 |
| | (0.2987) | (0.2621) | (0.2893) | (0.3532) | (0.2875) | (0.2731) | (0.1633) | (0.1996) | (0.2053) | (0.3145) |
| **2005** | 0.7553 | 0.7535 | 0.4468 | 0.4667 | 0.4357 | 0.6762 | 0.8120 | 0.9315 | 0.9634 | 0.9427 |
| | (0.3012) | (0.2942) | (0.1807) | (0.2188) | (0.2498) | (0.3245) | (0.3071) | (0.2232) | (0.1441) | (0.1769) |
| **2006** | 0.9342 | 0.7797 | 0.7320 | 0.4393 | 0.4435 | 0.4517 | 0.4578 | 0.6563 | 0.7772 | 0.9482 |
| | (0.1766) | (0.3052) | (0.3152) | (0.1761) | (0.1789) | (0.1855) | (0.1988) | (0.3587) | (0.3337) | (0.1631) |
| **2007** | 0.9457 | 0.9563 | 0.9019 | 0.7725 | 0.5291 | 0.4455 | 0.4720 | 0.4978 | 0.5703 | 0.8735 |
| | (0.1843) | (0.1595) | (0.2247) | (0.3003) | (0.2530) | (0.1837) | (0.1658) | (0.1389) | (0.2200) | (0.2680) |
| **2008** | 0.6584 | 0.6865 | 0.9094 | 0.9414 | 0.8248 | 0.6707 | 0.4517 | 0.4701 | 0.4811 | 0.4864 |
| | (0.2726) | (0.3322) | (0.2159) | (0.1759) | (0.2727) | (0.2991) | (0.1622) | (0.1836) | (0.1698) | (0.2538) |
| **2009** | 0.4845 | 0.4891 | 0.5514 | 0.6466 | 0.6862 | 0.6782 | 0.6019 | 0.4497 | 0.4627 | 0.4806 |
| | (0.1628) | (0.1991) | (0.2519) | (0.3087) | (0.3077) | (0.3016) | (0.2790) | (0.1613) | (0.1515) | (0.1536) |
| **2010** | 0.4397 | 0.4598 | 0.4721 | 0.4822 | 0.4741 | 0.4311 | 0.3419 | 0.4302 | 0.4533 | 0.4363 |
| | (0.1827) | (0.1814) | (0.1708) | (0.1773) | (0.1885) | (0.2395) | (0.3009) | (0.2676) | (0.1695) | (0.1780) |
| **2011** | 0.4585 | 0.4399 | 0.4397 | 0.4453 | 0.4638 | 0.4619 | 0.4168 | 0.4388 | 0.4217 | 0.4484 |
| | (0.1481) | (0.1749) | (0.1829) | (0.1781) | (0.1924) | (0.2192) | (0.2958) | (0.3440) | (0.3001) | (0.1528) |
| **2012** | 0.4462 | 0.4565 | 0.4628 | 0.4675 | 0.4564 | 0.4290 | 0.4592 | 0.4096 | 0.4162 | 0.4631 |
| | (0.1588) | (0.1410) | (0.1403) | (0.1315) | (0.1445) | (0.1882) | (0.1667) | (0.2393) | (0.2548) | (0.2430) |
| **2013** | 0.4589 | 0.4524 | 0.4718 | 0.4829 | 0.4743 | 0.4621 | 0.4496 | 0.4763 | 0.4741 | 0.4715 |
| | (0.1892) | (0.1789) | (0.1279) | (0.1188) | (0.1395) | (0.1522) | (0.1580) | (0.1751) | (0.1764) | (0.1950) |

Note: This table present the term structure of $\alpha$ between 1996 until 2013.

**Table 6.3: The Term Structure of $\delta$**
**Time To Maturity**

| **Year** | **10** | **9** | **8** | **7** | **6** | **5** | **4** | **3** | **2** | **1** |
|---|---|---|---|---|---|---|---|---|---|---|
| **1996** | 0.4535 | 0.5493 | 0.6443 | 0.5895 | 0.5895 | 0.4178 | 0.4014 | 0.3395 | 0.3288 | 0.3233 |
| | (0.3907) | (0.3529) | (0.3358) | (0.3858) | (0.3750) | (0.3983) | (0.3926) | (0.3903) | (0.3832) | (0.3732) |
| **1997** | 0.3127 | 0.4052 | 0.5780 | 0.6575 | 0.6946 | 0.5805 | 0.4662 | 0.4160 | 0.3529 | 0.3434 |
| | (0.3632) | (0.3633) | (0.3319) | (0.3181) | (0.2871) | (0.3645) | (0.3889) | (0.3966) | (0.3837) | (0.3893) |
| **1998** | 0.2308 | 0.2495 | 0.3900 | 0.5104 | 0.4949 | 0.5785 | 0.6735 | 0.3481 | 0.3738 | 0.2973 |
| | (0.3472) | (0.3475) | (0.3758) | (0.3225) | (0.3339) | (0.3093) | (0.3105) | (0.4010) | (0.4062) | (0.3888) |
| **1999** | 0.1739 | 0.2012 | 0.2253 | 0.2654 | 0.3102 | 0.4356 | 0.5907 | 0.6026 | 0.4469 | 0.1657 |
| | (0.3244) | (0.3353) | (0.3551) | (0.3496) | (0.3856) | (0.3607) | (0.2397) | (0.3510) | (0.3927) | (0.3299) |
| **2000** | 0.1007 | 0.0844 | 0.1275 | 0.1280 | 0.1638 | 0.3092 | 0.4081 | 0.5945 | 0.5438 | 0.1326 |
| | (0.2752) | (0.2362) | (0.2878) | (0.2762) | (0.3182) | (0.3630) | (0.3755) | (0.2773) | (0.3376) | (0.3033) |
| **2001** | 0.0537 | 0.0467 | 0.0938 | 0.1015 | 0.1148 | 0.1119 | 0.3045 | 0.4048 | 0.3820 | 0.4013 |
| | (0.2055) | (0.1914) | (0.2576) | (0.2598) | (0.2770) | (0.2573) | (0.3212) | (0.3678) | (0.3440) | (0.3873) |
| **2002** | 0.3212 | 0.0479 | 0.0682 | 0.1234 | 0.1623 | 0.1618 | 0.1930 | 0.3328 | 0.4128 | 0.4266 |
| | (0.3130) | (0.1955) | (0.2258) | (0.2957) | (0.3302) | (0.3045) | (0.3133) | (0.3176) | (0.2828) | (0.2766) |
| **2003** | 0.4166 | 0.3342 | 0.0368 | 0.0287 | 0.0327 | 0.1116 | 0.1381 | 0.1963 | 0.3578 | 0.4599 |
| | (0.2780) | (0.2919) | (0.1732) | (0.1574) | (0.1712) | (0.2749) | (0.2831) | (0.3264) | (0.3260) | (0.2437) |
| **2004** | 0.4586 | 0.4809 | 0.4219 | 0.2621 | 0.1627 | 0.1515 | 0.1747 | 0.2263 | 0.2422 | 0.3938 |
| | (0.2539) | (0.2228) | (0.2595) | (0.3112) | (0.2947) | (0.3151) | (0.3390) | (0.3656) | (0.3597) | (0.3179) |
| **2005** | 0.3839 | 0.4024 | 0.5176 | 0.5252 | 0.5334 | 0.3796 | 0.2374 | 0.1696 | 0.1599 | 0.1818 |
| | (0.3184) | (0.3228) | (0.1381) | (0.1824) | (0.1906) | (0.3175) | (0.3114) | (0.3339) | (0.3185) | (0.3270) |
| **2006** | 0.2402 | 0.3617 | 0.3889 | 0.5309 | 0.5269 | 0.5333 | 0.5166 | 0.3997 | 0.3428 | 0.2034 |
| | (0.3508) | (0.3348) | (0.3081) | (0.1446) | (0.1449) | (0.1438) | (0.1625) | (0.3307) | (0.3522) | (0.3357) |
| **2007** | 0.1413 | 0.1668 | 0.2340 | 0.3519 | 0.4718 | 0.5308 | 0.5168 | 0.5137 | 0.4559 | 0.2203 |
| | (0.3013) | (0.3173) | (0.3362) | (0.3331) | (0.1977) | (0.1344) | (0.1300) | (0.1116) | (0.1936) | (0.3253) |
| **2008** | 0.4101 | 0.3802 | 0.2172 | 0.1620 | 0.3005 | 0.4014 | 0.5187 | 0.5211 | 0.5231 | 0.5032 |
| | (0.2581) | (0.3011) | (0.3225) | (0.2958) | (0.3186) | (0.2681) | (0.1144) | (0.1346) | (0.1363) | (0.1988) |
| **2009** | 0.5130 | 0.5188 | 0.4940 | 0.4618 | 0.4369 | 0.4206 | 0.4560 | 0.5169 | 0.5206 | 0.5216 |
| | (0.1197) | (0.1572) | (0.1986) | (0.2882) | (0.2831) | (0.2730) | (0.2292) | (0.1190) | (0.1120) | (0.1233) |
| **2010** | 0.5214 | 0.5240 | 0.5229 | 0.5177 | 0.5343 | 0.5196 | 0.5598 | 0.5180 | 0.5106 | 0.5244 |
| | (0.1277) | (0.1333) | (0.1330) | (0.1366) | (0.1509) | (0.1682) | (0.2295) | (0.1976) | (0.1425) | (0.1366) |
| **2011** | 0.5046 | 0.5117 | 0.5151 | 0.5305 | 0.5207 | 0.5331 | 0.5662 | 0.5671 | 0.5501 | 0.5145 |
| | (0.1248) | (0.1351) | (0.1297) | (0.1364) | (0.1448) | (0.1727) | (0.2276) | (0.2624) | (0.2290) | (0.1172) |
| **2012** | 0.5177 | 0.5092 | 0.5236 | 0.5142 | 0.5309 | 0.5345 | 0.5165 | 0.5562 | 0.5589 | 0.5474 |
| | (0.1374) | (0.1145) | (0.1185) | (0.1135) | (0.1178) | (0.1375) | (0.1233) | (0.1836) | (0.1992) | (0.1688) |
| **2013** | 0.5288 | 0.5268 | 0.5092 | 0.5074 | 0.5091 | 0.5159 | 0.5210 | 0.5156 | 0.5258 | 0.5301 |
| | (0.1452) | (0.1494) | (0.1112) | (0.0977) | (0.1039) | (0.1190) | (0.1149) | (0.1316) | (0.1335) | (0.1579) |

Note: This table present the term structure of $\delta$ between 1996 until 2013.

**Table 6.4: The Term Structure of $\mu$**

**Time To Maturity**

| **Year** | **10** | **9** | **8** | **7** | **6** | **5** | **4** | **3** | **2** | **1** |
|---|---|---|---|---|---|---|---|---|---|---|
| **1996** | 0.6826 | 0.6188 | 0.5320 | 0.5583 | 0.5474 | 0.5476 | 0.5721 | 0.6128 | 0.6569 | 0.6950 |
| | (0.3179) | (0.3002) | (0.2658) | (0.2909) | (0.2840) | (0.3023) | (0.2981) | (0.3058) | (0.3006) | (0.3054) |
| **1997** | 0.7043 | 0.6745 | 0.6056 | 0.5340 | 0.5286 | 0.5756 | 0.5249 | 0.5617 | 0.6136 | 0.6594 |
| | (0.2907) | (0.2914) | (0.2863) | (0.2561) | (0.2381) | (0.2758) | (0.2915) | (0.2866) | (0.2905) | (0.2948) |
| **1998** | 0.6654 | 0.6885 | 0.6415 | 0.5941 | 0.5846 | 0.5127 | 0.5785 | 0.5160 | 0.5459 | 0.6177 |
| | (0.2755) | (0.2873) | (0.3052) | (0.2770) | (0.2714) | (0.2554) | (0.2376) | (0.2712) | (0.2801) | (0.2817) |
| **1999** | 0.6009 | 0.6665 | 0.7056 | 0.6927 | 0.6700 | 0.6678 | 0.5077 | 0.5837 | 0.5728 | 0.5820 |
| | (0.2803) | (0.3050) | (0.3001) | (0.3060) | (0.3223) | (0.2911) | (0.1709) | (0.2326) | (0.2350) | (0.2502) |
| **2000** | 0.5863 | 0.6235 | 0.6907 | 0.7003 | 0.7116 | 0.6612 | 0.6402 | 0.5118 | 0.5380 | 0.5358 |
| | (0.2202) | (0.2299) | (0.2368) | (0.2554) | (0.2769) | (0.2791) | (0.2808) | (0.1856) | (0.2044) | (0.1951) |
| **2001** | 0.5353 | 0.5714 | 0.6037 | 0.6539 | 0.6547 | 0.6988 | 0.6292 | 0.6203 | 0.5711 | 0.5849 |
| | (0.1822) | (0.2000) | (0.2392) | (0.2649) | (0.2681) | (0.2762) | (0.2599) | (0.2629) | (0.2253) | (0.1899) |
| **2002** | 0.5307 | 0.5186 | 0.5490 | 0.5975 | 0.6215 | 0.6597 | 0.6983 | 0.6562 | 0.6024 | 0.4833 |
| | (0.1742) | (0.1721) | (0.2119) | (0.2384) | (0.2712) | (0.2708) | (0.2926) | (0.2713) | (0.2582) | (0.1527) |
| **2003** | 0.4775 | 0.4924 | 0.4973 | 0.5352 | 0.5345 | 0.6129 | 0.6536 | 0.6729 | 0.6151 | 0.5674 |
| | (0.1552) | (0.1372) | (0.1593) | (0.1893) | (0.1838) | (0.2954) | (0.2918) | (0.2938) | (0.2668) | (0.2079) |
| **2004** | 0.5721 | 0.4992 | 0.4918 | 0.4727 | 0.4584 | 0.4955 | 0.6074 | 0.6304 | 0.6567 | 0.6029 |
| | (0.2080) | (0.1502) | (0.1390) | (0.1778) | (0.1838) | (0.2378) | (0.3043) | (0.3162) | (0.3110) | (0.2828) |
| **2005** | 0.6163 | 0.6097 | 0.5164 | 0.5060 | 0.5114 | 0.4630 | 0.4858 | 0.5449 | 0.6257 | 0.6969 |
| | (0.2669) | (0.2615) | (0.1247) | (0.1391) | (0.1657) | (0.1953) | (0.2132) | (0.2678) | (0.2877) | (0.2966) |
| **2006** | 0.6374 | 0.5939 | 0.5845 | 0.5166 | 0.5172 | 0.5051 | 0.4911 | 0.4479 | 0.5000 | 0.6233 |
| | (0.3060) | (0.2689) | (0.2604) | (0.1345) | (0.1377) | (0.1214) | (0.1348) | (0.2389) | (0.2740) | (0.3098) |
| **2007** | 0.5492 | 0.5916 | 0.5942 | 0.6100 | 0.5331 | 0.5029 | 0.5035 | 0.5004 | 0.4891 | 0.5035 |
| | (0.2648) | (0.2816) | (0.2808) | (0.2569) | (0.1762) | (0.1190) | (0.1163) | (0.0960) | (0.1433) | (0.2350) |
| **2008** | 0.5157 | 0.5675 | 0.6171 | 0.6744 | 0.6480 | 0.5741 | 0.5120 | 0.5061 | 0.4967 | 0.4874 |
| | (0.1790) | (0.2345) | (0.2876) | (0.2817) | (0.2786) | (0.2344) | (0.1044) | (0.1172) | (0.1189) | (0.1673) |
| **2009** | 0.5118 | 0.4950 | 0.5129 | 0.5397 | 0.5852 | 0.5819 | 0.5650 | 0.5160 | 0.5035 | 0.4964 |
| | (0.1052) | (0.1323) | (0.1731) | (0.2324) | (0.2448) | (0.2447) | (0.2211) | (0.1050) | (0.1098) | (0.1106) |
| **2010** | 0.5122 | 0.5037 | 0.5089 | 0.4919 | 0.5087 | 0.5162 | 0.5628 | 0.5493 | 0.5111 | 0.5151 |
| | (0.1150) | (0.1169) | (0.1004) | (0.1165) | (0.1213) | (0.1561) | (0.2077) | (0.1876) | (0.1300) | (0.1224) |
| **2011** | 0.5169 | 0.5082 | 0.5171 | 0.5142 | 0.5061 | 0.4944 | 0.5025 | 0.5075 | 0.5398 | 0.5068 |
| | (0.1161) | (0.1175) | (0.1162) | (0.1026) | (0.1105) | (0.1801) | (0.2223) | (0.2812) | (0.2317) | (0.1136) |
| **2012** | 0.5254 | 0.5259 | 0.5000 | 0.5144 | 0.4968 | 0.5121 | 0.4983 | 0.5095 | 0.4955 | 0.5438 |
| | (0.1292) | (0.1192) | (0.1165) | (0.1140) | (0.1045) | (0.1237) | (0.1224) | (0.1651) | (0.1925) | (0.1857) |
| **2013** | 0.5277 | 0.5226 | 0.5169 | 0.5017 | 0.5121 | 0.5032 | 0.5123 | 0.4990 | 0.4953 | 0.4870 |
| | (0.1425) | (0.1475) | (0.1210) | (0.0955) | (0.1024) | (0.1106) | (0.1043) | (0.1089) | (0.1179) | (0.1544) |

Note: This table present the term structure of $\mu$ between 1996 until 2013.

**Table 6.5: The Term Structure of $\varepsilon$**

**Time To Maturity**

| **Year** | **10** | **9** | **8** | **7** | **6** | **5** | **4** | **3** | **2** | **1** |
|---|---|---|---|---|---|---|---|---|---|---|
| **1996** | 0.4180 | 0.3983 | 0.4433 | 0.4559 | 0.5022 | 0.5236 | 0.4469 | 0.4824 | 0.4537 | 0.4505 |
| | (0.2939) | (0.2399) | (0.2219) | (0.3008) | (0.2785) | (0.3344) | (0.3431) | (0.3546) | (0.3429) | (0.3302) |
| **1997** | 0.4303 | 0.4161 | 0.4138 | 0.4720 | 0.4391 | 0.4752 | 0.4590 | 0.4096 | 0.4088 | 0.4006 |
| | (0.3281) | (0.3008) | (0.2411) | (0.2188) | (0.1878) | (0.2866) | (0.3095) | (0.3284) | (0.3328) | (0.3406) |
| **1998** | 0.4414 | 0.4908 | 0.4410 | 0.4473 | 0.4842 | 0.4589 | 0.4954 | 0.5319 | 0.5018 | 0.4334 |
| | (0.3693) | (0.3618) | (0.3240) | (0.2599) | (0.2547) | (0.2336) | (0.1996) | (0.3257) | (0.3390) | (0.3547) |
| **1999** | 0.5075 | 0.4977 | 0.5163 | 0.5377 | 0.5323 | 0.5250 | 0.5076 | 0.5465 | 0.5573 | 0.5157 |
| | (0.3673) | (0.3859) | (0.3849) | (0.3538) | (0.3563) | (0.2988) | (0.1504) | (0.2082) | (0.2439) | (0.3469) |
| **2000** | 0.5648 | 0.5380 | 0.6025 | 0.6524 | 0.6327 | 0.5675 | 0.5656 | 0.4973 | 0.5224 | 0.5441 |
| | (0.3181) | (0.3387) | (0.3958) | (0.3721) | (0.3896) | (0.3557) | (0.3204) | (0.1383) | (0.1753) | (0.2755) |
| **2001** | 0.6293 | 0.6844 | 0.6769 | 0.6708 | 0.6507 | 0.6719 | 0.6107 | 0.5842 | 0.5783 | 0.5787 |
| | (0.2460) | (0.2808) | (0.3250) | (0.3584) | (0.3737) | (0.3699) | (0.3323) | (0.3070) | (0.2604) | (0.1815) |
| **2002** | 0.6221 | 0.6801 | 0.6970 | 0.6839 | 0.6985 | 0.6641 | 0.5945 | 0.5748 | 0.5695 | 0.5648 |
| | (0.1689) | (0.2258) | (0.2677) | (0.3189) | (0.3570) | (0.3444) | (0.3553) | (0.3075) | (0.2608) | (0.1466) |
| **2003** | 0.5518 | 0.6066 | 0.7263 | 0.7431 | 0.7490 | 0.7027 | 0.6813 | 0.6503 | 0.5707 | 0.5343 |
| | (0.1535) | (0.1718) | (0.1766) | (0.2305) | (0.2144) | (0.3425) | (0.3496) | (0.3481) | (0.3087) | (0.2245) |
| **2004** | 0.5231 | 0.5206 | 0.5635 | 0.6433 | 0.6670 | 0.6847 | 0.7014 | 0.6732 | 0.6494 | 0.5584 |
| | (0.2335) | (0.1403) | (0.1577) | (0.1884) | (0.1953) | (0.2443) | (0.3476) | (0.3515) | (0.3589) | (0.3175) |
| **2005** | 0.5641 | 0.5725 | 0.4613 | 0.5008 | 0.5187 | 0.6296 | 0.6537 | 0.7211 | 0.6872 | 0.6181 |
| | (0.3011) | (0.2900) | (0.1284) | (0.1210) | (0.1448) | (0.1880) | (0.2295) | (0.3108) | (0.3422) | (0.3577) |
| **2006** | 0.6274 | 0.6044 | 0.5766 | 0.4445 | 0.4533 | 0.4841 | 0.5265 | 0.6419 | 0.6613 | 0.6133 |
| | (0.3551) | (0.3164) | (0.3050) | (0.1475) | (0.1333) | (0.1178) | (0.1222) | (0.2370) | (0.2885) | (0.3555) |
| **2007** | 0.6690 | 0.6567 | 0.6227 | 0.5901 | 0.4757 | 0.4587 | 0.5003 | 0.5143 | 0.5427 | 0.6882 |
| | (0.3515) | (0.3757) | (0.3712) | (0.3251) | (0.2121) | (0.1297) | (0.1109) | (0.0684) | (0.1647) | (0.2951) |
| **2008** | 0.6122 | 0.6559 | 0.7418 | 0.7132 | 0.6333 | 0.5898 | 0.4738 | 0.4973 | 0.5160 | 0.5585 |
| | (0.2271) | (0.2615) | (0.3059) | (0.3444) | (0.3343) | (0.2754) | (0.0859) | (0.0990) | (0.0944) | (0.1412) |
| **2009** | 0.5104 | 0.5352 | 0.5686 | 0.6160 | 0.5960 | 0.5619 | 0.5145 | 0.4743 | 0.4823 | 0.5030 |
| | (0.0754) | (0.1256) | (0.1786) | (0.2524) | (0.2742) | (0.2641) | (0.2230) | (0.1072) | (0.0995) | (0.1055) |
| **2010** | 0.4766 | 0.4923 | 0.4973 | 0.5191 | 0.5263 | 0.5281 | 0.5174 | 0.4590 | 0.4457 | 0.4550 |
| | (0.0972) | (0.1122) | (0.0906) | (0.0889) | (0.0971) | (0.1485) | (0.1848) | (0.1815) | (0.1341) | (0.1134) |
| **2011** | 0.4547 | 0.4503 | 0.4659 | 0.4827 | 0.5083 | 0.5366 | 0.5755 | 0.5900 | 0.5241 | 0.4522 |
| | (0.1206) | (0.1231) | (0.1018) | (0.0925) | (0.0871) | (0.1490) | (0.1822) | (0.2228) | (0.2000) | (0.1280) |
| **2012** | 0.4463 | 0.4531 | 0.4582 | 0.4598 | 0.4684 | 0.5039 | 0.5068 | 0.5434 | 0.5663 | 0.4724 |
| | (0.1512) | (0.1362) | (0.1300) | (0.1221) | (0.1047) | (0.1110) | (0.1024) | (0.1352) | (0.1549) | (0.1811) |
| **2013** | 0.4943 | 0.4861 | 0.4657 | 0.4692 | 0.4612 | 0.4746 | 0.4756 | 0.5052 | 0.5143 | 0.5368 |
| | (0.1515) | (0.1528) | (0.1200) | (0.1118) | (0.1181) | (0.1117) | (0.0951) | (0.0969) | (0.0997) | (0.1436) |

Note: This table present the term structure of $\varepsilon$ between 1996 until 2013.

low(high) of $\alpha$ does not follow the low(high) of PIN, however, they have a similar trend but are slightly overlapping for some period(I leave a full investigation for future research). Interestingly, after 2008, there is a small variation for all parameters, for instance from 2009 to 2013 the $\alpha$ for ten to one year to maturity shows an average value at 0.4, the average estimate of $\delta$, $\mu$ and $\varepsilon$ are 0.5. Finally, the average estimated PIN from 2009 has a small variation around 15% from the identical PIN at 0.667.

## 6.5 Analysis and Implications: PIN and VPIN comparison

The following section presents various results of PIN measures on individual ED futures contracts following the conditional imputation approach of Easley et al. [1996]. It will also present results for VPIN drawing on Easley et al. [2012] approach but with different $n$ and $V_i$. First, for the average PINs estimated they are in excess 0.5 for the majority of the sample. This is high relative to comparable equity market studies, see Table 6.6. However, the results coincide with those of Kim et al. [2014] on the CME Globex trades, for a short sample within my sample and Easley et al. [2012] for related US Dollar Treasury notes. Second, the average VPIN from my results is 0.126 which is lower than other estimates; however, the range of VPIN is wider and the minimum VPIN is 0.045 and the maximum is 0.998.

Although from this study the PIN estimates cover between 0.10 and 0.80 in the equities market, in the derivatives market the range is higher with minimum and maximum values of 0.04 and 0.99, respectively. Also, in the derivatives mar-

**Table 6.6: Comparison between average 'Probability of Informed Trading' estimates across various studies and markets.**

| Author | Asset | Sample Period | Min–max | Mean |
|---|---|---|---|---|
| Easley et al. [1996],(PIN) | NYSE – 90 stocks | Oct. 1, 1990 to Dec. 23, 1990 | 0.120–0.342 | 0.197 |
| Idier and Nardelli [2011](PIN) | Euro overnight interbank rate, Money market | Dec. 2000 to Mar. 2008 | 0.200–0.580 | 0.480 |
| Easley et al. [2012],(VPIN) | E-mini S&P 500 (CME) | Jan. 1, 2008 to Oct. 30, 2010 | 0.205–0.830 | 0.393 |
| | T-Note (CBOT) | Jan. 1, 2008 to Oct. 30, 2010 | 0.200–0.800 | 0.401 |
| | EUR/USD(CME) | Jan. 1, 2008 to Oct. 30, 2010 | 0.150–0.780 | 0.327 |
| | Brent Crude Oil (ICE) | Jan. 1, 2008 to Oct. 30, 2010 | 0.200–0.770 | 0.384 |
| | Silver (COMEX) | Jan. 1, 2008 to Oct. 30, 2010 | 0.200–0.840 | 0.411 |
| Abad and Yagüe [2012],(PIN) | Spanish Stock Exchange – 15 stocks | Jan. 1, 2009 to Dec. 31, 2009 | 0.104–0.501 | 0.227 |
| Kim et al. [2014],(VPIN) | Intra day trading – Eurodollar Futures (CME) | Jan. 3, 2005 to Dec. 29, 2006 | 0.760–0.970 | 0.880 |
| Yan and Zhang [2014],(PIN) | NYSE/AMEX stocks that have data in the ISSM and TAQ databases | Jan, 1, 1983 to Dec. 31, 2004 | 0.177–0.227 | 0.201 |
| **My PIN** | 40 Eurodollar Future contracts | Jan. 1, 1996 to Jan 1, 2014 | 0.369–0.992 | 0.688 |
| **My VPIN** | 40 Eurodollar Future contracts | Jan. 1, 1996 to Jan 1, 2014 | 0.045–0.998 | 0.126 |

Note: this table compares the PIN and VPIN from this research with others for various products. It can be seen that my PIN is higher than other PINs, which might be affected by the trading mechanism in the futures market which is different from the stock exchange; however, my VPIN is lower than other VPINs.

ket the min–max spread is higher than in the equities market, particularly in the interest rate derivatives market. The highest mean of PIN appears in Kim et al. [2014] with a value of 0.88 ;yet, the highest actual value of PIN is shown in this work with the maximum value of 0.992. The minimum VPIN is 0.040 and the maximum is 0.998. Interestingly, the value of PIN in the Eurodollar Futures is higher than the equity market and always high as presented in this work the PIN is mostly higher than 0.50. However, despite my PIN being constantly high, the average VPIN in this study is lower than others such as in Kim et al. [2014] in which it varies between 0.76 and 0.97. Table 6.6, illustrates this point and this aspect of my study would be interesting to investigate in the future.

### 6.5.1 Following the VPIN Dispute

The VPIN as a tool for detecting excessive levels of order imbalance has been subject to considerable academic discussion and it is worth reviewing some of the key input choices in determining its value. As previously noted, Andersen and Bondarenko [2013, 2014c] and Abad et al. [2015] have criticized the VPIN primarily due to the lack of a good mechanism to be able to choose the correct bulk volume classification. Hence the number of buckets ($n$) and the bucket size ($V_i$) are nuisance parameters with no simple method of constructing an appropriate statistical loss function to provide guidance on their values. My solution to this problem is to repeat the analysis over a wide range of different $n$ and $V_i$ and, indeed, I will show that the VPIN, at times, can be sensitive to these choices. Furthermore, this work computes the VPIN for VPIN20, VPIN50, VPIN100 and VPIN200, then compares them with original point estimates of the PIN. I can observe from this that there are substantively different values of VPIN during January 19-29 2010; the minimal value of VPIN is VPIN20 and maximal VPIN is VPIN200 almost continuously through this snapshot of data. For example, PIN and VPIN estimates on January 23, 2010, for the PIN is 0.7521 and 0.0132, 0.0096, 0.0096, 0.0048 for VPIN20, VPIN50, VPIN100 and VPIN200, respectively. From this example we can see that the absolute value of VPIN is sensitive to the choice of $n$ and $V_i$ is certainly not in question. However, the correlations and cumulative differences across choices of $n$ and $V_i$ indicate a high level of agreement in direction, if not in level, which is in keeping with the commentary made in Easley et al. [2014].

Table 6.7 presents the correlation matrix and p-values of VPIN and PIN. Noticeably, there is a positive and significant correlation between all five variables. This indicates that there is a significant relationship between all five variables:

**Table 6.7: Long run correlation coefficients of PIN, VPIN20, VPIN50, VPIN100 and, VPIN200 on ED?0.**

| | VPIN20 | VPIN50 | VPIN100 | VPIN200 | PIN |
|---|---|---|---|---|---|
| **VPIN20** | 1 | | | | |
| **VPIN50** | 0.9250 | 1 | | | |
| | (0.0000) | | | | |
| **VPIN100** | 0.3127 | 0.349 | 1 | | |
| | (0.0000) | (0.0000) | | | |
| **VPIN200** | 0.2938 | 0.3297 | 0.2484 | 1 | |
| | (0.0000) | (0.0000) | (0.0000) | | |
| **PIN** | 0.1156 | 0.1362 | 0.1343 | 0.0843 | 1 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0051) | |

Note: This table illustrates the matrix of correlation between PIN, VPIN20, VPIN50, VPIN100 and, VPIN200 also a matrix of p-values for testing the hypothesis of no correlation with ($p < 0.05$)

the highest correlation is between VPIN20 and VPIN50 with $r = 0.9250, n = 1100, p = 0.000$ this decreases to $r = 0.3127, 0.2938$ and 0.1156 for VPIN100, VPIN200 and PIN respectively.

Figure 6.2 also illustrates the variability of the VPIN as a function of different choices of $n$ and $V_i$. The solid line represents the VPIN estimates for March 2002 maturing futures, the dotted line represents VPIN for June 2002 maturing contracts, and the dashed line represents the VPIN for September 2002 contracts. For each contract I have five different colors that represent different $n$ and $V_i$ choices ($n = 20$ with $V_i = 20$, $n = 50$ with $V_i = 50$, $n = 100$ with $V_i = 100$, and $n = 200$ with $V_i = 200$). The sensitivity, in level, of the VPIN to $n$ and $V_i$ is self-evident however, the cumulative differences are highly consistent.

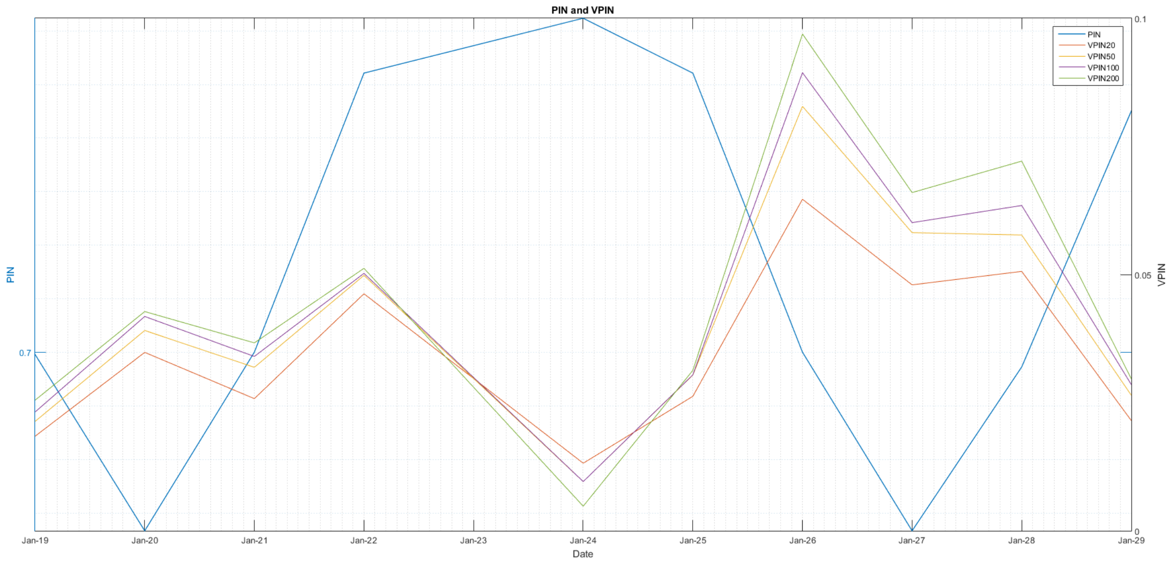



**Figure 6.2: Comparison of a four different types of VPIN on EDH2, EDM2, and EDU2.**
Note: This figure presents the compared plots of five different types of VPIN such as VPIN20, VPIN50, VPIN100 and VPIN200 between February 27 until March 09, 2012. It can be seen that VPIN is sensitive to different $n$ and $V_i$, however there is systematic trend between four types of VPINs. They can be described as, VPIN with $n = 20$ with $V_i = 20$, $n = 50$ with $V_i = 50$, $n = 100$ with $V_i = 100$, and $n = 200$ with $V_i = 200$, also the thick line represents EDH2, the dotted line represents EDM2 and the dash line represents EDU2.

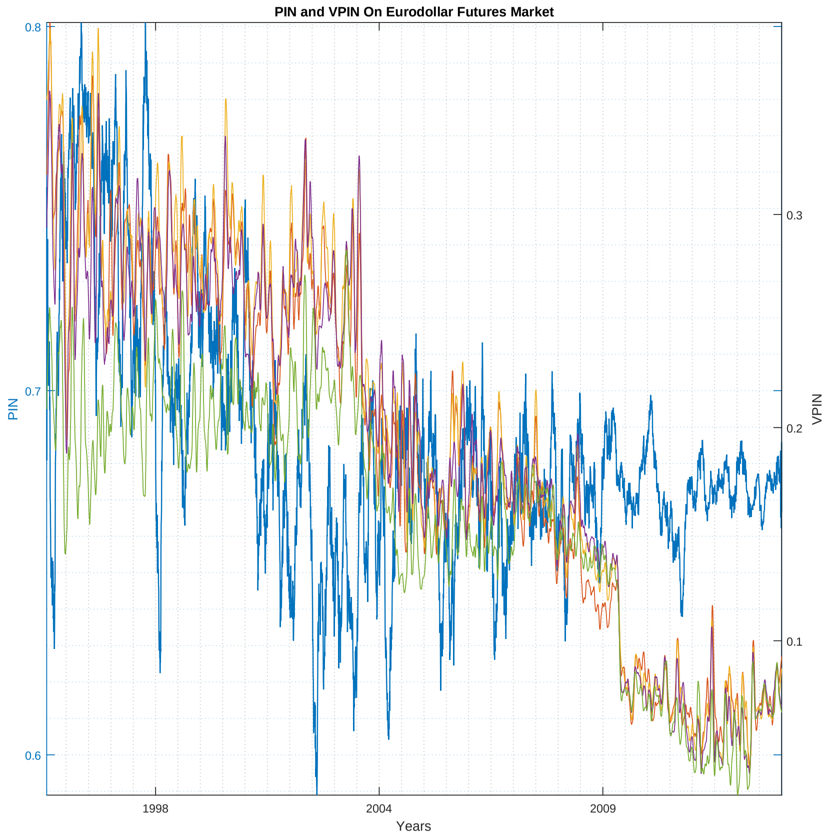
### 6.5.2 Empirical evidence, PIN and VPIN on Eurodollar Futures Market

This section presents the results of cross-sectional PIN and VPIN on the Eurodollar Futures market across 40 contracts from 1996 to 2013 (Figure 6.3). For the entire sample, I find the PIN and VPIN fluctuate with a downward sloping trend.

The left subplot of Figure 6.3 shows the average PIN on the left axis, and average VPIN20, VPIN50, VPIN100, and VPIN200 on the right axis. It also shows fluctuations in PIN and VPIN. The highest PIN occurs during 1996 to 1998 with a variation around 0.80-0.82. It decreases after 1998 with its lowest value at 0.58 around mid 2002. The PIN, after a downward trend from 1996 to mid 2002, bounced back and remains constantly high until 2013, at between 0.065 to 0.070. I assume the high value of PIN between 1996 to mid 2002 is caused by trading activity that operates on an open-outcry platform. This platform, due to its greater information content, has more effect than the electronic market, and can increase the PIN. This has been noted by previous studies on the effect of migration from open-outcry to electronic platforms (Shah and Brorsen [2011]; Ates and Wang [2005]; Aitken et al. [2004]; Tse and Zabotina [2001] ). This pattern, albeit with some variation, also appeared on the VPIN.

Overall, the VPIN shows a monotonic decrease from a high point of about 0.40 in 1996 to between 0.01 and 0.02 in 2012. It gradually drops after CME launched the Globex trading system in 2002. However, around 2004, the VPIN bounces back. This pattern appears after routine manipulation of the LIBOR as noted in the FSA and CFTC documents. At the same time, the VPIN is also constantly high although not higher than pre-2002. However, I find a significant decrease in the VPIN around May 2009, when it dropped from about 0.14 to 0.08 compared to the previous period. This decrease is a result of the launch of the LIBOR manipulation investigation. Thus, historical analysis of both PIN and VPIN shows variations in the ED futures market consistent with the FSA and CFTC documents. The same pattern can be observed in PIN and four types of VPIN.

The right subplot of Figure 6.3 presents cumulative ΔPIN and ΔVPIN. As the PIN and VPIN have differences in their range, it is hard to compare them, so, I had to normalize PIN and VPIN by investigating their differences - 'ΔPIN' and 'ΔVPIN'. Using the same analysis as in the left subplot of Figure 6.3, I find the



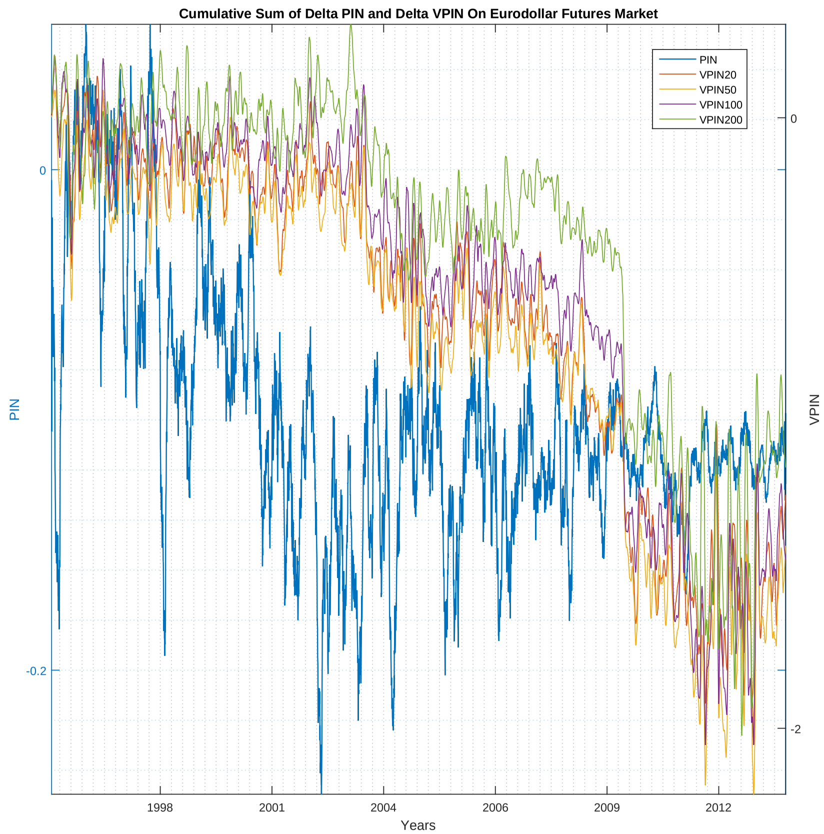




**Figure 6.3: The PIN (left-axis) and VPIN (right-axis) on Eurodollar Futures Market**

Note: This Figure presents the PIN and VPIN estimates for the whole cross section of the Eurodollar Futures Market, estimated directly from the limit order book data. See Table 2.1 for the sample description.

The left subplot of Figure 6.3, presents the levels of the PIN and VPIN estimates and the right subplot presents the cumulative change in the PIN and VPIN estimates from 1996 to 2013 for ease of interrogation. The VPIN is decomposed into four plots, and each represents a different choice of volume bucket size (20, 50, 100 and 200 contracts, with each contract having a notional value of $1million).

All of the PIN and VPIN measures have a substantial downward trend, with an approximate 20% decline over the sample. Interestingly, the measure of PIN drops several years prior to the measure of VPIN. It is worth noting that the drop in PIN corresponds to the period just prior to the introduction of the 'Globex' electronic trading platform and then stays relatively low thereafter. The pattern of the trend in VPIN is slightly different, albeit the level of reduction (around a 20% decline from 1996 to 2013) is very similar. The drop off in VPIN occurs in two sharp declines, one in 2003 and a second sharper decline in 2008/9.

same results in the right. That is the major pattern of ΔPIN and ΔVPIN showing a downward trend. Also, the ΔPIN and ΔVPIN had a significant drop after the Globex was launched in 2002, then increased again around 2004. Similarly, I find these two indicators increasing again around the time of the financial crisis, then dropping. However, ΔVPIN appears to be more sensitive than ΔPIN, as at the beginning of 2004 it dropped around 0.5 or 50% compared to 2002. During the same period, ΔPIN dropped only 10%. Also, there was a rapid decrease in VPIN after the LIBOR investigation with ΔVPIN dropping from -0.05 to -0.15 compared to the previous period.

### 6.5.3 Term Structure of PIN and VPIN

The left subplot of Figure 6.4 presents a cross-section of the term structure of PIN for 40 Eurodollar futures contracts for 3,653 days. In this figure I overlay plots of historical PIN against Days-To-Maturity (X axis) to investigate the variation of long term relationships. Although I have 40 different historical PINs, all show a similar pattern. The figure shows that from 3,653 to 3,000 days to maturity the PIN shows values between 0.62 to 0.92. After this period, the PIN slightly decreases and then drops to its lowest value at 0.32 at around 1,400 days to maturity. However, at this point PIN are highly volatile with a range of 40% with the highest PIN about 0.72 and the lowest 0.32. However, after 365 days to maturity the PIN had a much lower variation of around 0.15 to 0.10. One final finding from this plot is that from 1,400 days to maturity, the term structure of PIN tended to be V-shaped similar to the term structure of the price of futures.

The right subplot of Figure 6.4 presents a term structure of cumulative Δ PIN for 40 Eurodollar futures contracts for a 10 year period. In this figure, there is

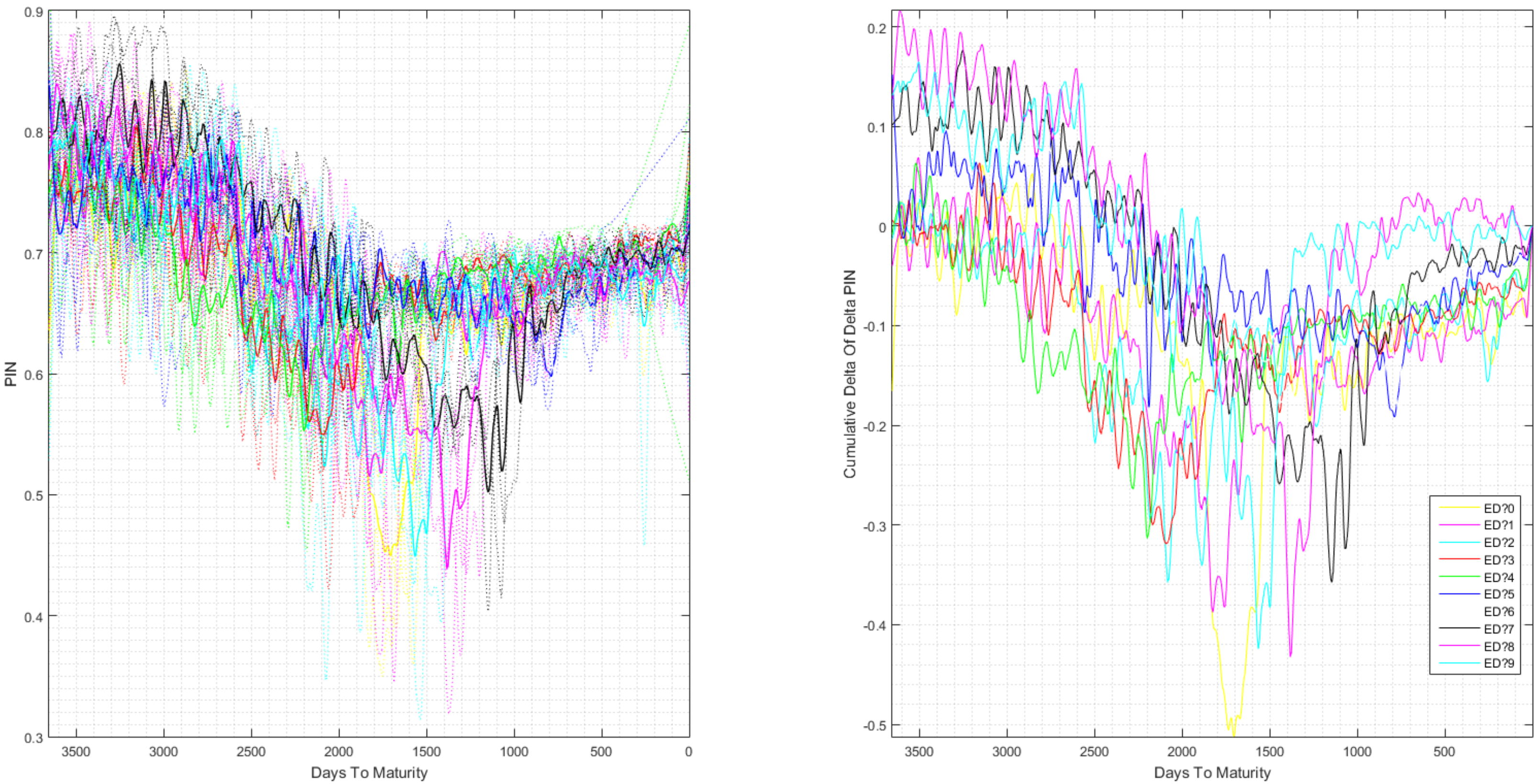



**Figure 6.4: Term Structure of PIN**

Note: This figure presents the term structure of PIN that varies between 0.72 to 0.85 from the first trading day to 3,000 days to maturity, then slightly decreases and drops to bottom at around 1,400 days or 3 to 4 years to maturity. The PIN in this period are highly volatile with a wider range compared to the previous period, the highest PIN at 0.72 and the lowest at 0.32. Furthermore, after three years to maturity an upward trend of PIN is shown with less volatility, with the lowest at 0.50 and the highest at 0.75, when it comes to shorter than one year to last trading day.

a similar pattern in the term structure of PIN and cumulative $\Delta$ PIN. Using the same format as in the left subplot of Figure 6.4 to illustrate the term structure of cumulative $\Delta$ PIN, I find a downward trend from the first trading day until 1,800 days to maturity. From the beginning of this trading period the cumulative $\Delta$ PIN varies around 20% between 0 to 0.2, then drops to its lowest value of between -0.5 and -0.1. Furthermore, at this point of time the cumulative $\Delta$PIN were extremely volatile with a range of around 40% with a high of -0.08 and a low of -0.52. Furthermore, from around 1,400 days to maturity, the cumulative $\Delta$ PIN increases with less variation which narrows to 10% in comparison to 40% in the last period. Finally, it is interesting to note that the term structure of PIN and cumulative $\Delta$ PIN from 1,400 days to expiration have the same V-shape as the typical term structure of futures.

Using the same format analysis from Figure 6.4, Figure 6.5 presents the term structure of VPIN on the left subplot and term structure of cumulative $\Delta$VPIN on the right subplot. Although I have shown the term structure of PIN from 3,653 days to expiration, I can present only 1,100 days to maturity for VPIN due to the volume bucket effects. That is to say, as described in the methodology section, for the VPIN algorithm to work, the volumes have to completely fill the volume bucket.

It can be seen from the left subplot of Figure 6.5 that from 1,100 to 700 days to maturity the term structure of VPIN tends to be U-shaped. The VPIN fluctuates from nearly 0 to just above 0.50 or 50% difference. Later, the variation range decreases from 50% to 20% on 700 days to expiration. Next, between 700 to 400 days to maturity, the VPIN slightly increases from 0.15 to 0.27 respectively, then decreases afterwards. The VPIN drops after 400 days and slightly increases again at around 120 days to expiry. During 400 to 120 days to expiry, the VPIN varies around 10% between 0.10 to 0.20. Interestingly, the differences between the lowest and highest tend to be narrower closer to maturity. This result is consistent with (Ballocchi et al. [2001]) in that the range of variation on ED futures price also decreases as the contracts approach expiration. A remarkable result from this plot indicates that from 365 days to expiration , the term structure of VPIN tends to be V-shaped as the term structure of futures prices.

The right subplot of Figure 6.5 presents the term structure of cumulative $\Delta\ VPIN$ across 40 ED contracts with four different types of VPIN. Overall, the term structure tends to be V-shaped similar to the term structure of futures prices, as the range of variation decreases over the period. The highest $\Delta\ VPIN$ was at the beginning of this period with the minimum at -2.5 and maximum at nearly 1. From 700 days to maturity, the range of variation increases, which is consistent with the term structure of VPIN when it slightly increases at around 700-days to maturity and decreases afterwards.

For clear comparison, this analysis focuses on PIN and VPIN from 1,100 days to maturity. Therefore, I find a similar pattern of the term structure of $\Delta\ PIN$ and $\Delta\ VPIN$, which tends to be V-shaped in respect to the term structure of futures contracts. To this end, the behavior of PIN in Figure 6.4 and VPIN in Figure 6.5 leads us to investigate whether time to maturity can help to explain the change in the variation of PIN and VPIN, $\Delta\ PIN$ and $\Delta\ VPIN$. The findings show that time to expiration tends to have an effect on the value and variation of both PIN and VPIN.

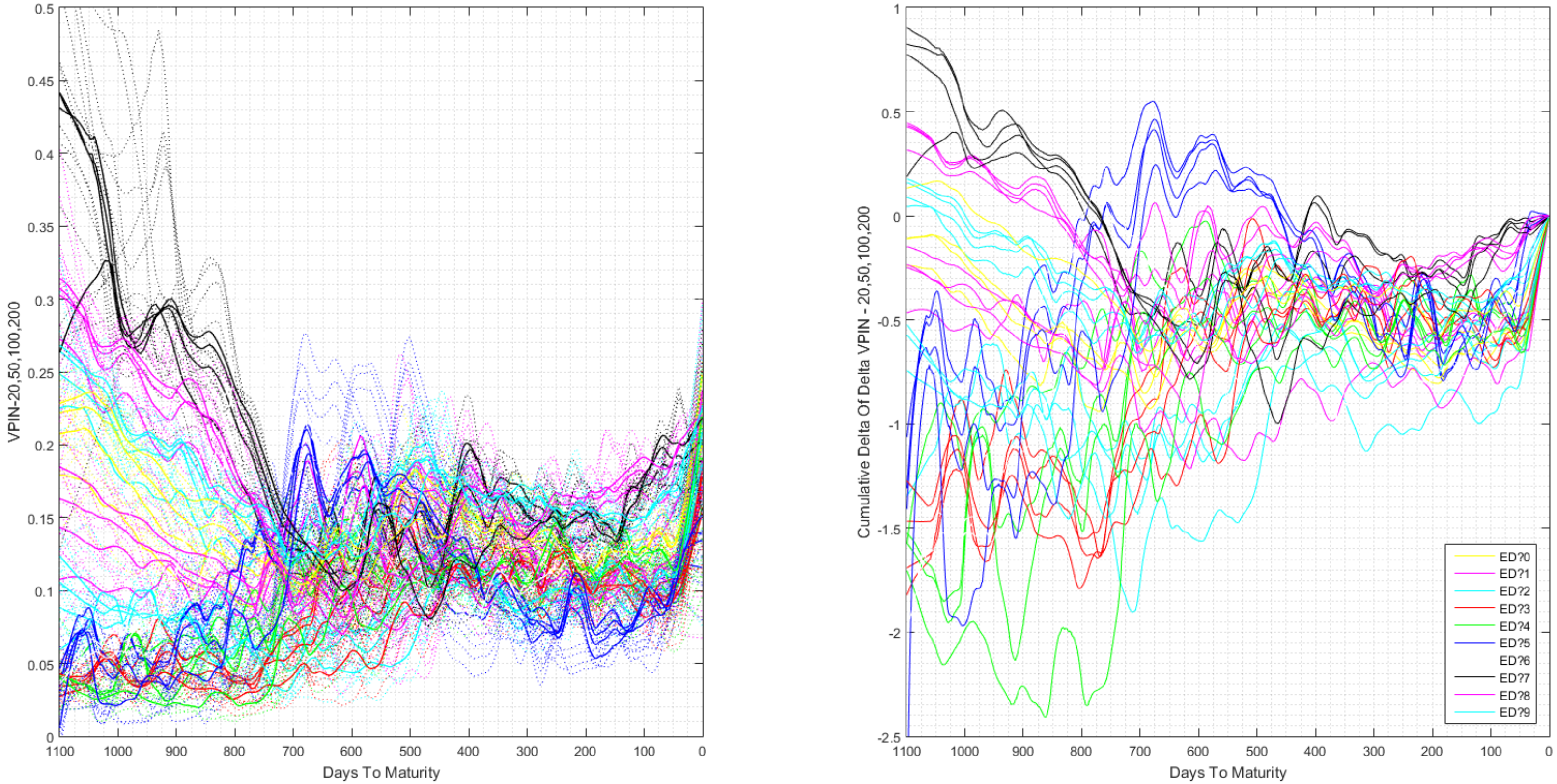



**Figure 6.5: Term Structure of VPIN and Term Structure of $\Delta VPIN$**

Note: This Figure presents the term structure of VPIN and term structure of $\Delta VPIN$. Despite the term structure of PIN starting from ten years to maturity, the VPIN can be observed from 1,100 days or around three years to maturity, and this is because there is not enough trading volume per day that can fill the volume bucket on the VPIN calculation. The term structure of VPIN tends to be U-shaped from 1,100 days or around three years to maturity and the variation of VPIN tends to be narrow for a shorter time to maturity. The VPIN drops after 400 days to maturity then has a slight monotonic increase from around 120 days left before the futures contract expires. The result of the narrow variation of VPIN may be consistent with a previous study that the variation of ED futures prices also decreases as the contracts approach expiration (Ballocchi et al. [2001]).

### 6.5.4 Correlation Surface

The similar patterns of PIN and VPIN against day-to-maturity leads us to investigate their relationship with time to maturity. To investigate this, Pearson's correlation is used to test the correlation between the PIN and four different types of VPIN from 1,000 days to maturity.

Figure 6.6 presents the correlation between PIN and VPIN20, VPIN50, VPIN100, VPIN200 from 1,000 days to maturity to expiry date. All types of VPIN positively correlate with PIN over time to maturity. However, this varies over the whole period as it increases nearer to maturity. The highest correlation appears between PIN and VPIN20 around 7 days with a correlation of 0.72. Inspecting the relationship between PIN and four VPIN, I find the PIN and the VPIN50 have a higher correlation compared to VPIN20, VPIN100, VPIN200 except after 15 days to maturity when the relationship drops lower than the relationship between PIN and VPIN20. Moreover, around 15 days before the final trading day, the highest correlation changes from VPIN50 to VPIN20 then reaches a peak before the end of the contract. In this period, the lowest correlation appears on VPIN200 followed by VPIN100. This figure pinpoints conclusively a strong effect of time to maturity for PIN and VPIN, as results show a higher correlation nearer to maturity.

### 6.5.5 PIN, VPIN and the LIBOR Manipulation

This section presents PIN and VPIN with the LIBOR manipulation case. Most market manipulation studies use general publicly available data. However, I use specific dates of LIBOR manipulation as cataloged in the CFTC and FSA documents presented in Table 5.3. To further investigate risk surrounding these

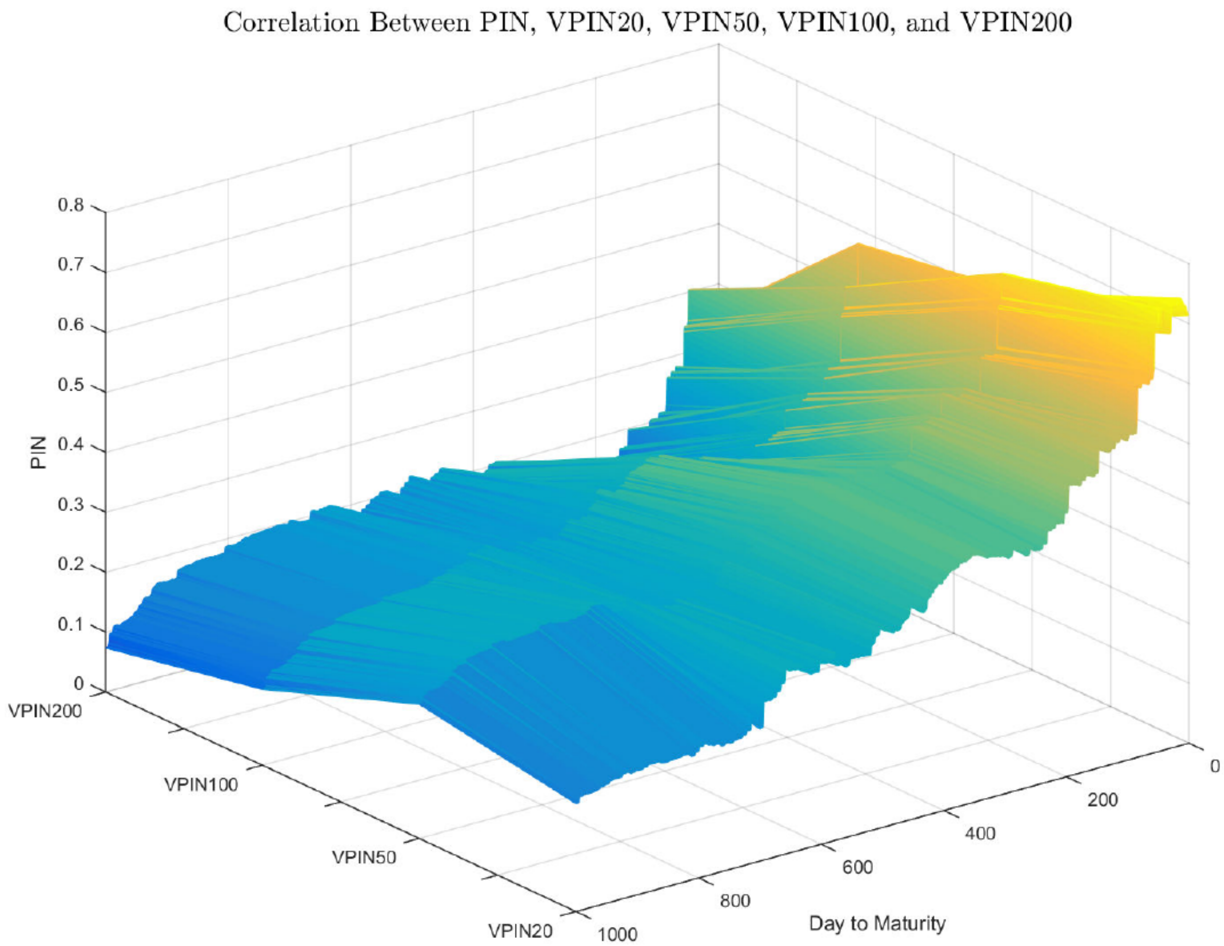



**Figure 6.6: Surface Plot of the correlation surface between the estimated PIN, VPIN20, VPIN50, VPIN100, and VPIN200.**

Note: This figure presents the correlation matrix between PIN and VPIN20, VPIN50, VPIN100, VPIN200 from 1,000 days to maturity. It can be seen that the correlation between PIN and VPIN significantly increases for a shorter time to maturity.

specific events, I calculate PIN, VPIN20, VPIN50, VPIN100, and VPIN200 to identify short term risk. I also observe long run behavior from 1996 to 2014.

I make the conjecture that LIBOR manipulation has an effect on PIN and VPIN in the futures market as ED trading is based on the LIBOR rate. This analysis begins with an analysis of FSA and CFTC documents regarding dishonest or manipulative submissions of the LIBOR rate issued to major banks, such as Barclays and RBS. The findings indicate that there are two different purposes of manipulation before and after the financial crisis. During the pre-crisis, the banks manipulated the LIBOR, attempting to generate competitive advantage for their trading positions to boost profits. However, during the financial crisis, LIBOR manipulation was an attempt to reduce customer perception of the banks high borrowing costs as this made them look desperate for cash. This was part of a beauty contest designed to 'reassure' investors and regulators.

According to the CFTC settlement document, Barclays' traders attempted to manipulate the US Dollar LIBOR from at least mid-2005 to the autumn of 2007, and thereafter sporadically until 2009. For instance, on March 31, 2006, Barclays employees, including the bank's senior Treasury managers, sent a request to a US dollar LIBOR submitter to submit a higher rate than normal. The submitter replied he would submit the rate requested.[1] At this point, I assume that the PIN and VPIN after the manipulation date should be higher than the previous period, then lower afterwards.

The top of the left panel of Figure 6.7 displays PIN and VPIN around March

[1] CFTC, Order instituting proceedings pursuant to sections 6(c) and 6(d) of the Commodity Exchange Act, as amended, making findings and imposing remedial sanctions; In the Matter of Barclays plc, pp. 9–10.

31, 2006 as recorded in the LIBOR manipulation document. In this case, I can investigate informed trading around this event by studying the variation of PIN and VPIN by using data from EDM6 contract, which has the closest expiry date. I investigate PIN and VPIN on EDH6 because this contract is the nearest expiration futures contract with the largest trading volume. Overall, the findings from this analysis show a significant increase in PIN and VPIN. Three days before the event date, the PIN, VPIN20, VPIN50 VPIN100 and VPIN200 move around 0.67, 0.16, 0.16, 0.165 and 0.125 respectively. Then, from the event date the PIN and VPIN sharply increases, achieving a peak on April 02, with the PIN at 0.82, VPIN20 at 0.50 and VPIN50 also at 0.50. However, there is no significant movement on VPIN10 and VPIN200.

For clarification, in the middle row of the left panel, the figure compares ΔPIN, ΔVPIN and the last row presents the PIN and VPIN based on the event date. On the day after the event date, the PIN increases by 0.22 or 22 %, VPIN20 and VPIN50 increases nearly three times when compared with the event date. Furthermore, three days after the event, the PIN and VPIN have significant decrease as the PIN, VPIN20 and VPIN50 shrinks by as much as it increased. Despite PIN VPIN20 and VPIN50 having a significant movement around the event, VPIN100 and VPIN200 have no significant changes before and after.

The left panel of Figure 6.7, presents the variation in PIN, VPIN on September 07, 2007, when Rabobank's senior US Dollar trader requested a US Dollar submitter to keep the 3M LIBOR high for the rest of that week.[1] This panel shows that between 03 and 13 September 2007, PIN and VPIN have a significant

[1]CFTC, Order instituting proceedings pursuant to detection(c) and 6(d) of the commodity exchange act, as amended, making findings and imposing remedial sanctions in the matter of Coöperative Central Raiffeisen Boerenleenbank B.A., p. 10.

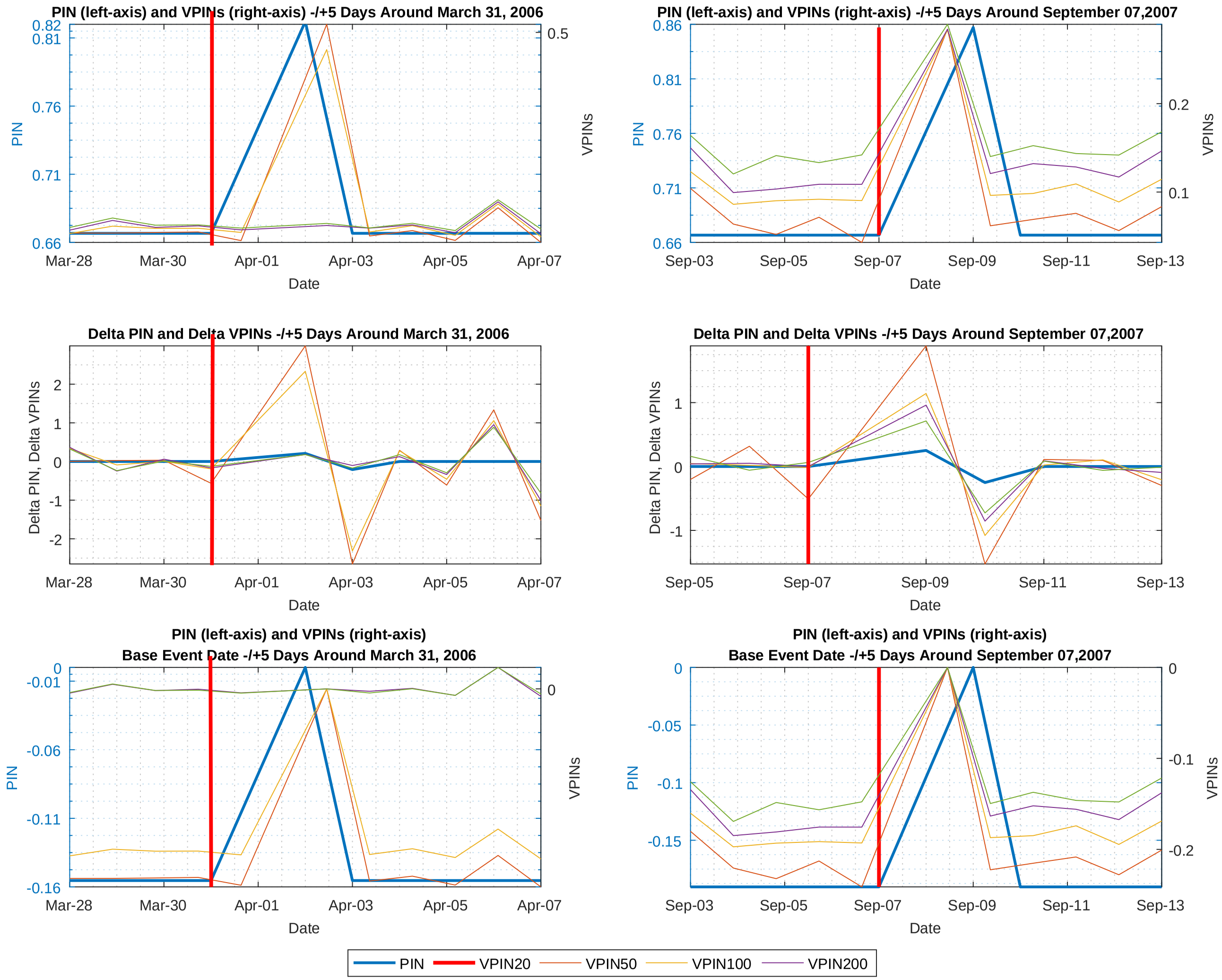



**Figure 6.7: The PIN and VPIN Around Identified LIBOR Manipulation.**

Note A: The left panel of figure 6.7 presents historical PIN on the left axis with a blue line and VPIN on the right axis for the LIBOR manipulation event, between March 28 and April 07, 2006. According to the CFTC document, Barclays' traders attempted to manipulate the US Dollar LIBOR on March 31, 2006, when Barclays' employees, including senior Treasury managers requested the Barclays' LIBOR submitter to submit a higher US Dollar LIBOR rate than the normal rate.

Note B: The right panel of figure 6.7 presents the PIN on the left axis with a blue line and VPIN on the right axis for LIBOR manipulation event between September 04 to 14, 2007. In this event, Rabobank's US Dollar LIBOR traders asked their submitter on September 07, 2007 to keep a high LIBOR rate for the rest of that week.

movement characterized by a spike of both on the day after the event. This manipulation influences PIN and VPIN on the EDU7 contract expiring on September 19, 2007. The PIN increases from 0.68, then reaches a peak on September 09 at 0.80, falling to 0.695 on the following day. Also, VPIN20, VPIN50, VPIN100 and VPIN200 increase from around 0.08 before the event to 0.17 on the event date, then reached a peak at 0.29 one day after, as a nearly threefold increase. Finally, similar to the PIN, the VPIN drops to its previous level.

I find PIN and VPIN have a systematic pattern around the LIBOR manipulation events as they increase on the event date, reach a peak the day after, and finally decrease to the previous level. However, it is interesting to note that normally all 40 Eurodollar futures contracts trade within the same time frame. As a result, informed traders may manipulate the ED in different contracts. The limitation of this study is that I cannot identify specific Eurodollar futures contract undertaken by specific banks as this information is not publicly available. Hence, the appropriate way to study PIN and VPIN is to investigate via a term structure and a term structure of variation.

Figure 6.8 presents the term structure of PIN (Top-Panel)and VPIN (Lower-Panel) for -/+60 days around LIBOR manipulation dates. This figure presents the overlay 2,080 plots of PIN and 8,320 plots of VPIN (dotted line) with average PIN and VPIN (thick black line), and also presents 95% confidence interval(black dashed line) into one plot to investigate how the PIN and VPIN react to LIBOR manipulation. The top panel of Figure 6.8 illustrates that the PIN has a 20% variation from 0.55 to 0.75. The cumulative differences of PIN or cumulative $\Delta$PIN has around a 50% variation of -0.02 to 0.50. However, the average PIN only slightly varies with the highest on the event day from 0.68 to 0.64. The

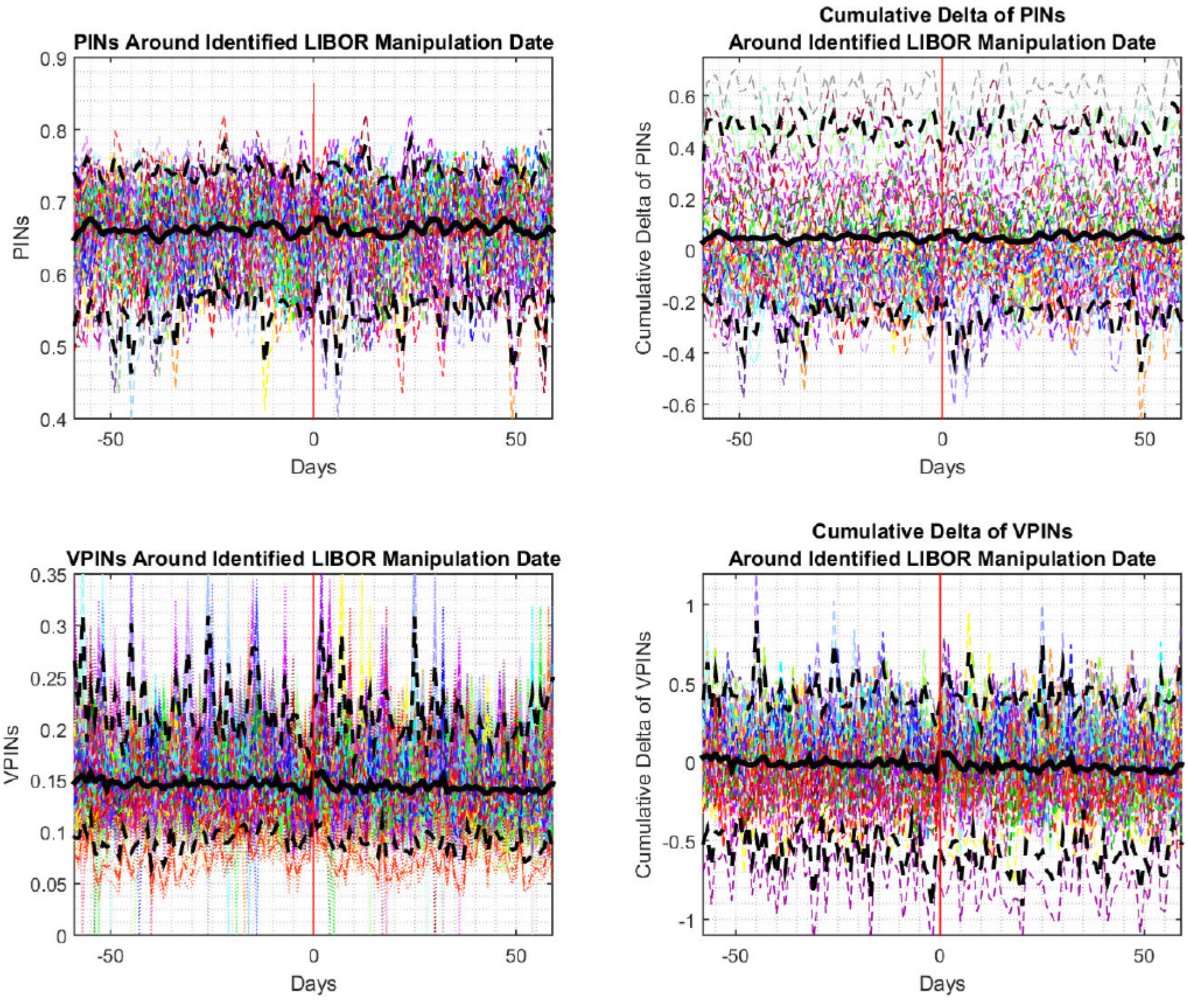



**Figure 6.8: The Term Structure of PIN, VPIN, Delta PIN and Delta VPIN for -/+60 days around identified LIBOR manipulation date.**
Note: This figure presents the term structure of PIN and VPIN for -/+60 days around the identified LIBOR manipulation date with the overlay 2,080 plots of PIN and 8,320 VPIN presents on a dotted line, the average PIN and VPIN on a thick black line, and also 95% confident interval on a black dashed line. It can be seen from this figure that the variation of PIN and VPIN around the LIBOR manipulation event in the average daily value is slightly fluctuates.

For the upper panel, the average PIN slightly fluctuates between the lowest at 0.65 and the highest at 0.68. Also, the PIN increases on the event date, then peaks between one or two days after the event date (day zero). Finally, the PIN drops to normal three days after the event.

In the lower panel of figure 6.8 presents individual VPIN on a dotted line and average VPIN on a thick black line. It can be seen that the average VPIN slightly fluctuates between the lowest at 0.13 and the highest at 0.18. Whilst the PIN is at a peak after the event, the VPIN spiked on the event date then drops between two and three days after the event.

cumulative average $\Delta$PIN also has a small variation between 0.02 and 0.08 with the highest on the event date.

The lower panel of Figure 6.8 presents the term structure of VPIN and cumulative $\Delta$VPIN with cross sectional average (thick line) across all events, all ED contracts and all types of VPIN (dotted line). It can be seen that the VPIN has higher variation than the PIN. The term structure of VPIN during the event has a minimum value of 0.02 and a maximum of 0.35 or around 35% variation. However, if I focus on the 95% confidence interval, I find VPIN at around 12% variation, mostly moving between 0.10 to 0.20. The average VPIN during LIBOR manipulation varies between 0.135 to 0.16 with the highest VPIN one day after event at 0.16 and the second highest on event day at 1.58. Finally, the average VPIN drops to normal around three days after the event. One remarkable result is the variation witnessed is lowest from the event date to two days later. This shows that during LIBOR manipulation all types of VPIN have a similar pattern, as they rapidly increase on the event date, remain high for two days then drop to normal.

The right subplot of the lower panel of Figure 6.8 presents the cumulative $\Delta$VPIN. In this plot, the cumulative $\Delta$VPIN has variation between -0.5 to 0.5 with average value at -0.0194. Also, the highest cross sectional average VPIN on the event day, at 0.07. This increases around 10% compared with the average cumulative $\Delta$VPIN. This result confirms that VPIN has more sensitivity to toxicity events, as the average VPIN and the average cumulative $\Delta$VPIN on the event date are higher than their average. In addition the $\Delta$ VPIN is constantly high for at least two days after the event, then drops to normal.

Overall, Figure 6.8 demonstrates that PIN and VPIN are able to detect toxicity events in the Eurodollar Futures market as they spike around the manipulation date. However, Table 6.8 shows that the cross sectional average PIN and VPIN have no statistically significant differences around these events, both pre and post 2008. In the pre-2008 period, the highest PIN appears for 60 days after the event and the lowest for 60 days before at 0.7069 and 0.6938 respectively. The highest VPIN in this period appears on VPIN50 for 30 days before the event and the lowest is VPIN20 for 60 days before the event at 0.1713 and 0.0487 respectively. For the post 2008 period, the highest PIN appears for 60 days before and the lowest for 60 days after the event at 0.6752 and 0.6663, respectively. The highest VPIN is VPIN200 for 30 days and, on the contrary, is VPIN20 for 60 days at 0.1773 and 0.0722 respectively. This shows how PIN and VPIN can act partially as an early warning signal for toxic events.

Table 6.9 below illustrates PIN and VPIN based on event dates both pre and post 2008. In the case of pre 2008, the first column shows that the PIN after day zero is higher than the previous period as the PIN is monotonically increasing from -0.0030 for sixty days before the event and reaches a peak 60 days after, when the PIN is 0.0101 higher than the PIN on event day. The second four columns present four different types of VPIN in the same period as PIN. Unlike the PIN, VPIN is not monotonically increasing. However, all four types of VPIN increase after day zero then drop from three to ten days after the event. The highest VPIN based on the event date is on VPIN20 for ten days after the event, at 0.0896 higher than event day.

Post 2008, the PIN and VPIN have a similar pattern as they rapidly increase on the event date and are constantly high for two to three days, then drop back

**Table 6.8: PIN, VPIN20, VPIN50, VPIN100,and VPIN200 -/+ 60 days around LIBOR manipulation events date.**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Informed trading around LIBOR manipulation events date** | | | | | | | | | |
| Days | PIN-Pre2008 | VPIN - Pre 2008 | | | | PIN-Post2008 | VPIN - Post 2008 | | | |
| | | 20 | 50 | 100 | 200 | | 20 | 50 | 100 | 200 |
| **-60** | **0.6938** | **0.0487** | **0.0600** | **0.0652** | **0.0722** | **0.6752** | **0.0490** | **0.0656** | **0.0815** | **0.0917** |
| | (0.0196) | (0.0584) | (0.0590) | (0.0505) | (0.0400) | (0.0089) | (0.0350) | (0.0356) | (0.0360) | (0.0329) |
| **-30** | **0.6946** | **0.1625** | **0.1713** | **0.1188** | **0.0810** | **0.6724** | **0.1400** | **0.1691** | **0.1773** | **0.1896** |
| | (0.0237) | (0.0630) | (0.0633) | (0.0563) | (0.0413) | (0.0089) | (0.0297) | (0.0305) | (0.0285) | (0.0290) |
| **-20** | **0.6957** | **0.0868** | **0.0972** | **0.1042** | **0.1193** | **0.6717** | **0.0925** | **0.1132** | **0.1192** | **0.1187** |
| | (0.0260) | (0.0662) | (0.0668) | (0.0566) | (0.0445) | (0.0088) | (0.0332) | (0.0337) | (0.0328) | (0.0325) |
| **-10** | **0.6967** | **0.0493** | **0.0886** | **0.0648** | **0.0706** | **0.6731** | **0.0741** | **0.0993** | **0.1121** | **0.1259** |
| | (0.0287) | (0.0623) | (0.0627) | (0.0537) | (0.0412) | (0.0081) | (0.0338) | (0.0344) | (0.0336) | (0.0330) |
| **0** | **0.6968** | **0.0500** | **0.0573** | **0.0673** | **0.0724** | **0.6732** | **0.0578** | **0.0786** | **0.0991** | **0.1174** |
| | (0.0325) | (0.0510) | (0.0627) | (0.0499) | (0.0375) | (0.0117) | (0.0332) | (0.0327) | (0.0326) | (0.0311) |
| **10** | **0.6995** | **0.1396** | **0.0602** | **0.0814** | **0.0884** | **0.6738** | **0.1241** | **0.1447** | **0.1635** | **0.1781** |
| | (0.0324) | (0.0364) | (0.0500) | (0.0302) | (0.0303) | (0.0053) | (0.0290) | (0.0266) | (0.0259) | (0.0209) |
| **20** | **0.7020** | **0.0570** | **0.1519** | **0.0712** | **0.0780** | **0.6686** | **0.0684** | **0.0947** | **0.1126** | **0.1314** |
| | (0.0318) | (0.0447) | (0.0376) | (0.0432) | (0.0451) | (0.0053) | (0.0290) | (0.0266) | (0.0250) | (0.0230) |
| **30** | **0.7043** | **0.0549** | **0.0650** | **0.6590** | **0.0701** | **0.6671** | **0.0700** | **0.0997** | **0.1271** | **0.1483** |
| | (0.0322) | (0.0492) | (0.0457) | (0.0435) | (0.0397) | (0.0043) | (0.0281) | (0.0271) | (0.0252) | (0.0228) |
| **60** | **0.7069** | **0.0911** | **0.0593** | **0.1263** | **0.1343** | **0.6663** | **0.0995** | **0.1128** | **0.1288** | **0.1389** |
| | (0.0350) | (0.0425) | (0.0412) | (0.0374) | (0.0338) | (0.0017) | 0.0316 | (0.0297) | (0.0288) | (0.0284) |

Note: This table compares the average PIN, VPIN20, VPIN50, VPIN100, and VPIN200 including its standard deviation for -/+60, 30, 20 and 10 days around the identified LIBOR manipulation event according to the CFTC and FSA documents from 1996 to the end of 2007, and from 2008 to the end of 2013. It can be seen from this table, that there is a small scale of variation before and after events date for both PIN and VPIN. However, the variation for both parameters for the first sub period(Pre-2008) is higher than post-2008. Despite the fact that there is not a statistically significant difference for PIN and VPIN, both , on the event date, are high compared to ten days before the event.

to normal. The highest PIN appears for ten days after the event at 0.0006, higher than the level on the event date. However, the PIN in this period is lower than pre 2008. The highest VPIN based on the event date appears for 30 days to maturity at 0.0822, 0.0905, 0.0782 and 0.0722, higher than the event date, for VPIN20 follow by VPIN50, VPIN100 and VPIN200 respectively. Both Table 6.8 and 6.9 demonstrate that the PIN is weaker than VPIN as a signal of market manipulation post 2008. For example, the average VPIN ten days after the event is mostly 0.06 or 6% higher than the VPIN on the event date. This is compared to 0.0027 or 0.27% for the PIN. However, overall there is no statistically significant difference for PIN and VPIN based on the event date around these events, for either pre or post 2008.

Despite the PIN and VPIN not strongly capturing the market manipulation for -/+ 60 days from the event date, focus on the variation of PIN and VPIN for -/+ 10 days, shows evidence of market manipulation, as this toxicity event is short lived. In general, ten days after the event PIN and VPIN are higher than their level on the event date. Especially in the post 2008, all types of VPIN perform well compared to the PIN. Focus on variation of PIN and VPIN based on the event date, gives us more information as PIN and VPIN for 10 days after the event are always higher than the level on the event date.

The results from table 6.8 and 6.9 suggest conclusively, the PIN and VPIN are sensitive to toxicity events or informed trading in the futures market, as both increase on the LIBOR manipulation event date. However, these parameters still have limited ability to detect any informed trading as the statistics show no significant difference before and after the event date for both pre and post 2008.

**Table 6.9: PIN, VPIN20, VPIN50, VPIN100,and VPIN200 based on events date -/+ 60 days around identified LIBOR manipulation events.**

| Informed trading around LIBOR manipulation events date(based event date) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Day | PIN-Pre2008 | VPIN - Pre 2008 | | | | PIN-Post2008 | VPIN - Post 2008 | | | |
| | | 20 | 50 | 100 | 200 | | 20 | 50 | 100 | 200 |
| **-60** | **-0.0030** | **-0.0013** | **0.0027** | **-0.0021** | **-0.0002** | **0.0020** | **-0.0088** | **-0.0130** | **-0.0176** | **-0.0257** |
| | (0.0196) | (0.0584) | (0.0590) | (0.0505) | (0.0400) | (0.0089) | (0.0350) | (0.0356) | (0.0360) | (0.0329) |
| **-30** | **-0.0022** | **0.1125** | **0.1140** | **0.0515** | **0.0086** | **-0.0008** | **0.0822** | **0.0905** | **0.0782** | **0.0722** |
| | (0.0237) | (0.0630) | (0.0633) | (0.0563) | (0.0413) | (0.0089) | (0.0297) | (0.0305) | (0.0285) | (0.0290) |
| **-20** | **-0.0011** | **0.0368** | **0.0399** | **0.0369** | **0.0469** | **-0.0015** | **0.0347** | **0.0346** | **0.0201** | **0.0013** |
| | (0.0260) | (0.0662) | (0.0668) | (0.0566) | (0.0445) | (0.0088) | (0.0332) | (0.0337) | (0.0328) | (0.0325) |
| **-10** | **-0.0001** | **-0.0007** | **0.0313** | **-0.0025** | **-0.0018** | **-0.0002** | **0.0163** | **0.0207** | **0.0130** | **0.0085** |
| | (0.0287) | (0.0623) | (0.0627) | (0.0537) | (0.0412) | (0.0081) | (0.0338) | (0.0344) | (0.0336) | (0.0330) |
| **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| | (0.0325) | (0.0510) | (0.0627) | (0.0499) | (0.0375) | (0.0117) | (0.0332) | (0.0327) | (0.0326) | (0.0311) |
| **10** | **0.0027** | **0.0896** | **0.0029** | **0.0141** | **0.0160** | **0.0006** | **0.0663** | **0.0661** | **0.0644** | **0.0607** |
| | (0.0324) | (0.0364) | (0.0500) | (0.0302) | (0.0303) | (0.0053) | (0.0290) | (0.0266) | (0.0259) | (0.0209) |
| **20** | **0.0052** | **0.0070** | **0.0946** | **0.0039** | **0.0056** | **-0.0046** | **0.0106** | **0.0161** | **0.0135** | **0.0140** |
| | (0.0318) | (0.0447) | (0.0376) | (0.0432) | (0.0451) | (0.0053) | (0.0290) | (0.0266) | (0.0250) | (0.0230) |
| **30** | **0.0075** | **0.0049** | **0.0077** | **0.5917** | **-0.0023** | **-0.0061** | **0.0122** | **0.0211** | **0.0280** | **0.0309** |
| | (0.0322) | (0.0492) | (0.0457) | (0.0435) | (0.0397) | (0.0043) | (0.0281) | (0.0271) | (0.0252) | (0.0228) |
| **60** | **0.0101** | **0.0411** | **0.0020** | **0.0590** | **0.0619** | **-0.0069** | **0.0417** | **0.0342** | **0.0297** | **0.0215** |
| | (0.0350) | (0.0425) | (0.0412) | (0.0374) | (0.0338) | (0.0017) | 0.0316 | (0.0297) | (0.0288) | (0.0284) |

Note: This table presents PIN, VPIN20, VPIN50, VPIN100 and VPIN200 based on their value on the event date for -/+ 60, 30, 20, 10 days around the event. The full data is separated into two parts: firstly, 1996 to the end of 2007 and secondly from 2008 to the end of 2013. Both PIN and VPIN based on its value on the event date for all sub-periods is high after the event compared with the previous period. In this case, it can be assumed that the PIN and VPIN somewhat detected some informed trading in the ED futures market on the LIBOR event, as these two parameters are high on the event date, then drops to normal. Finally, the VPIN may have better performance than the PIN, it can be seen in this table that all types of VPIN show a higher positive variation then the PIN on the event date. However, there is not a statistically significantly difference for either based on their value on the event date for the whole period.

### 6.5.6 PIN, VPIN and the Maturity Effect

This section contains an illustration of PIN and VPIN before and after expiration to investigate the relationship between the indicators around the maturity date known as the maturity effect. The most striking result from this figure is an aggressive decline after the spike of PIN and VPIN around the maturity date. Both gradually increase and reach a peak within one week of the last trading day. When the ED starts trading again, they continuously decrease to a lower level, although increase later. However, there is some variation between PIN and VPIN, the VPIN continually decreasing over a longer period than the PIN, and finally slightly increasing around thirty days after the event when the ED begins trading again.

Figure 6.9 presents the performance of PIN and VPIN around the maturity date. In this event, I use the EDH0 contract which has expiration on March 17,2010 as an example. This figure is separated into four subplots, first, PIN and VPIN: second, delta PIN and VPIN: third, PIN and VPIN(based on event date) +/-7 days from maturity date: finally PIN and VPIN(based on event date) +/-30 days from maturity date.

On the top panel of the left column, the PIN is seen on the left axis and VPIN on the right for -/+7 days around expiration. There is similar variation between both as the two measurements increase before the maturity date then decrease during the five trading days before expiration. Finally, they bounce back. The highest PIN in this contract appears on March 15, 2010 which is two days to maturity at a value of 0.82 then rapidly decreases to 0.70 on the last trading day. After EDH0 starts trading on the following day, the PIN continuously decreases to around 0.50 and stays constantly low, then finally bounces back five day later.

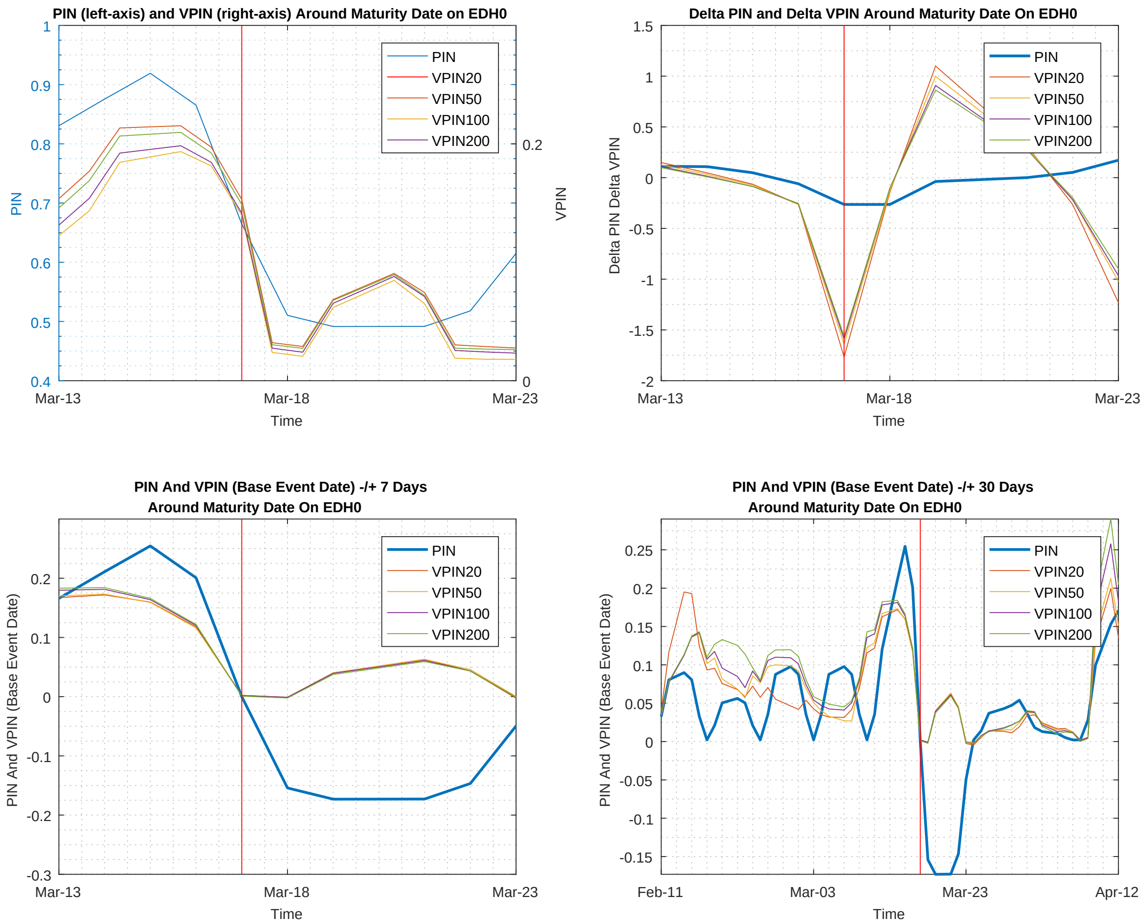



**Figure 6.9: The historical PIN and VPIN of EDH0 with the expiry date.** Note: This figure presents an example of PIN, VPIN, ΔPIN, and ΔVPIN for the EDH0 futures contract. It can be seen from the top left panel that the PIN and the VPIN increase before the maturity date, then rapidly drop around two to three days before the last trading day. Finally, the PIN and VPIN bounce back to normal after the expiration when EDH0 starts trading again on the following day. The top right panel presents ΔPIN and ΔVPIN. This sub-figure illustrates that the differences of VPIN are higher than thatof PIN. A high level of ΔVPIN assumes that the VPIN has a higher degree of sensitivity to this toxicity event than the PIN. The lower panel presents PIN and VPIN based on their value on the event date for -/+7 and 30 days from the maturity date. Both show a strong variation around this event compared to its value on the LIBOR event. Clearly, there is a significant pattern as the PIN and VPIN increase, then reach a peak within the last week of the trading period, and dropping to normal the following day.

Moreover, VPIN is somewhat similar to PIN, the highest VPIN appears around two to three days before expiry at 0.25 then decreases to 0.125 on the last trading day with the lowest at 0.025 between one to two days after the ED starts trading again, and finally bounces back.

The top panel of the right column presents the differences between PIN and VPIN or ΔPIN and ΔVPIN for -/+7 days around the maturity date. Despite, ΔPIN and ΔVPIN having a similar pattern, the ΔVPIN has better performance as it has higher variation than PIN. This subplot shows the latter has a small drop of around 10% as the ΔPIN slides from 0 to -0.02 then bounces back to normal on the following day. However, the VPIN drops around one and a half times within seven days compared to the previous period, as the ΔVPIN falls from 0 to -1.5, then rebounds to a level around two times higher than its lowest point. Additionally, the lower panel of Figure 6.9 presents PIN and VPIN based on the event date for -/+7 and -/+30 days around the event date. Similar to the upper panel, there are similar patterns for both PIN and VPIN. The VPIN performs better than PIN as a signal of market manipulation, however, the PIN shows a higher variation for the manipulation event.

To gain further insight into the effect of maturity date on PIN and VPIN, both have been investigated for a longer period. Figure 6.10 illustrates cross sectional averages across all 40 ED contracts on PIN and VPIN for -/+30 and -/+60 days around the expiry date. Overall, there is a similar pattern to Figure 6.9. The PIN and the VPIN reach a peak within the last week of the trading period, then drop. Finally, PIN and VPIN bounce back to normal after the last trading day when the ED starts trading again. It can be seen from the top left panel of Figure 6.10 that from ten days to maturity, the average PIN increases from 0.67 to 0.725

then drops to 0.70 around two days after the last trading day. Also, similar to the PIN, the VPIN increases from 0.07 to around 0.10 three days before the maturity date, then drops to 0.06 two days after. Finally, the VPIN slightly increases to a normal level.

Top right panel of Figure 6.10 presents the differences of PIN and VPIN or $\Delta$PIN and $\Delta$VPIN for -/+30 days around the maturity date. In this period the $\Delta$VPIN has higher variation than $\Delta$PIN. The $\Delta$PIN fluctuates between -0.01 and 0.01, however, the $\Delta$VPIN fluctuates around 40% between -0.25 and 0.15. The highest $\Delta$VPIN appears on the seven days before the last trading day at 0.15, then gradually drops to the bottom on the expiry date at -0.25. Finally, the VPIN bounces back to normal. Additionally, from seven days to maturity $\Delta$VPIN increases around 25%, from -0.1 to 0.15. Whilst these two parameters have slightly different variation before the maturity date, they have a similar pattern. The $\Delta$PIN and $\Delta$VPIN reach a peak within the last week of the trading period, then rapidly decrease, and finally bounce back after the last trading day when the ED futures start trading the following day.

The lower panel of Figure 6.10 presents PIN and VPIN based on the PIN and the VIN on the event date for +/- 30 and +/-60 days around maturity date. These sub-figures show a similar pattern PIN and VPIN based on the event date and normal PIN and VPIN. Additionally, there is a similar trend between both types . The VPIN is constantly high sixty days before expiry, then reaches a peak at around seven days before expiration, as the VPIN based on the event date shows a positive result before the last trading day. However, later between two to three days to the expiration, PIN and VPIN gradually decrease continuously dropping after the last trading day. Finally, there is a move back to normality. However,

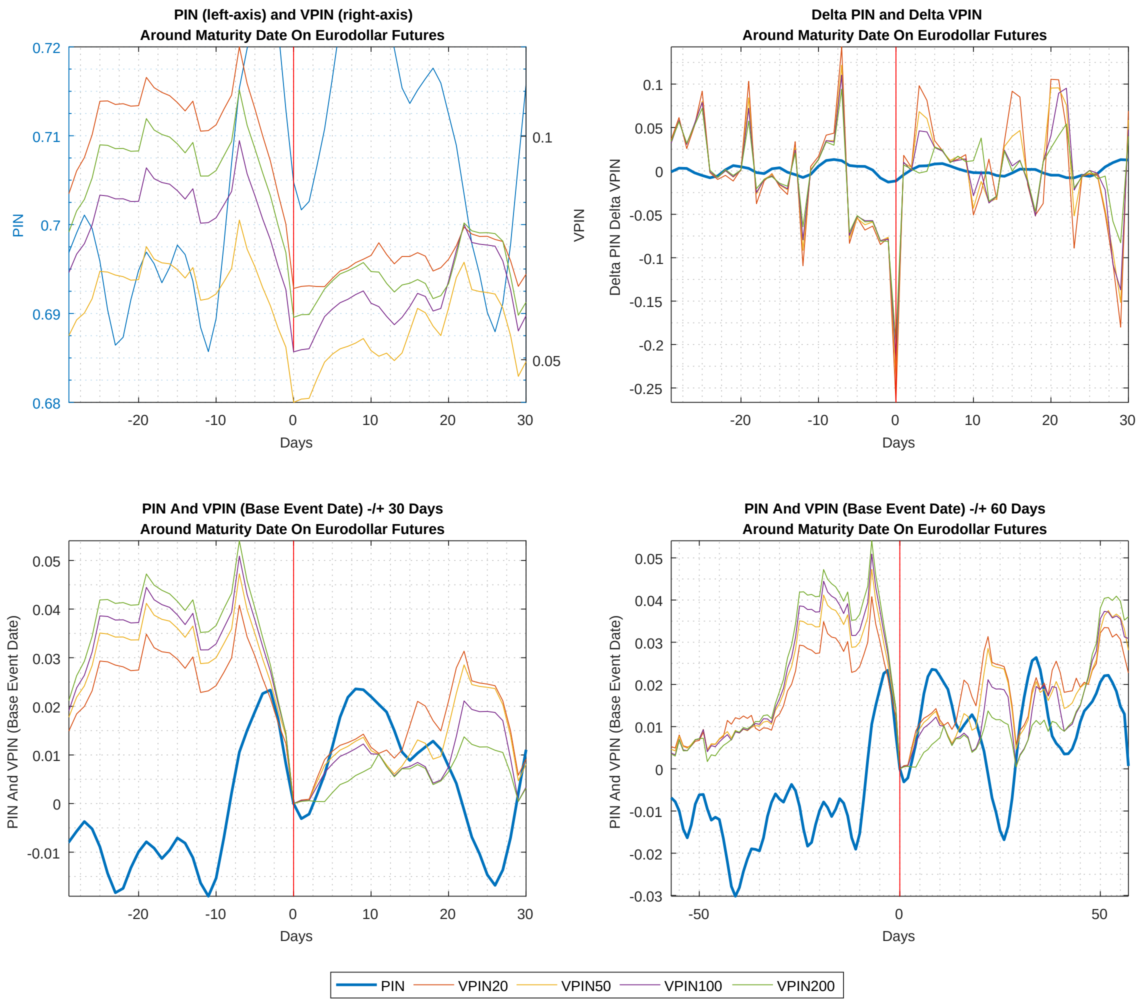



**Figure 6.10: The cross-sectional PIN, VPIN, $\Delta$ PIN and $\Delta$ VPIN across all 40 ED futures contracts with the expiry date.**

Note: This figure presents cross sectional average PIN, VPIN, $\Delta$PIN and $\Delta$VPIN across all 40 ED futures contracts around the expiry date. The red solid vertical line indicates the last trading day on the ED futures contracts. The top left panel presents PIN and four different types of VPIN for -/+30 days around maturity date. During the last week of the trading period, the PIN gradually increases then drops around two to three days until the last trading day. Finally, the PIN and VPIN bounce back. There is a significant pattern between these two parameters. However, the VPIN shows a slightly different pattern from the PIN, as it gradually increases for 30 days before the last trading day, reaching a peak within the last week before the maturity date. The PIN rapidly increases from eleven days to maturity then peaks around two to three days to maturity. Finally, both bounce back after the last trading day when the ED commences trading again.

The top right panel presents the differences of PIN and the differences of VPIN ($\Delta$PIN and $\Delta$VPIN). This figure shows a higher degree of difference on the VPIN($\Delta$VPIN) than the $\Delta$PIN. In this case, I can assume that the VPIN has more sensitivity than the PIN on this event. The lower panel presents PIN and VPIN based on the event date for -/+30 and -/60 days from maturity date. These two plots show a similar result as the top panel. There is a systematic trend between PIN and VPIN based on their value on the event date and the latter is more sensitive to this event than the PIN.

the PIN based on its value on the event date shows a somewhat negative value from 60 to 5 days to maturity. However, they reach a peak within one week before the maturity date, then rapidly drop.

The following section presents the result of the term structure of PIN and VPIN on the maturity effect as presented for the LIBOR manipulation events. Figure 6.11 presents the term structure of PIN and VPIN, also cumulative ΔPIN and cumulative ΔVPIN for the maturity effect. The figure presents individual PIN, VPIN20, VPIN50, VPIN100, and VPIN200 as a dotted line and average PIN and VPIN as a thick line. It also presents 95% range as a 95% confident interval as a thick dotted line. Overall, I present 40 PIN and 200 VPIN for the term structure.

The top panels of Figure 6.11 present cross-sectionals of PIN and cumulative ΔPIN. Overall, there is a small variation on average PIN around the maturity date with a small spike about three days to expiration. This spike has a variation of 2% higher than the average. Despite there being a small scale of variation on average PIN before the maturity date, the individual PIN has a higher variation after the last trading day. The major trend of PIN (dotted lines) is constantly moving around the identical PIN at about 0.67 (thick line) with a small spike and there is the widest variation of PIN at three days to expiration. The lowest PIN is 0.2 and the highest is 0.995, which is nearly 80% variation.

However, after three days to expiration the variation of PIN drops to 50%, with the highest recording at 0.97 and the lowest around 0.30. The variation is wider at the beginning than the previous trading period, this may be consistent with the term structure of PIN (figure 6.4) as it has a high variation from the first trading day; then the variation becomes narrow when the ED futures are nearer to maturity.

The lower panel of Figure 6.11 illustrates VPIN and cumulative of ΔVPIN for -/+60 days from the maturity date across all 40 ED futures contracts. This panel presents not only forty individual VPIN and forty cumulative ΔVPIN from all ED contracts, but each contract also calculates four different types of VPIN and cumulative ΔVPIN presented on a dotted line with the average value in a thick line, and 95% confident interval in a thick dotted line. It can be seen from this figure that there is quite naturally the same pattern of average of VPIN with the PIN. The average VPIN gradually increases from 60 days from the last trading days and reaches a peak at around two days to maturity, then rapidly drops. Finally, the VPIN increases to a normal level after the expiration when the ED start trading again. Also, there is a lower scale of variation on VPIN than PIN, which is notably smaller and more systematic than the PIN. The lowest VPIN is near to zero and the highest is about 0.8, with the average upper bound at about 0.2.

In comparison, the average PIN has a small variation around 0.67 and the average cumulative ΔPIN is near to zero. However, the average VPIN gradually increases from 0.05 to just above 0.20 and the average cumulative ΔVPIN increases from just around zero to one on two days to the maturity date. Finally, VPIN and cumulative ΔVPIN plunges to nearly zero on the last trading day. The average VPIN and the average cumulative ΔVPIN reaches a peak on two days to expiration which increases by around twofold, compared to the previous period. the average VPIN has a higher variation than the PIN and there is a systematic trend between different VPINs. One final concluding remark is that

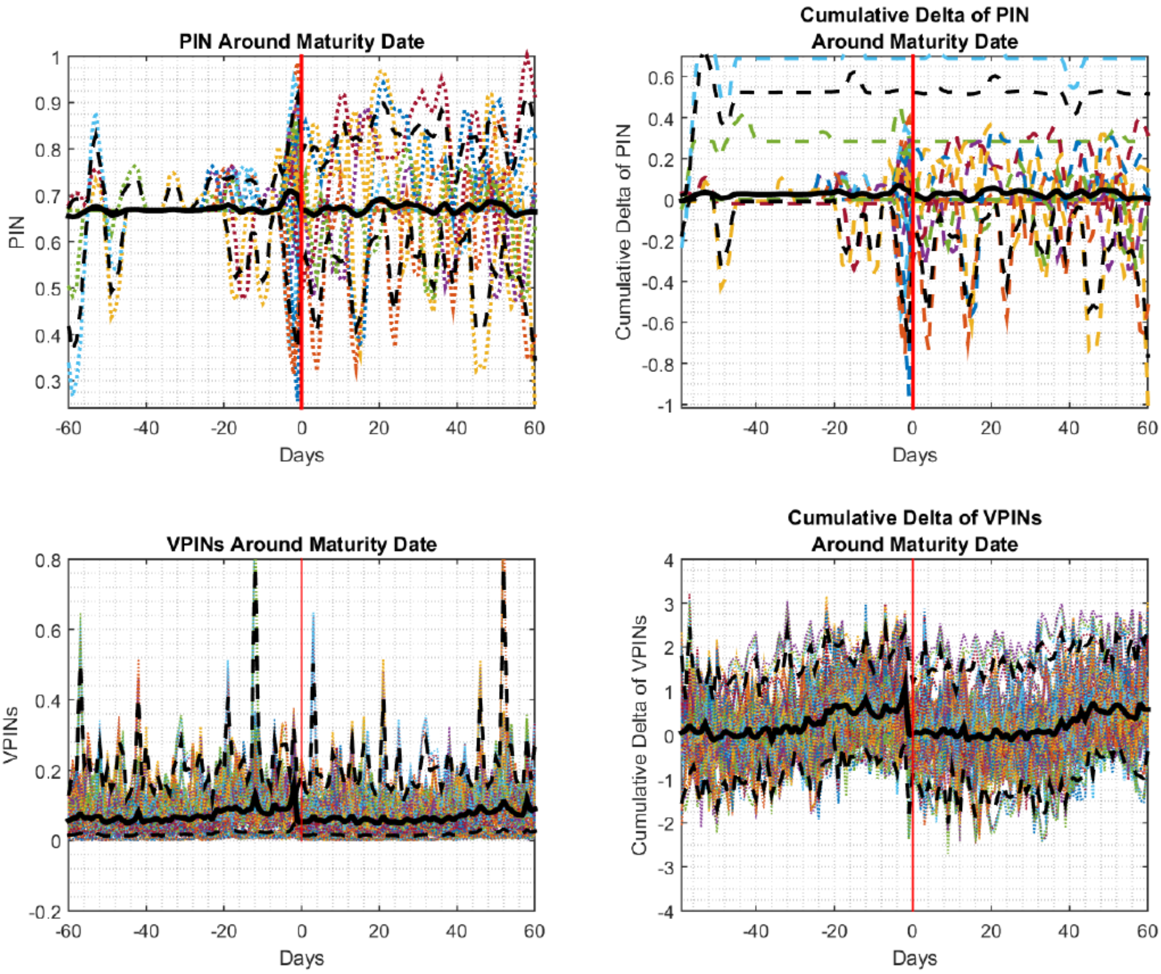



**Figure 6.11: The variation of PIN and VPIN -/+60 days around maturity date**

Note: This figure presents the term structure of PIN, VPIN, cumulative of ΔPIN and cumulative of ΔVPIN with their average value for -/+60 days around maturity date. Overall, there is a systematic pattern between the average PIN and VPIN around this event, as these parameters show a spike on two days to maturity then rapidly drop on the last trading day. Finally, these two parameters bounce back after the last trading day when the ED futures start trading again. However, the lower panel shows that all types of VPIN have a higher scale of variation than the PIN for both before and after the maturity date.

the term structure of ΔVPIN tends to be a 'sine curve'.

After studying PIN and VPIN via the term structure around the maturity date, I then investigate these two parameters from their standard deviation(SD). Table 6.10 reports cross sectional average PIN and VPIN across all 40 ED contracts, constructed for 60, 30, 20 and 10 days for fixed time windows around the maturity dates with their SD. Then, the result of these two parameters is divided into two sub-periods. The first sub-period is between January 1, 1996 and December 31, 2007 and the second sub-period is between January 1, 2008 until December 31, 2013. Moreover, this time period is separated into two parts because there will be different results for PIN and VPIN between these two periods as Figure 5.2 shows the break in the relationship of the bank rank of LIBOR quotes after 2008.

Overall, the variation of PIN is lower than VPIN for both periods as its standard deviation is smaller than the VPIN. However, there is a similar pattern between these two parameters. They both increase before the last trading day, then rapidly drop. Finally, they bounce back after the last trading day when the ED futures starts trading again. The lowest PIN appears on the last trading day (Day 0) at 0.6695 and 0.6552 for pre 2008 and post 2008 samples respectively. The highest appears for 20 and 60 days after the maturity date for pre-2008 and post-2008 respectively. For the VPIN, the lowest appears on VPIN20 for both pre and post 2008, at 0.0471 and 0.0309 respectively. The highest VPIN mostly appears for thirty days to maturity for all different types of VPIN, with the highest of 0.1639 on VPIN200 for pre 2008. This table indicates that there is a similar trend for PIN and VPIN for both periods, however, their variation and standard deviation (SD) are different.

Pre 2008, the SD on PIN and VPIN are slightly different. The SD on PIN shows that the PIN is not statistically significantly different, as the SD varies between 0.02 to 0.05. However, the SD on VPIN during this period is slightly higher than PIN. The highest SD for 10 days after the maturity date is 0.088, 0.089, 0.095 and 0.099 for VPIN20, VPIN50, VPIN100 and VPIN200 respectively. Post 2008, the highest SD for PIN is 0.055 for 60 days after the maturity date. However, the highest SD for VPIN is for 30 days to maturity at 0.108, 0.0833 and 0.085 for VPIN20, VPIN100 and VPIN200 respectively. The highest SD for VPIN20 is 0.127 for ten days to maturity which is the highest SD for all types of VPIN. Finally, there are two remarkable results from this table; first the maturity effect tends to have an impact on informed trading as the PIN and VPIN have shown a similar pattern. They slightly increase for 60 days to maturity, then fall to the lowest PIN and VPIN on the last trading day. Next, the smallest VPIN on this event for both the pre 2008 and post 2008 period, is VPIN20 followed by VPIN50, VPIN100 and VPIN200.

Next, I analyze PIN and VPIN based on their value on the event date for -/+60 days around the maturity which is presented in Table 6.11. The results indicate that the PIN and VPIN around the maturity date are higher than the those on regular trading days. Using the same format analysis as Table 6.10, I find the same result. First, the VPIN has a greater magnitude of variation than the PIN for both pre and post 2008. Secondly, the PIN and VPIN gradually increase for 60 days to maturity then peak within five trading days. After three to two days to maturity, they rapidly drop and finally bounce back. On pre-2008, the highest PIN is for ten and twenty days after the maturity date at 0.0283 or 2.83%, higher than the PIN on the maturity date. Moreover, the highest VPIN

**Table 6.10: PIN, VPIN20, VPIN50, VPIN50, VPIN100, and VPIN200 -/+ 60 days around maturity date.**

| | **Informed trading around maturity date** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Days | PIN -Pre 2008 | VPIN - Pre 2008 | | | | PIN - Post 2008 | VPIN - Post 2008 | | | |
| | | 20 | 50 | 100 | 200 | | 20 | 50 | 100 | 200 |
| **-60** | **0.6926** | **0.1104** | **0.1238** | **0.1335** | **0.1423** | **0.6933** | **0.0753** | **0.0797** | **0.0809** | **0.0841** |
| | (0.0223) | (0.0851) | (0.0785) | (0.0734) | (0.0721) | (0.0324) | (0.0924) | (0.0858) | (0.0764) | (0.0754) |
| **-30** | **0.6933** | **0.1207** | **0.1396** | **0.1488** | **0.1639** | **0.6975** | **0.0911** | **0.0937** | **0.0968** | **0.1027** |
| | (0.0278) | (0.0881) | (0.0817) | (0.0714) | (0.0775) | (0.0400) | (0.1088) | (0.0953) | (0.0833) | (0.0850) |
| **-20** | **0.6926** | **0.1208** | **0.1368** | **0.1459** | **0.1564** | **0.6999** | **0.0833** | **0.0859** | **0.0873** | **0.0924** |
| | (0.0313) | 0.0903 | (0.0836) | (0.0752) | (0.0751) | (0.0433) | (0.1001) | (0.0875) | (0.0758) | (0.0774) |
| **-10** | **0.6930** | **0.1126** | **0.1285** | **0.1375** | **0.1466** | **0.7006** | **0.0749** | **0.0783** | **0.0793** | **0.0839** |
| | (0.0354) | (0.0853) | (0.0796) | (0.0721) | (0.0720) | (0.0459) | (0.0920) | (0.1276) | (0.0699) | (0.0715) |
| **0** | **0.6695** | **0.0471** | **0.0591** | **0.0724** | **0.0827** | **0.6552** | **0.0309** | **0.0384** | **0.0417** | **0.0444** |
| | (0.0368) | (0.0734) | (0.0682) | (0.0662) | (0.0639) | (0.4656) | (0.0767) | (0.0726) | (0.0626) | (0.0616) |
| **10** | **0.6978** | **0.1012** | **0.1176** | **0.1323** | **0.1397** | **0.7066** | **0.0497** | **0.0532** | **0.0531** | **0.0576** |
| | (0.0420) | (0.0888) | (0.0892) | (0.0965) | (0.0998) | (0.0448) | (0.0641) | (0.0586) | (0.0484) | (0.0568) |
| **20** | **0.6979** | **0.0981** | **0.1105** | **0.1222** | **0.1279** | **0.7097** | **0.0427** | **0.0473** | **0.0482** | **0.0507** |
| | (0.0466) | (0.0727) | (0.0740) | (0.0764) | (0.0750) | (0.0477) | (0.0469) | (0.0341) | (0.0359) | (0.0413) |
| **30** | **0.6970** | **0.0931** | **0.1049** | **0.1170** | **0.1223** | **0.7116** | **0.0501** | **0.0556** | **0.0563** | **0.0572** |
| | (0.0488) | (0.0665) | (0.0666) | (0.0700) | (0.0658) | (0.0501) | (0.0522) | (0.0517) | (0.0471) | (0.0489) |
| **60** | **0.6891** | **0.0981** | **0.1106** | **0.1210** | **0.1262** | **0.7157** | **0.0518** | **0.0580** | **0.0574** | **0.0591** |
| | (0.0530) | (0.0593) | (0.0558) | (0.0578) | (0.0537) | (0.0550) | (0.0546) | (0.0547) | (0.0417) | (0.0401) |

Note: This table compares average PIN, VPIN20, VPIN50, VPIN100, and VPIN200, with their standard deviation for -/+60, 30, 20 and 10 days from maturity date. The time period is separated into two periods: the first sub-period is from 1996 to the end of 2007 and the second from 2008 to the end of 2013. Overall, there is small scale variation for both PIN and all types of VPIN before and after the event dates. However, their standard deviation pre-2008 are higher than post-2008. Additionally, VPIN has a higher standard deviation than PIN. It can be assumed that VPIN is more sensitive to the maturity effect than PIN, as it increases before the last trading day then drops dramatically by over 50% on the last day. Finally, there is a similar pattern between these two parameters. However, neither are statistically significantly different during the LIBOR manipulation event, as their standard deviation varies between 0.02 to 0.09.

**Table 6.11: The PIN, VPIN20, VPIN50, VPIN50, VPIN100, and VPIN200 based events date -/+ 60days around maturity date.**

| | | **Informed trading around maturity date** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Days | PIN -Pre 2008 | VPIN - Pre 2008 | | | | PIN - Post 2008 | VPIN - Post 2008 | | | |
| | | 20 | 50 | 100 | 200 | | 20 | 50 | 100 | 200 |
| **-60** | **0.0231** | **0.0633** | **0.0647** | **0.0611** | **0.0596** | **0.0381** | **0.0444** | **0.0413** | **0.0392** | **0.0397** |
| | (0.0223) | (0.0851) | (0.0785) | (0.0734) | (0.0721) | (0.0324) | (0.0924) | (0.0858) | (0.0764) | (0.0754) |
| **-30** | **0.0238** | **0.0736** | **0.0805** | **0.0764** | **0.0812** | **0.0422** | **0.0602** | **0.0553** | **0.0551** | **0.0583** |
| | (0.0278) | (0.0881) | (0.0817) | (0.0714) | (0.0775) | (0.0400) | (0.1088) | (0.0953) | (0.0833) | (0.0850) |
| **-20** | **0.0231** | **0.0737** | **0.0777** | **0.0735** | **0.0737** | **0.0447** | **0.0524** | **0.0475** | **0.0456** | **0.0480** |
| | (0.0313) | (0.0903) | (0.0836) | (0.0752) | (0.0751) | (0.0433) | (0.1001) | (0.0875) | (0.0758) | (0.0774) |
| **-10** | **0.0235** | **0.0655** | **0.0694** | **0.0651** | **0.0639** | **0.0454** | **0.0440** | **0.0399** | **0.0376** | **0.0395** |
| | (0.0354) | (0.0853) | (0.0796) | (0.0721) | (0.0720) | (0.0459) | (0.0920) | (0.1276) | (0.0699) | (0.0715) |
| **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| **10** | **0.0283** | **0.0541** | **0.0585** | **0.0599** | **0.0570** | **0.0514** | **0.0188** | **0.0148** | **0.0114** | **0.0132** |
| | (0.0420) | (0.0888) | (0.0892) | (0.0965) | (0.0998) | (0.0448) | (0.0641) | (0.0586) | (0.0484) | (0.0568) |
| **20** | **0.0283** | **0.0510** | **0.0514** | **0.0498** | **0.0452** | **0.0545** | **0.0118** | **0.0089** | **0.0065** | **0.0063** |
| | (0.0466) | (0.0727) | (0.0740) | (0.0764) | (0.0750) | (0.0477) | (0.0469) | (0.0341) | (0.0359) | (0.0413) |
| **30** | **0.0275** | **0.0460** | **0.0458** | **0.0446** | **0.0396** | **0.0564** | **0.0192** | **0.0172** | **0.0146** | **0.0128** |
| | (0.0488) | (0.0665) | (0.0666) | (0.0700) | (0.0658) | (0.0501) | (0.0522) | (0.0517) | (0.0471) | (0.0489) |
| **60** | **0.0196** | **0.0510** | **0.0515** | **0.0486** | **0.0435** | **0.0605** | **0.0209** | **0.0196** | **0.0157** | **0.0147** |
| | (0.0530) | (0.0593) | (0.0558) | (0.0578) | (0.0537) | (0.0550) | (0.0546) | (0.0547) | (0.0417) | (0.0401) |

Note: This table presents a comparison of average PIN, VPIN20, VPIN50, VPIN100, VPIN200 based on their value on the event date, and also standard deviation (SD) for -/+60, 30, 20 and 10 days from the maturity date. The time period is separated into two periods: the first sub period from 1996 to the end of 2007 and the second is from 2008 to the end of 2013. Overall, there is a similar result to Table 6.10 - a small scale variation before and after the event dates for both PIN and VPIN. Despite the variation of VPIN being higher for the first pre-2008 period, the variation of PIN for post-2008 is higher than pre-2008. Nevertheless, there is a systematic trend between these two parameters as the PIN and all types of VPIN increase before the last trading day then drop dramatically more than 50% on the last day. Finally, they increase to the normal level.

based on the VPIN on the event date appears for thirty and twenty days before the maturity date, at 0.0737, 0.0805, 0.0764 and 0.0812 for VPIN20, VPIN50, VPIN 100 and VPIN200 respectively. For the second sub period, the highest PIN based on the PIN on the event date is for sixty days after the expiration at 0.605 or 6% higher than that on the expiry date. Similar to the first period, the highest VPIN appears for thirty and twenty days before the maturity date at 0.0602, 0.0553, 0.0551 and 0.0583 for VPIN20, VPIN50, VPIN 100 and VPIN200 respectively.

Conclusively, I find the VPIN has a higher scale of variation than PIN except after the maturity date for the second sub period. For the first sub period, the average PIN for sixty days before the maturity date is at 0.023 or 2.3% higher than the last trading day. The average VPIN from sixty days to maturity is at 0.07 or 7% higher than the VPIN on the last trading day, which is around three times higher when compared to the PIN. The clear indication is that both have a similar pattern around the maturity event as they increase then reach a peak before it. Finally, they continuously drop after the last trading day when the ED starts trading again, before bouncing back to normal. Also, the results are in the line with the evidence, as described in the previous analysis section, that the VPIN has more predictability for truly toxicity events than the PIN.

## 6.6 Summary Chapter

This paper takes a comprehensive dataset of the Chicago Mercantile Exchange (CME) tape data, which covers every inside quote and trade from 1996 to 2015 for the 40 LIBOR referenced, quarterly dated Eurodollar futures contracts. First, I apply different types of PIN and VPIN metrics over a variety of estimation windows. I have undertaken an empirical microstructure model of Easley and O'Hara [1992] and Easley et al. [1996] for the PIN and Easley et al. [2011] and Easley et al. [2012] for VPIN. I then have constructed a variety of tests to see if the pattern of PIN and VPIN exhibit structural changes around documented cases of manipulation of the LIBOR reference rate and the maturity event. Finally, I compare this to systematic fluctuations in these measures relative to the futures term structure.

Unsurprisingly, given the scale of the task, therefore, the results are very mixed. Both PIN and VPIN vary systematically and in a statistically significant pattern in respect to the term structure of the futures contracts. PIN varies in a v-shaped pattern, with long (2000 to 3500 days) and short maturity (0 to 500 days) contracts, having significantly higher PIN than intermediate contracts (which are actually the most heavily traded). However, VPIN tends to be a v-shaped pattern from 900 days to maturity to the last trading day. Similar to the PIN, the VPIN on long and short maturity contracts, has significantly higher PIN than intermediate contracts. The former is substantially lower than the latter over the entire range of calculation measures. However, when I move to documented cases of market manipulation in the LIBOR reference rate, the results are ambiguous. There are definitive examples when the PIN and the VPIN shift systematically around a relevant, documented case of a LIBOR manipulation. However, when I build cross sectional averages across events, there is no significant evidence of systematic shifts in either the PIN or VPIN metric. Consistent with Andersen and Bondarenko [2014a] Abad et al. [2015], these findings clearly show that the PIN and VPIN are mostly correlated with trading activity as a result of the term structure analysis. Moreover, they have less predictive power as an early warning signal of market manipulation. It should be noted that whilst I have included every documented case of manipulation directly linked to the relevant reference rates, the list is necessarily incomplete as the regulatory actions have tended to focus on sample charges to the firms involved, rather than documenting every occurrence and its motivation.

The investigation extends to the maturity effect. The results are remarkable and show PIN and VPIN have a significant pattern around this event. They reach a peak at three days to maturity, then drop within the last week of the trading period. Finally, both gradually increase to a normal level after the ED starts trading again.

Another remarkable result in this study is that the cumulative of $\Delta$PIN and the cumulative of $\Delta$VPIN show a clear pattern on the event study, as they spike on the toxicity manipulation event. In line with Easley et al. [2012], the results also show that the cumulative value of $\Delta$VPIN performs better than the normal value in providing information about toxicity events. Therefore, I also apply this method to the PIN. The cumulative of $\Delta$PIN and the cumulative of $\Delta$VPIN show a clear pattern on the event study, as they spike on LIBOR manipulation dates, which are higher than the spike on the normal PIN and VPIN.

Finally, despite VPIN having a higher deviation than PIN and their cumulative differences performing better than the normal value to capture these toxicity events, they are not significantly statistically different for both the LIBOR manipulation and the maturity event. The results may be because the PIN and VPIN approach is unsuitable or because it may have been used incorrectly for this type of data. Alternatively, the LIBOR manipulation and expiration may have only a minor effect. Another remark finding from the higher degree of differentiation and a lower degree of correlation between PIN and VPIN is that the PIN cannot be substituted by the VPIN. However, I leave full investigation of this issue to future work, as more fine detail emerges from the current round of court cases and provides more direct evidence of channels of informed trading.

# Chapter 7

# Term-structure analysis of hidden order in the limit order book: evidence from the E-Mini S&P 500

The hidden order is currently increasingly popular as a standard feature of electronic limit order book markets. The invisible order allows traders to hide all, or partially hide their orders to avoid exposure to risk. I propose a new hidden order detection algorithm for the limit order book to investigate the impact of invisible orders on the market environment using E-mini S&P500 data. The algorithm shows 43% all of the trade volume is involved with invisible liquidity. This work also finds that price impact decreases and market quality is improved with the presence of a hidden order both during high and low-frequency trading periods. I use this measure to study the association between hidden order and other observed market environments. The analysis finds aggressively hidden order activity when trading volume is increased.

## 7.1 Introduction

Hidden liquidity, also known as iceberg order[1], or Max Show in the CME Globeox trading platform[2], is the new standard feature of the electronic limit order book market. Most trading platforms allow entirely invisible or partially invisible orders. As traders can choose whether to make their orders visible or invisible, this inevitably has an impact on market liquidity which will be addressed in some detail later in the chapter. Although market participants who use hidden orders commonly lose time priority to traders who submit displayed orders, this time disadvantage is offset by the secrecy afforded by the hidden orders strategy. However, the more serious cost is that because of the time lag, some hidden orders are not possible to execute. Clearly, hidden orders have both costs and benefits compared to visible orders. However, the advantage of entirely or partially hidden orders is that it reduces the risk of being undercut by aggressive or high-speed traders and this potential loss is always greater than the cost of losses though time priority, especially for agents who want to submit a large order. Using hidden order, agents can lower the incentives for incoming parasite or front runner traders, who quote more competitive prices on the same side of the market, to undercut their order.

There is increasingly popular debated, such as; De Winne and D'hondt [2007b], Hautsch and Huang [2012b] and Buti and Rindi [2013], about the benefits of hidden order and what sort of traders that are likely to use hidden order. Slow traders may use hidden orders as a tool to overcome their speed disadvantage when they compete with high speed traders. Also, uninformed traders who have

[1]As in the wider literature, these terms will be used interchangeably in this paper; hidden order, hidden liquidity, invisible order, non-displayed order, iceberg order (De Winne and D'hondt [2007a], Bessembinder et al. [2009],, Frey and Sandås [2009], Bloomfield et al. [2015]

[2]CME Globex Reference Guide, CME Group, accessed March 7, 2016, <http://www.cmegroup.com/globex/files/GlobexRefGd.pdf>)

no private information on asset values may rationally submit hidden orders to reduce their costs of adverse selection when they trade against informed traders. Strategic traders who use the tactic of entering small marketable orders to fulfilled their large orders use reserve order to avoid risk of position exposure problem. The usage of non-displayed order for this type of traders is to ensure about their profit from being losses to predatory traders such as the HFT traders who use the "Pinging" strategy. Buti and Rindi [2013] and Hautsch and Huang [2012b] also provide empirical evidence that traders use hidden orders to manage picking off risk and the cost of non-execution. They also find that the aggressiveness of hidden order decreases when the HFT traders' pinging strategy that operates on the opposite side of the market increases.

The evolution of hidden liquidity in financial markets has gained momentum in recent years. This growth is surprisingly large in terms of volume. For instance, Aitken et al. [2001] report that 28% of all trading volume on the Australian Stock Exchange is hidden order. De Winne and D'hondt [2007a] show that in 40 stocks belonging to CAC40 index from Euronext venue has a hidden order volume accounting for 39%. Bessembinder et al. [2009] report that on Euronext Paris, 44% of all order volume is hidden. Frey and Sandås [2009] show that the average share of iceberg orders is 9% of all non-marketable orders and 16% of overall volume in German XETRA trading. Pardo and Pascual [2012] find that 18% of trade on the Spanish Stock Exchange involves hidden order executions. Hautsch and Huang [2012b] studied 99 NASDAQ stocks and found that 14% of all trading volume originates from hidden order.

The growth number of hidden liquidity is continually increasing, however, it is important to note that this trading feature existed before the advance of fully electronic trading platforms. Blume and Goldstein [1997] explained in Bloomfield and O'Hara [2000] that the floor brokers in the NYSE sometimes use "not held" orders to allow time to deliberate the order following a customer's instruction. In this case, the order is not put into the book by the floor brokers but instead the "not held" order allows the floor broker time and price discretion in transacting on a best efforts basis. The "not held" orders from floor brokers are currently being replaced with invisible orders in electronic trading platforms in global financial exchange, such as Australian Stock Exchange, Chicago Mercantile Exchange (CME), Deutsch Börse (XETRA), Euronext, NASDAQ, and SWX Swiss exchange.

The issue of cost and benefits to the market is intensely debated, as the hidden liquidity in an electronic order book is associated with a variety of complicated trading strategies which may be profitable. A key benefit is improved liquidity but this comes with increased uncertainty for this liquidity. However, it is complicated to explain how hidden order impacts on financial markets. The benefits of a hidden order ($HD$) are not only for opportunistic traders, but market operators also gain a benefit from this feature. During market stress or when traders are reluctant to reveal their trading positions, $HD$ allows traders to provide some liquidity into the market by submitting a hidden order instead of taking liquidity out of the market. However, this procedure is a real trade-off between liquidity and market transparency. On the negative side, an increase in $HD$ activity also imposes a certain degree of opacity in the market mechanism. This is consistent with the finding of Bloomfield and O'Hara [2000], and Madhavan [2000].

In this analysis, I draw on the Chicago Mercantile Exchange (CME) tape data, a very comprehensive dataset, covering every trade and quote in the limit order book for ten level above (below) best offer (best bid) from 2008 to 2015 for 32 E-mini S&P futures contract. The original data set includes 428,540,484 rows with 4 columns from trades data and 3,797,936,125 rows with 44 columns from limit order book data. This analysis modulates the time-series data into a term-structure style. The term-structure analysis helps us to avoid the bias of the econometric statistical values. This bias is caused by the extreme trade and quote volume and a large number of trades and quotes recorded in limit order book when the futures contract has a short time to expiration. Furthermore, I focus on term-structure analysis, hoping to avoid any seasonal effect that can occur in the time series studies. Also, for futures trading, a key factor that impacts on futures pricing models is time to maturity. Next, a variety of tests are constructed to investigate the association between market quality and hidden order, whilst also investigating the impulse response of hidden order to other observed variables.

Unsurprisingly, given the scale of the task, I find comprehensive results. The portion of hidden order activity in terms of volume during high-frequency trading periods is higher than during low-frequency periods. The highest portion of hidden order volume per day appears in high trading activity periods which are between ten to two weeks before maturity at about 43% of trades volume. The lowest portion is 39% for a low frequency period (see Table 7.2). The association between hidden order and market quality shows similar results for low and high-frequency trading periods. However, the impact of $HD$ is greater during low frequency than high-frequency trading periods, and it is always positive. Furthermore, when I expand the experiment into individual buy and sell hidden order, the result is unambiguous. It is clear that hidden order from both buyers and sellers has improved the quality of the market, the greater impact is from buy hidden orders but the level of impact varies.

The contribution of this analysis is threefold. First, this is the only research study to provide a comprehensive innovation of signed-hidden order detection algorithm for E-min S&P500 for limit order book data. Second, to implement the detection algorithm, this analysis applies the Volume-Weighted Average Price (VWAP) approach with the signed-trade direction indicator to introduce the Volume Weighted Average Price-Trade Direction (VWAPTD) indicator. Last, the empirical application of a term-structure analysis is a new contribution to the literature in the field, and my algorithms are available for other researchers to implement in such studies in a different market setting. Additionally, the advantage of the algorithm is that it can be constructed from publicly available data, therefore, it does not rely on special data.

The empirical study begins by systematizing E-mini S&P500 time series data in the form of the term-structure. After this, the E-mini data is separated into 18 sets of data, which is from eighteen to one week to expiration. From this term-structure, I find the E-mini trading becomes highly active from fourteen weeks until two weeks to maturity. Surprisingly, from fifteen to fourteen weeks, the trade updated jumps from 0.64 to 2.16 million updated, results for a period longer than fifteen weeks to maturity, showing that trade updated declines continuously. For instance, in sixteen, seventeen and eighteen weeks to maturity, the average trade updated per week is 0.019, 0.008, and 0.005 million updated respectively. These very small numbers allow us to focus only on the data from eighteen weeks to maturity. Next, using the new proposal of a combination between VWAPTD and a hidden order algorithm to classify hidden order as buy-initiated or sell-initiated hidden order, I can investigate the price impact from both buy and sell hidden orders. From this, the findings show a similar pattern of price impact for both buy and sell hidden order. One remarkable result from this analysis is that the price impact significantly drops after the hidden order is submitted. This result is consistent with Harris [1996], Aitken et al. [2001], Anand and Weaver [2004], De Winne and D'hondt [2007a], and Frey and Sandås [2009].

Then, the study moves to investigate the association between two sides of hidden order and the traditional market quality measures, finding the benefit of invisible liquidity to the market illustrated through a negative relationship between the hidden order and effective spread, realized spread and the price impact. However, surprisingly, the widening spread occurs when hidden orders are submitted onto the market. Parlour [1998], Buti and Rindi [2008] explained this spectacle by the behavior of large traders who normally prefer to use hidden order to reduce their exposure to being undercut by predatory traders by posting their price away from the midpoint to convince new entrants to join the queue at prices far from the best bid or best ask. At this point, I extend the empirical study by employing a signed-hidden order volume to investigate the relationship between hidden order and other interested variables by using Vector Auto Regression (VAR) and impulse response function (IRF) analysis. For the VAR analysis, I find traditional results from quote revision studies as trades are serially correlated with the same direction of previous trade and quote mid-point increase (decrease) after buy (sell). For the signed-hidden order ($\widetilde{HDV}$) equation in a VAR system, I find the greatest impact measured by the estimated coefficient is from signed-trade volume ($\widetilde{TV}$), and the lowest is from lagged signed-hidden order. Also, from this data set, the results from Granger's causality test shows there is a two-way effect in my VAR system. Furthermore, the result of an impulse response analysis shows that the response of $\widetilde{HDV}$ following a shock of $\widetilde{TV}$

is greater than the response of $\widetilde{HDV}$ to the shock of quote return and signed-trade direction. Finally, from the IRF analysis, the results indicate that hidden liquidity has an immediate positive impulse response caused by an innovation in signed-trading volume ($\widetilde{TV}$) followed by a decrease with an exponential decay function. Also, the impulse of $\widetilde{HDV}$ response to the $\widetilde{TV}$ is higher than the response to $\widetilde{QR}, \widetilde{TD}$, and $\widetilde{HDV}$.

The remainder of the study is organized as follows. First §.7.2 outlines the debate on hidden order, in the academic literature, with regard to the impact of hidden order on financial markets. §.7.3 presents the adaptation of the new approach of signed-hidden order detection algorithm and demonstrates how it works. I also introduce a classic version of market quality measurement and the association of hidden order to other observed variables. Subsequently §.7.4 presents the analysis of the study dataset. Finally, §.7.5 summarizes the key findings and presents some brief concluding comments and directions for future research.

## 7.2 Literature review

The research complements the theoretical and empirical literature of the market microstructure on the electronic limit order book and hidden liquidity. This study uses limit order book data that I believe offers a more comprehensive data set than trades and quotes data. The main reason for using limit order book (LOB) is the pre-trade transparency, as the LOB has a higher degree of the transparency than market order (Madhavan et al. [2005], Baruch [2005] and Boulatov and George [2013]). Passive traders use the limit order book to minimize the transaction cost, as the order can be executed at a better price. Also, strategic traders participate in the limit order book to signal trading intention to other market participants by posting flickering orders to attract a counterparty to increase the execution probability and reduce the time of posted position. However, Harris [1996], Hautsch and Huang [2012b] illustrate that traders who use this strategy may be faced with adverse effects such as: adverse price reaction (Hautsch and Huang [2012a]), the revoking of defensive market order in the face of a larger displayed order in LOB (Moinas [2010]), the strategic gain of parasitic traders over traders who submit a large visible order to LOB (Harris [1996]). These adverse effects of limit order book can be relieved by submitting a partially or fully invisible order (Buti and Rindi [2013]). For example, Esser and Mönch [2007] reveals that large position traders can optimize their order by submitting a fraction of hidden orders to maintain their order continuously and balancing exposure risk against execution risk.

Instances of hidden liquidity in financial markets are increasingly popular, and I provide some evidence in the following section. Aitken et al. [2001] show 28% of trading volume on the Australian Stock Exchange is hidden. They also find a similar price impact between non-displayed orders and displayed orders. De Winne and D'hondt [2007a] show that for 40 stocks belonging to CAC40 index from the Euronext venue, hidden orders accounted for 39%. Moreover, they find a positive relationship between hidden orders and order aggressiveness on the opposite side. Bessembinder et al. [2009] report that 44% of order volume submitted in Euronext Paris is hidden liquidity; they also document that the hidden order reduces trading costs but increases execution time. Frey and Sandås [2009] shows that the average share from hidden orders is 9% of all non-marketable orders and 16% of all trading volume in the German XETRA exchange. They also find that in this market, the price impact of the hidden order depends on its size, also, the price impact increases with a higher probability of hidden order. Pardo and Pascual [2012] find that 18% of trade volume in the Spanish Stock Exchange is involved with hidden liquidity and their results show no significant relationship between hidden order and price impact. A study of 99 NASDAQ stocks (Hautsch and Huang [2012b]) find that 14% of trading volume comes from hidden orders.

Currently, empirical studies of hidden liquidity have been increasing significant. Generally, most are related to the association between hidden liquidity and market quality (Aitken et al. [2001]). The empirical findings from Bessembinder et al. [2009] and Harris [1996]) show that traders prefer to hide their trading position and size of order when the tick size is small and the order size is large. Tuttle [2003] shows that market depth significantly increased after a hidden order procedure was introduced on NASDAQ, and furthermore this invisible order had more predictive power than a visible order for future price movement for this market. He further found the presence of quote revision after the hidden order is submitted.

To further investigate the association between hidden order and market quality, this study looks at the impact of hidden liquidity as a function of signed-hidden order volume and other market properties, such as quote return, signed-trade direction, and signed-trading volume. There are, so far, only a few studies of market impact and hidden order. The main reason is that, to study the hidden order, a researcher has to identify this invisible liquidity from trading accounts which require a special dataset that contains this information. These unique data sets are difficult to obtain (Torre [1997], Almgren et al. [2005]). However, to study hidden order by using publicly available data, Moro et al. [2009] and Vaglica et al. [2008] propose a signed-hidden order detection algorithm developed from Bernaola-Galván et al. [2000]. They find a concave function of temporary price impact from intraday hidden liquidity. Pardo and Pascual [2012], using their hidden order algorithm to study the Spanish Stock Exchange, find evidence of no significant differences in price impact between non-displayed order compared and displayed order.

Another area of hidden liquidity study is the market reaction to the presence of hidden order. For example, De Winne and D'hondt [2007a] demonstrate that traders are more aggressive when they find a signal of hidden order on the opposite side of the market. Frey and Sandås [2009] find a negative relationship between price impact and a fraction of hidden liquidity and Pardo and Pascual [2012] find no significant relationship between hidden order and return or volatility. Even though most empirical studies focus on hidden order motivated by market liquidity, there are some papers focusing on the information content of invisible orders (Biais et al. [1995], Griffiths et al. [2000], Bisière and Kamionka [2000], Ranaldo [2004], Beber and Caglio [2005], and Veredas and Pascual [2004]).

To understand the behavior of traders who submit hidden liquidity, Gozluklu et al. [2009] and Bloomfield et al. [2015] set up an experimental market in laboratory studies to examine how hidden order activity affects traders' behavior and the market outcome. They find no significant difference in the market outcome between traders who use hidden liquidity and traders who use displayed order. Also, the result shows that both informed and uninformed traders tend to use invisible orders when they detect an increase in trading aggressiveness.

## 7.3 Methodology

### 7.3.1 Detecting Hidden Orders

This study uses the E-Mini S&P 500 data from the CME Globex electric trading platform. On the Globex, price and volume are normally submitted through the limit order book (LOB). Traders who seek immediate execution need to price the limit order to be marketable, a buy (sell) order needs to be priced at or above (below) the prevailing ask (bid) price. Also, traders who prefer to price their order at a specific price and volume can place their order in the limit order book then the order is waiting to be executed. However, this order might not be executed if the price cannot be matched during the period of time in which the order is left open.

The Globex provides a 10-deep level of limit order for their futures products with a display quantity system. This system allows traders to decide whether to display all or partially display their order in the limit order book. In the limit order book, displayed order is shown to the market participants on their trading screen and the reside orders in the book are invisible to all traders. However, the execution priority is as follows: price, displayed order, and time, so at the same price all visible orders are executed before invisible orders. For example, a trader wants to buy 100 futures contracts with a hidden order, in this case, the trader may separate his buy order into 2 different orders: 10 contracts of a displayed order and 90 contracts of a hidden order. Therefore, no more than 10 buy contracts are exposed to the market and the remainder is booked at the bottom of the book but does not appear on the trading screen.

Since Globex provides a non-display system, in which the invisible or hidden order is not marked on the customer trading screen, traders cannot directly identify hidden liquidity or this type of order as it is not publicly available. Nonetheless, to solve this problem, I have developed a detection algorithm based on a combination of Vaglica et al. [2008] and Pardo and Pascual [2012]. Vaglica et al. [2008] and Moro et al. [2009] who developed their hidden order proxy from Bernaola-Galván et al. [2000] in which the algorithm is working with signed-traded volume. I then adopted this signed proxy with Pardo and Pascual [2012] algorithm. Therefore, the detection algorithm is carried out using the following steps S1-S6.

Step 1: First, the limit order book (LOB) and trade files are matched.

Step 2: Next, trade is classified as a buy or sell by using volume weighted average price trade direction (VWAPTD) with +1 for a buy trade and -1 for a sell trade; this trade classification is explained in the following section.

Step 3: At this point, if the trade is classified as buy-initiated then the algorithm compares the trade reported size with the corresponding updates volume (changes in the accumulated volume) on offer or bid side in the LOB.

Step 4: If the trade is classified as a sell-initiated, then the algorithm compares the trading volume with the corresponding updates volume on the ask side.

Step 5: To infer the volume of hidden order or invisible order, the trade volume is compared with the volume update in LOB. If the trade size is larger than the corresponding updates volume, a deviation between these two volumes can only be explained by the presence of invisible or a hidden volume.

Step6: However, for this algorithm, if the corresponding updates volume in LOB is positive or larger than the reported trade size, the algorithm classifies this as a modification order.

**Table 7.1: Examples of the procedure of hidden order detection algorithm for E-mini S&P 500 on August 10, 2014 (ESU4).**

| | | Trades | | Volume Updated | | Invisible Volume | |
|---|---|---|---|---|---|---|---|
| Time | B/S | Price | Volume | Bid | Offer | Bid | Offer |
| 15:49:33.706 | Buy | 1914.75 | 2 | 0 | 1 | 0 | 0 |
| 15:49:33.706 | Buy | 1914.75 | 2 | 0 | 3 | 0 | 0 |
| 15:49:33.707 | Buy | 1914.75 | 3 | 0 | 2 | 0 | 0 |
| 15:49:33.707 | Buy | 1914.75 | 2 | 0 | -1 | 0 | 0 |
| 15:49:33.707 | Buy | 1914.75 | 1 | 10 | 0 | 0 | 0 |
| 15:49:33.707 | Buy | 1914.75 | 2 | 0 | -55 | 2 | 0 |
| 15:49:33.707 | Buy | 1914.75 | 16 | 19 | -15 | 0 | 0 |
| 15:49:33.707 | Sell | 1914.75 | 2 | -960 | 841 | 0 | 0 |
| 15:49:33.707 | Sell | 1914.75 | 1 | 63 | -8 | 0 | **9** |
| 15:49:33.707 | Sell | 1914.75 | 5 | 20 | -250 | 0 | **255** |
| 15:49:33.707 | Sell | 1914.75 | 50 | 131 | -102 | 0 | **52** |
| 15:49:33.707 | Sell | 1914.75 | 7 | 2 | 0 | 0 | 0 |
| 15:49:33.707 | Sell | 1914.75 | 1 | 3 | 0 | 0 | 0 |
| 15:49:33.718 | Sell | 1914.75 | 1 | 0 | 0 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 1 | -2 | 0 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 1 | 1 | -1 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 50 | 0 | 1 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 1 | 2 | -3 | 0 | **2** |
| 15:49:33.770 | Buy | 1915.00 | 2 | -8 | 0 | **6** | 0 |
| 15:49:33.770 | Buy | 1915.00 | 5 | 4 | -9 | 0 | 0 |
| 15:49:33.770 | Buy | 1915.00 | 3 | 0 | 2 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 1 | 13 | -20 | 0 | **19** |
| 15:49:33.770 | Sell | 1914.75 | 19 | -31 | 16 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 4 | -46 | 24 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 8 | 40 | 21 | 0 | 0 |
| 15:49:33.770 | Sell | 1914.75 | 1 | 0 | 0 | 0 | 0 |
| 15:49:33.804 | Sell | 1914.75 | 1 | 27 | -2 | 0 | **1** |
| 15:49:33.900 | Sell | 1914.75 | 1 | 31 | -33 | 0 | **32** |
| 15:49:33.900 | Sell | 1914.75 | 2 | 0 | 0 | 0 | 0 |
| 15:49:33.900 | Sell | 1914.75 | 10 | 0 | 0 | 0 | 0 |
| 15:49:33.900 | Sell | 1914.75 | 7 | 1 | 0 | 0 | 0 |

Note: This table presents an example of the procedure of the hidden order detection algorithm that was taken from the activity in the E-mini futures on August 10, 2014 (ESU4). The table reports order activity starting around 15:49:33.706 pm and ending at 15:49:33.900 am. Shading identifies hidden order volume corresponding between 1 and 255 volumes for bid and offer. In total, during this roughly 300 millisecond period, there are 9 volumes of hidden order from the buyer and 370 volumes from the seller.

Since traders cannot observe hidden orders as it is invisible on their screen, one might wonder how my algorithm can detect this type of order. It is instructive to present the particular detection algorithm that I believe expresses the volume of hidden order. Table 7.1 illustrates the procedure of the hidden order detection algorithm from E-mini S&P500 on August 10, 2014 (ESU4) beginning at 15:49:33.706 and ending at 15:49:33.900. Over this period, there were 31 trade transactions' updated, and I highlight in gray some of the hidden order during this period to make it easier to scrutinize how the algorithm engaged in this type of order. After trade data and the LOB data were matched, the algorithm classified trade as a buy or sell, then I recorded the corresponding volume updates or changes in accumulated volumes in a limit order book for both bid and offer. From this table, at 15:49:33.707, after a trade is classified as a sell-initiated with one unit of trading volume, then this trading volume is compared with the volume updated on offer side in the limit order book. At this point, if the corresponding volume updated is positive, the algorithm classifies this trade as a modification. However, if the corresponding volume is negative, which means the liquidity in the limit order book is consumed, then the algorithm identifies this trade with the presence of invisible liquidity. In this case, the corresponding updated volume on the sell side or offer is compared with the trade volume, then the difference between 1 from trade volume and -8 volume updated can only be explained by the presence of invisible liquidity, which is 9 hidden order volumes.

With this hidden order detection algorithm, I can identify the hidden liquidity in the limit order book for E-mini S&P500. Then, I report this hidden volume as a term-structure from eighteen to one week to maturity. Table 7.2 provides the results of hidden liquidity on E-mini S&P5000 limit order book, regarding respectively the mean, maximum and minimum of the hidden order from bid and

Table 7.2: **The number of hidden volume from bid-ask side for E-Mini S&P500 futures reported by number of week to maturity from July 2, 2008 to June 19, 2015.**

| Number of Week to Maturity | Mean ask hidden order volume | Max ask hidden order volume | Min ask hidden order volume | Mean bid hidden order volume | Max bid hidden order volume | Min bid hidden order volume |
|---|---|---|---|---|---|---|
| 1 | 313,748 | 1,049,860 | 2,505 | 297,364 | 1,112,845 | 3,660 |
| 2 | 2,791,823 | 6,592,515 | 10,370 | 2,776,721 | 6,464,745 | 10,035 |
| 3 | 3,831,544 | 6,963,570 | 16,800 | 3,897,564 | 7,443,275 | 16,810 |
| 4 | 3,179,783 | 6,766,500 | 6,220 | 3,237,435 | 6,837,425 | 14,705 |
| 5 | 3,145,950 | 6,422,880 | 10,120 | 3,198,121 | 6,746,140 | 13,740 |
| 6 | 3,444,619 | 6,636,360 | 6,860 | 3,477,611 | 7,115,870 | 9,535 |
| 7 | 3,596,351 | 8,602,990 | 7,550 | 3,638,719 | 8,862,410 | 8,885 |
| 8 | 3,555,755 | 6,749,675 | 12,515 | 3,622,222 | 7,525,500 | 10,935 |
| 9 | 3,348,175 | 7,404,300 | 9,290 | 3,407,632 | 8,360,245 | 10,685 |
| 10 | 3,484,405 | 7,150,105 | 10,165 | 3,535,457 | 7,373,200 | 12,285 |
| 11 | 3,255,571 | 5,444,755 | 11,480 | 3,297,365 | 5,563,050 | 11,165 |
| 12 | 2,563,291 | 6,303,800 | 5,300 | 2,600,693 | 6,073,980 | 3,855 |
| 13 | 3,049,775 | 6,376,945 | 7,205 | 3,051,061 | 6,551,260 | 3,420 |
| 14 | 3,250,240 | 6,852,730 | 6,285 | 3,306,488 | 7,026,035 | 8,110 |
| 15 | 1,249,334 | 5,713,315 | 330 | 1,259,627 | 5,564,375 | 170 |
| 16 | 118,136 | 3,173,330 | 540 | 118,187 | 3,149,560 | 195 |
| 17 | 6,713 | 34,230 | 130 | 6,278 | 30,200 | 105 |
| 18 | 5,872 | 17,310 | 1,445 | 5,947 | 18,420 | 1,475 |

Note: This table presents average number of quotes volume within limit order book, trade volume, quotes updated and trade updated from July 2, 2008 to June 19, 2015.

offer. Furthermore, the last column of this table compares a portion of hidden order volume with trade volume. Over the sample period, in an average term, the result shows 42% of trades executed on E-mini S&P500 involved with hidden liquidity, the minimum at 39% and the maximum at 43%. For high trading activity periods (between fourteen to two weeks to maturity) the hidden order is 43% and 39% for low trading activity period. Also, the average hidden order on the bid is larger than the offer for most of the sample periods. The results of the hidden order in E-mini S&P500 is higher than reported in other research such as: 28% in the Australian Stock Exchange (Aitken et al. [2001]), 16% in the German XETRA(Frey and Sandås [2009]), 18% in the Spanish Stock Exchange(Pardo and

Pascual [2012]); however, it is similar to De Winne and D'hondt [2007a] with 39% in the CAC40 and 44% in Euronext(Bessembinder et al. [2009]).

The figure 7.1 presents Number of Hidden Order on the E-Mini S&P 500 between July 2008 to June 2015. It can be seen from the upper sub-figure that the average daily number of hidden orders jumps from around 50,000 orders per day between 2008 and 2009 to around 150,000 orders per day after 2010, until 2015. The lower sub-figure presents an average weekly number of hidden orders presented as a term-structure for eighteen to one week to maturity. The highest number of hidden order appears on week seven at about 800,000 orders follow by 798,000 orders for week eight, the lowest is 713 orders for week eighteen. Interestingly, it can be seen from this sub-figure that the number of hidden order is plunged from around 588,000 orders in week two to 60,000 orders for the last week of the trading period. These hidden orders are consistent with the number of hidden orders that jumped from about 200,000 for week fifteen to about 680,000 for week fourteen. This pattern can assume that the strategic trader uses hidden order roll over futures contracts to switch from the front month contract that is close to maturity, which is around two weeks to expiration, to another contract in a further-out month which is around fourteen weeks to expiration.

### 7.3.2 Liquidity, Market Quality

This chapter is divided into three parts with three different models; in section 7.3.2.2 the OLS model is modified from Hasbrouck and Saar [2013]; in section 7.3.2.3 the vector auto-regression (VAR) model; and in section 7.3.2.4 the impulse response analysis.

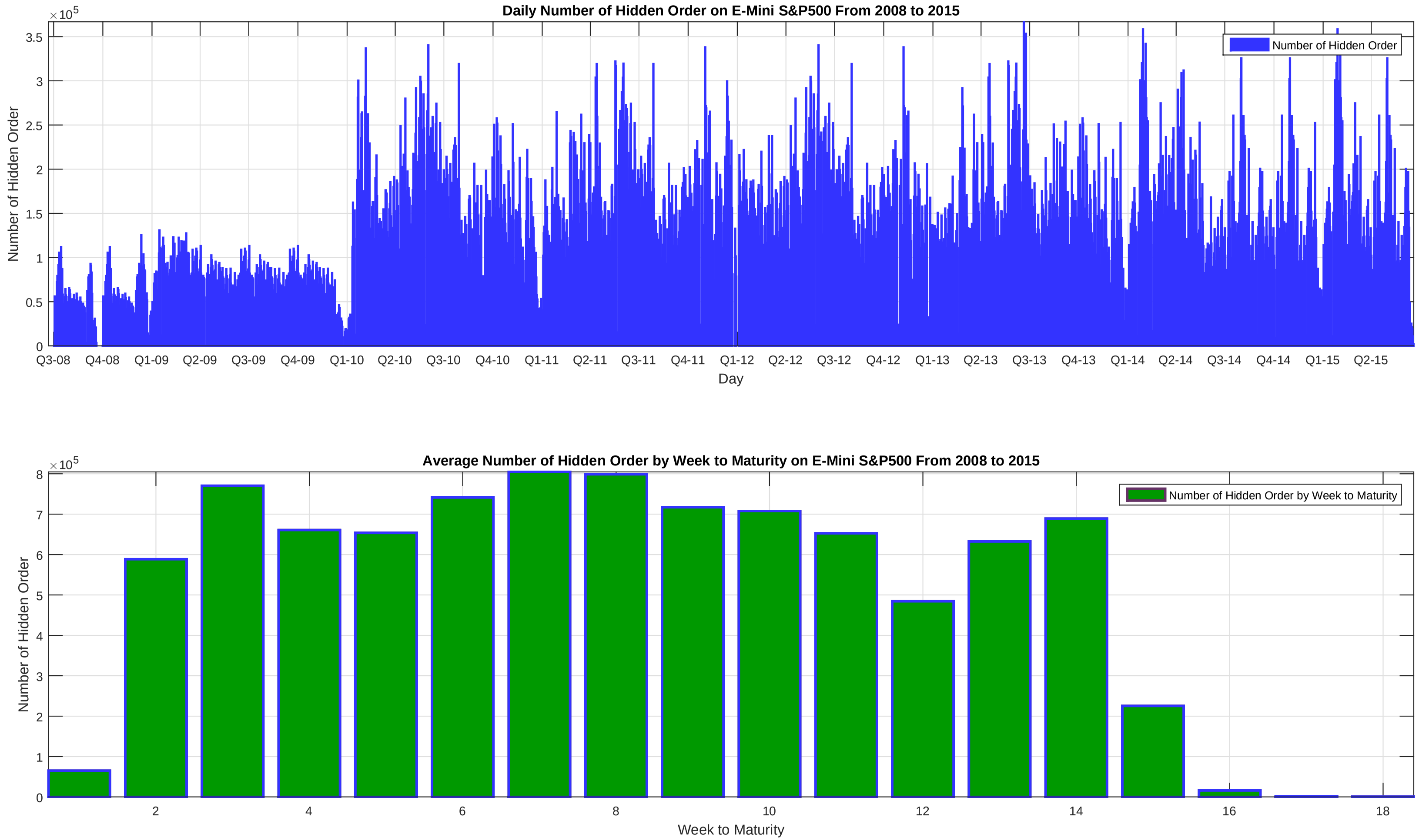



**Figure 7.1: Number of Hidden Order from Trade data on the E-Mini S&P 500 From 2008 to 2015.**
Note: This figure presents the number of hidden orders on the E-Mini S&P 500 from 2008 to 2015. The upper sub-figure presents a daily number of hidden orders and the lower sub-figure presents an average number by week from eighteen to one week to maturity.

#### 7.3.2.1 The Volume-Weighted Average Price Trade Direction (VWAPTD)

The hidden order allows market participants to submit their orders partially or totally invisibly, so traders can choose to display only a small fraction of their order simultaneously. Copeland and Galai [1983] explain that agents who engage with a hidden limit order are normally believed to be large liquidity traders who need to reduce their exposure risk. These traders submit hidden orders to protect themselves from front-running traders, undercutting and stealing price priority in the knowledge that the price will rise or fall when the large order is submitted to the market (Pardo and Pascual [2012],Harris and Panchapagesan [2005]). In addition to the benefit to large liquidity traders, hidden orders also benefit the financial market. To clearly understand this advantage, we investigate a relationship between hidden order and liquidity risk or market quality by using quoted spread, effective spread, realized spread and price impact, all of which are measured based on volume-weighted average price (VWAP) from both bid and offer in a limit order book. Berkowitz et al. [1988] explain the VWAP is a ratio between traded value multiplied by traded price to total traded volume over a particular time horizon. However, we adopt this VWAP into LOB data with a ten tick level above and below the last bid and offer price. We explain this measure as the average price at which a stock is quoted over the quoting horizon from ten levels of LOB, so we calculate VWAP for both bid and ask side of LOB. The VWAP is often used as a benchmark and a good approximation of expected price by passive investors and is supported by Madhavan [2002] and Li and Ye [2013].

To illustrate VWAP for bid and ask price, we give the following example. Assume that at time $t$ on a trading day, traders submit sell orders in a limit order book. For level 1(L1) is 5 volume, L2 is 10, L3 is 12, L4 is 12, L5 is 10 and the price is 98, 98.5, 99, 99.2 and 99.5 for L1 to L5 respectively. Thus, the VWAP for sell ($VWAP_{ask_t}$) is equal to:

$$\frac{(98 \times 5) + (98.5 \times 10) + (99 \times 12) + (99.2 \times 12) + (99.5 \times 10)}{5 + 10 + 12 + 12 + 10} = 98.95$$

For buy orders in limit order book at level 1(L1) it is 6 volume, L2 is 10, L3 is 13, L4 is 12, L5 is 10 and the price is 97, 96.5, 96, 95.5 and 95 for L1 to L5 respectively. Thus, the VWAP for buy ($VWAP_{bid_t}$) is equal to:

$$\frac{(97 \times 6) + (96.5 \times 10) + (96 \times 13) + (95.5 \times 12) + (95 \times 10)}{6 + 10 + 13 + 12 + 10} = 96.84$$

In general, if we let $V_{j,t}$ and $P_{j,t}$ represent the submitted volume and submitted price at time $t$ at level $j$ in limit order book, then the VWAP from bid(B) and ask(A) are formulated as:

$$VWAP_{bid_t} = \frac{\sum_{j=1}^{n} PB_{jt}.VB_{jt}}{\sum_{j=1}^{n} VB_{jt}}. \tag{7.3.1}$$

where, $n$ is the number of level on limit order book, $PB_{jt}$ is bid price at $j$ level at time $t$, $VB_{jt}$ is bid volume at $j$ level at time $t$, and ask(A) are formulated as:

$$VWAP_{ask_t} = \frac{\sum_{j=1}^{n} PA_{jt}.VA_{jt}}{\sum_{j=1}^{n} VA_{jt}}. \tag{7.3.2}$$

where, $n$ is the number of level on limit order book, $PA_{jt}$ is ask price at $j$ level, $VA_{jt}$ is ask volume at $j$ level both occur at time $t$. Furthermore, this work uses $VWAP_{bid_t}$ and $VWAP_{ask_t}$ to calculate mid price at time $t$ equal to:

$$mid_t = \frac{VWAP_{ask_t} - VWAP_{bid_t}}{2}. \tag{7.3.3}$$

Then, trade classification work with the requirement of two sets of data. First, mid price from $mid_t = \frac{VWAP_{ask_t} - VWAP_{bid_t}}{2}$, and second data is trade price. So, this method classifies a trade as a buy when it occurs above the mid price and conversely, it will be classified as a sell when it occurs below the mid price. When the trade appears at the mid price it is unclassified.

Next, I examine the liquidity and market quality measured by quoted spread, effective spread, realized spread and price impact. I follow the VWAP approach to calculating these measurements. The quoted spread is the difference between $VWAP_{ask_t}$ and $VWAP_{bid_t}$ divided by $mid_t$, and the effective spread is the difference between the midpoint from $VWAP_{ask_t}$ and $VWAP_{bid_t}$ and the trade price $p_t$. $EffSpread$ is used to measure the market condition because it captures liquidity flow. For instance, the wider effective spread means the market is in a condition of less liquidity and also the market has less quality. Hendershott et al. [2011] recommends this spread for traders or institutions who want to trade at the inside quote because the effective spread is more meaningful as a transaction cost at point of execution than the normal spread. Therefore, we can define the proportion of quoted spread as,

$$Spread_t = \frac{VWAP_{ask_t} - VWAP_{bid_t}}{mid_t} \tag{7.3.4}$$

and effective spread defined as:

$$EffSpread_t = \frac{TD_t(p_t - mid_t)}{mid_t}. \tag{7.3.5}$$

where $TD_t$ is signed-trade direction indicator that equals 1 for buy-initiated and -1 for sell-initiated.

The $EffSpread$ can be decomposed into two components. First, realized spread ($RSpread$) represents theoretical profits of a liquidity provider or a cost for liquidity suppliers who want to keep their position at the midpoint after a trade. The $RSpread$ is defined as

$$RSpread_t = \frac{TD_t(p_t - mid_{t+1})}{mid_t}. \tag{7.3.6}$$

where $p_t$ is trade price, $TD_t$ is trade indicator, and $mid_t$ is midpoint from VWAP approach.

The second component is the adverse selection measured by price impact. The $PImpt$ shows how much mid-price moves within a given look-ahead window. Also it, can be represented as the liquidity suppliers lost to informed traders due to adverse selection (Chakrabarty et al. [2014]). This work uses the next midpoint instead of time windows. So, $PImpt$ is defined as:

$$PImpt_t = \frac{TD_t(mid_{t+1} - mid_t)}{mid_t}. \tag{7.3.7}$$

This $EffSpread$, $RSpread$, and $PImpt$ is a standard measurement of market liquidity and market quality, so when this metric becomes narrow we assume that the market has more liquidity and quality (Chordia et al. [2008]).

#### 7.3.2.2 The OLS Model

To gauge the effect of the hidden order to market liquidity and market quality, I adapt the OLS from Hasbrouck and Saar [2013] with our hidden order study. Also, each data set is standardized to have zero mean and unit variance to facilitate aggregation, then I present as in Hasbrouck's work. In my model, we assume that all measures should have a negative relationship with market quality, for instance, when the market quality improves, the $EffSpread$ should be decreased, and the market quality should have a positive association with the hidden order proxy or hidden order intensity ($HDI$). The $HDI$ is the fraction between hidden order volume and trade volume at time $t$. Therefore, the first model is:

$$MktQuality_{i,t} = a1MarketConcentration_{i,t} + a2HiddenOrderIntensity_{i,t} + \nu_{i,t} \tag{7.3.8}$$

where $i = 1, ..., N$ indexes the number of week to maturity, $t = 1, ...T$ indexes of time, $MktQuality$ represents one of the market quality measures ($Spread$, $EffSpread$, $RSpread$, and $PImpt$), and

$$MarketConcentration_{i,t} = \frac{TotalQuoteVolumes_{i,t}}{NumberOfTraders_{i,t}}. \tag{7.3.9}$$

This concentration ratio indicates whether the E-mini futures market is comprised of a few large or many small traders. The $MarketConcentration$ is assumed to have a negative relationship with the market quality.

To allow for the possibility of hidden order from bid and offer having a different impact on market quality, we separate the hidden order intensity into two components from buy and sell. Therefore, we have defined the second model as:

$$\begin{aligned} MktQuality_{i,t} =& a_1 MarketConcentration_{i,t} + a_2 AskHDC_{i,t} \\ &+ a_3 BidHDC_{i,t} + \nu_{i,t}. \end{aligned} \tag{7.3.10}$$

where $AskHDC_t$ is the function of hidden liquidity from seller and trading volume at time $t$ and $BidHDC_t$ is the function of hidden liquidity from buyer and trading volume at time $t$.

#### 7.3.2.3 Vector Auto-Regression Model

The Vector Auto-Regression or VAR model is one of the most effective for multiple time series analysis. It is an extension analysis on multiple time series regression to the univariate autoregressive model. The VAR system indicates the association amongst a set of interested variables, so it is used to analyze a certain aspect of the relationship between observed variables. Leung et al. [2000] and Jiang [2015] state that the VAR has proven to be a successful technique for forecasting systems and is appropriate for characterizing the dynamic behavior of financial time series. However, Sims [1980] ,who won a Nobel Prize in 2011, argues that the restricted VAR conventional model has no substantive justification, and economists should use the unrestricted VAR instead. Consequently, this work uses unrestricted VAR as Sims recommends by including all interesting variables in each equation.

We start with the basic VAR. The VAR is an $n$ equation with $N$ variable model, and each variable is explained by its own lag with the values of the remaining $N$-1 variables. The basic $n$-lag vector autoregression model can be written as:

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_n Y_{t-n} + \nu_t \tag{7.3.11}$$

where $Y_t$ is denoted as the vector values time series variables, $t$ is time index, $c$ is constant matrix, $\Phi$ is the coefficient matrices , and $\nu$ is an error term. The reduced form of the VAR model is:

$$Y = \mathit{\Phi}\mathbb{L} + \nu \tag{7.3.12}$$

where $Y$ is the dependent variables, $\mathbb{L}$ is the lagged operator variables, and $\mathit{\Phi}$ is the coefficient matrices of lagged terms,

$$\mathit{\Phi} = [c, \Phi_1, \Phi_2, \ldots, \Phi_n] = \begin{bmatrix} c^1 & \phi_{1,1}^1 & \phi_{1,2}^1 & \cdots & \phi_{1,1}^{N-1} & \phi_{1,2}^{N-1} & \ldots & \phi_{1,n}^{N-1} \\ c^2 & \phi_{2,1}^1 & \phi_{2,2}^1 & \cdots & \phi_{2,1}^{N-1} & \phi_{2,2}^{N-1} & \ldots & \phi_{2,n}^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c^{N-1} & \phi_{1,n}^1 & \phi_{2,n}^2 & \cdots & \phi_{2,n}^{N-1} & \phi_{2,n}^{N-1} & \ldots & \phi_{n,n}^{N-1} \end{bmatrix} \tag{7.3.13}$$

Following the description of Hamilton [1994], the VAR($n$) can be rewritten as a VAR(1) with companion matrix representing the vector projection. The $mean - adjusted$ form of the VAR($n$) is then

$$Y_t - u = c + \Phi_1(Y_{t-1} - u) + \Phi_2(Y_{t-2} - u) + \cdots + \Phi_n(Y_{k-n} - u) + \nu_t, \tag{7.3.14}$$

therefore, the VAR($n$) can be equivalently rewritten as:

$$\hat{Y}_t = F\hat{Y}_{t-n} + V_t \tag{7.3.15}$$

where $\hat{Y}_t = [Y_t - u, Y_{t-1} - u, \ldots, Y_{t-n+1} - u]^{-1}$, $F$ is the companion matrix

$$F = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{n-1} & \Phi_n \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix} \tag{7.3.16}$$

where $I$ is an identity matrix and $V_n \equiv [u_n, 0, \ldots, 0]^{-1}$. If the eigenvalues of the companion matrix are within the boundary of the unit circle, $F^s \to 0$; $s \to \infty$, then a stable VAR($n$) process is stationary and ergodic with time invariant means, variances, and auto-covariances.

At this point, I adapted the basic VAR model to provide empirical evidence on the response or dynamic relationship of variables, such as quote revision ($\widetilde{QR}$), signed-trade direction ($\widetilde{TD}$), signed-hidden order ($\widetilde{HDV}$) and signed-trades volume ($\widetilde{TV}$), to various exogenous impulses. At this point, the basic structure of restricted VAR model was modified , similar to Dufour and Engle [2000], Zebedee [2001], Chung et al. [2005], Viljoen et al. [2014] who studied quotes revision by using Hasbrouck [1991]'s model to examine the dynamic relationship of trades and quotes through a vector autoregression (VAR)system with five lags. The Hasbrouck [1991]'s model can be defined as

$$\begin{aligned} r_t &= \textstyle\sum_{i=1}^{n} a_i r_{t-1} + \sum_{i=1}^{n} b_i x_{t-1} + \nu_t^r, \\ x_t &= \textstyle\sum_{i=1}^{n} c_i r_{t-1} + \sum_{i=1}^{n} d_i x_{t-1} + \nu_t^x, \end{aligned} \tag{7.3.17}$$

where $r_t$ is the difference in mid-quote return; $t-i$ is the time between $r_t$ and $r_{t-i}, i = 1, ..., n$ and the error term is $\nu_t$.

This section start with a unit root test for all interested variables including $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$ and $\widetilde{TV}$ examining whether time series are stationary. Taking the existence of a unit root as the null hypothesis, the Augmented Dickey-Fuller (ADF) test is used for examination. The result from the ADF test for all series presents a strong rejection of the null hypothesis on non-stationary with a significant p-value at 1% level. The rejection indicates that the data samples are trend stationary. After testing a unit root test, i further modify the VAR to unrestricted VAR to differentiate the associated impact of hidden order and interested data series for E-mini S&P500. The modification system allows one to distinguish the impact of hidden order from other variables. The standard proposition of the modification system in the matrix form is

$$\begin{pmatrix} \widetilde{QR}_t \\ \widetilde{TD}_t \\ \widetilde{HDV}_t \\ \widetilde{TV}_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} + \sum_{i=1}^{n} \begin{pmatrix} \phi_{11,i} & \phi_{12,i} & \phi_{13,i} & \phi_{14,i} \\ \phi_{21,i} & \phi_{22,i} & \phi_{23,i} & \phi_{24,i} \\ \phi_{31,i} & \phi_{32,i} & \phi_{33,i} & \phi_{34,i} \\ \phi_{41,i} & \phi_{42,i} & \phi_{43,i} & \phi_{44,i} \end{pmatrix} \cdot \begin{pmatrix} \widetilde{QR}_{t-i} \\ \widetilde{TD}_{t-i} \\ \widetilde{HDV}_{t-i} \\ \widetilde{TV}_{t-i} \end{pmatrix} + \begin{pmatrix} \nu 1, t \\ \nu 2, t \\ \nu 3, t \\ \nu 4, t \end{pmatrix} \tag{7.3.18}$$

To estimate the VAR system, I estimate consistently by least squares using five lags and the covariance stationary is also assumed. The first 5 lags ($n = 5$) is choose as used in Hasbrouck [1991]; Dufour and Engle [2000]; Chung et al. [2005]. Therefore, the multi-equations unrestricted VAR model is reduced to a simple OLS form which allows easier understanding. This VAR system can be defined as: the quote revision equation

$$\widetilde{QR}_t = \sum_{i=1}^{5} \alpha_{1,i} \widetilde{QR}_{t-i} + \sum_{i=1}^{5} \alpha_{2,i} \widetilde{TD}_{t-i} + \sum_{i=1}^{5} \alpha_{3,i} \widetilde{HDV}_{t-i} + \sum_{i=1}^{5} \alpha_{4,i} \widetilde{TV}_{t-i} + \nu_t^{qr}. \tag{7.3.19}$$

The signed-trade direction equation

$$\widetilde{TD}_t = \sum_{i=1}^{5} \alpha_{1,i} \widetilde{QR}_{t-i} + \sum_{i=1}^{5} \alpha_{2,i} \widetilde{TD}_{t-i} + \sum_{i=1}^{5} \alpha_{3,i} \widetilde{HDV}_{t-i} + \sum_{i=1}^{5} \alpha_{4,i} \widetilde{TV}_{t-i} + \nu_t^{td}. \tag{7.3.20}$$

The signed-hidden order volume equation

$$\widetilde{HDV}_t = \sum_{i=1}^{5} \alpha_{1,i} \widetilde{QR}_{t-i} + \sum_{i=1}^{5} \alpha_{2,i} \widetilde{TD}_{t-i} + \sum_{i=1}^{5} \alpha_{3,i} \widetilde{HDV}_{t-i} + \sum_{i=1}^{5} \alpha_{4,i} \widetilde{TV}_{t-i} + \nu_t^{hdv} \tag{7.3.21}$$

The signed-trade volume equation

$$\widetilde{TV}_t = \sum_{i=1}^{5} \alpha_{1,i}\widetilde{QR}_{t-i} + \sum_{i=1}^{5} \alpha_{2,i}\widetilde{TD}_{t-i} + \sum_{i=1}^{5} \alpha_{3,i}\widetilde{HDV}_{t-i} + \sum_{i=1}^{5} \alpha_{4,i}\widetilde{TV}_{t-i} + \nu_t^{tv}. \tag{7.3.22}$$

Where the $\widetilde{QR}_t$ is quote revision at time $t$, $\widetilde{QR}_t = 100 \times (\ln mid_t - \ln mid_{t-1})$, $\widetilde{TD}_t$ is the signed-trade direction indicator at time $t$, $\widetilde{HDV}$ is signed-hidden order volume and $\widetilde{TV}$ is signed-trading volume at time $t$, $\nu_t$ is white noise, which may be correlated with a variance-covariance matrix $\Sigma$.

#### 7.3.2.4 Impulse Response Analysis

In the previous section, the VAR models is introduced in which represent the correlations among a set of variables. These are often used to analyze certain aspects of the relationships between the observed variables. Also, I am interested to know the response of one variable to an impulse in another variable in a system that involves a number of further variables. The impulse response analysis is an important type of structure analysis based on the VAR model. This model provides key insights on the reaction of any dynamic system in response to external change or it can provide the magnitude and direction of market impact.

The VAR system is transformed to an impulse response analysis in order to examine how rapidly the $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$, and $\widetilde{TV}$ movements are transmitted to the other observed variables. This impulse response is used to explain the speed of adjustment as an indicator of the degree of the association or the feedback effect of observed variables in this VAR system. The innovation of the impulse response can obtain information over the concern of the interactions among the variables.

Lütkepohl and Reimers [1992] suggest that there is no difference between VAR and VECM models in estimating the impulse response. Now, we start with a basic of VAR, $Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_n Y_{t-n} + \nu_t$ (Eq.7.3.11), where $\Phi_i$ are $4 \times 4$ coefficient metrics in our case. The estimation result from multivariate least squares has an asymptotic normal distribution where a covariance matrix can be estimated by the normal formula for stationary processes (Phylaktis [1999]).

Under the Wold Theorem, if the VAR($n$) is stationary then it has an infinite order moving average. Hence, we transform Eq.7.3.11 to a vector moving average(VMA) to examine the interaction between our four variables from the E-mini market. The transformation model is defined as:

$$\begin{aligned} Y_t &= u + \sum_{i=1}^{\infty} \Phi_i \nu_{t-i} \\ &= \hat{\mathit{\Phi}}(\mathbb{L}) \nu_t \end{aligned} \tag{7.3.23}$$

where $\mathbb{L}$ is the lag operator and $\hat{\mathit{\Phi}}(\mathbb{L})$ is matrix polynomial in $\mathbb{L}$. Assuming the time is $t+s$, so the VMA system can be described as

$$\begin{aligned} Y_{t+s} &= u + \nu_{t+s} + \hat{\Phi}_1 \nu_{t+s-1} + \hat{\Phi}_1 \nu_{t+s-2} + \cdots + \hat{\Phi}_s \nu_t + \hat{\Phi}_{s+1} \nu_{t-1} + \dots \\ &= \hat{\mathit{\Phi}}(\mathbb{L}) \nu_{t+s}, \end{aligned} \tag{7.3.24}$$

thus, the VMA system can be rewritten as;

$$Y_t - u = \hat{\Phi}(\mathbb{L})_t \tag{7.3.25}$$

where $\hat{\mathit{\Phi}} \in \{\hat{\Phi}_1, \{\hat{\Phi}_2, \dots, \{\hat{\Phi}_s, \{\hat{\Phi}_{s+1}, \dots\}$, the operators $\hat{\mathit{\Phi}}(\mathbb{L})$ and $\mathit{\Phi}(\mathbb{L})$ are related with $\hat{\mathit{\Phi}}(\mathbb{L}) = [\mathit{\Phi}(\mathbb{L})]^1$. $\hat{\Phi} = F_{11}^{(s)}$ is the $s$ order moving average matrix, and the $F_{11}^{(s)}$ denotes the upper left block of $F^s$. Consequently, the $\hat{\Phi}_{ij}^{(s)}$ is the impulse response of $Y_{i,t+s}$ to an innovation of one unit change in standard deviation of the $Y_{jt}$ at the $s$ step.

In the unrestricted VAR system, however, of the $\widetilde{QR}$, the $\widetilde{TD}$, the $\widetilde{HDV}$, and the $\widetilde{TV}$, the main interesting is not only when all four responses decay in each VAR system, but there are two interesting issues from this multivariate system. First, the quote revision analysis in quote equation and signed-trade direction equation and secondly the impulse response following a shock in hidden order and trade volume equation. The issue of interest is separated into two sets because in the primary test I do not find a strong impact of the shock of $\widetilde{HDV}$ and $\widetilde{TV}$ to $\widetilde{QR}$ and $\widetilde{TV}$. Furthermore, this study focus on the speed of coverage following a shock, and then compare this speed of decline between high and low frequency trading periods following a term-structure study. In examining the speed of adjustment in this unrestricted VAR model, I take into account the interaction between the $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$, and $\widetilde{TV}$ in the quote, signed-trade direction, signed-hidden order volume and signed-trade volume equations.

## 7.4 Analysis

### 7.4.1 Price impact of hidden order

Recent financial market literature has identified two interpretations of the risk of order exposure. Firstly, traders' risk from the front running liquidity, and secondly, the markets' risk from a large order submission. However, this risk can be decreased by submitting an invisible order as studied in Harris [1996], Aitken et al. [2001], Anand and Weaver [2004], De Winne and D'hondt [2007a], and Frey and Sandås [2009]. To understand whether the hidden order can manage the exposure risk of front running for traders who want to keep their trading position secret or can protect the market from the risk of market stress when a large trading volume appears on a trading screen, I had to study the impact of a hidden liquidity on the market quality. This section focuses on the price impact of hidden order using the tick-by-tick price impact to investigate the price improvement granted by this type of order. If the nearby trade after the hidden order occurs is not granted much, then we will see only one-period price impact. However, if the hidden order is coincident with a subsequent decrease or increase in price, we should see a long-term price impact. This work follows Reiss and Werner [2005] in presenting the price impact in terms of median cumulative difference before and after the hidden order occurs in the market. The result from this analysis also agrees with Reiss and Werner [2005], that the median cumulative differences price impact is less sensitive to outliers, and it is not different from the mean cumulative differences price impact.

Figure 7.2 presents the median cumulative differences price impact surrounding twenty trade transactions before and after hidden orders submitted from seller and buyer. For comparison, I also include the median cumulative differences price impact from ordinary trades, from which I expect little or no short term or long term price impact. The figures reveal that the median cumulative differences price impact of hidden order is significantly larger than the median cumulative price impact of ordinary trades ($*$-line). Moreover, the sell hidden order has a higher positive price impact than the buy hidden order for twenty trades before hidden orders occur in the market. However, after this event, the price impact shows an adverse effect. The figure also shows a similar pattern of the average median cumulative price impact for both buy and sell hidden orders.

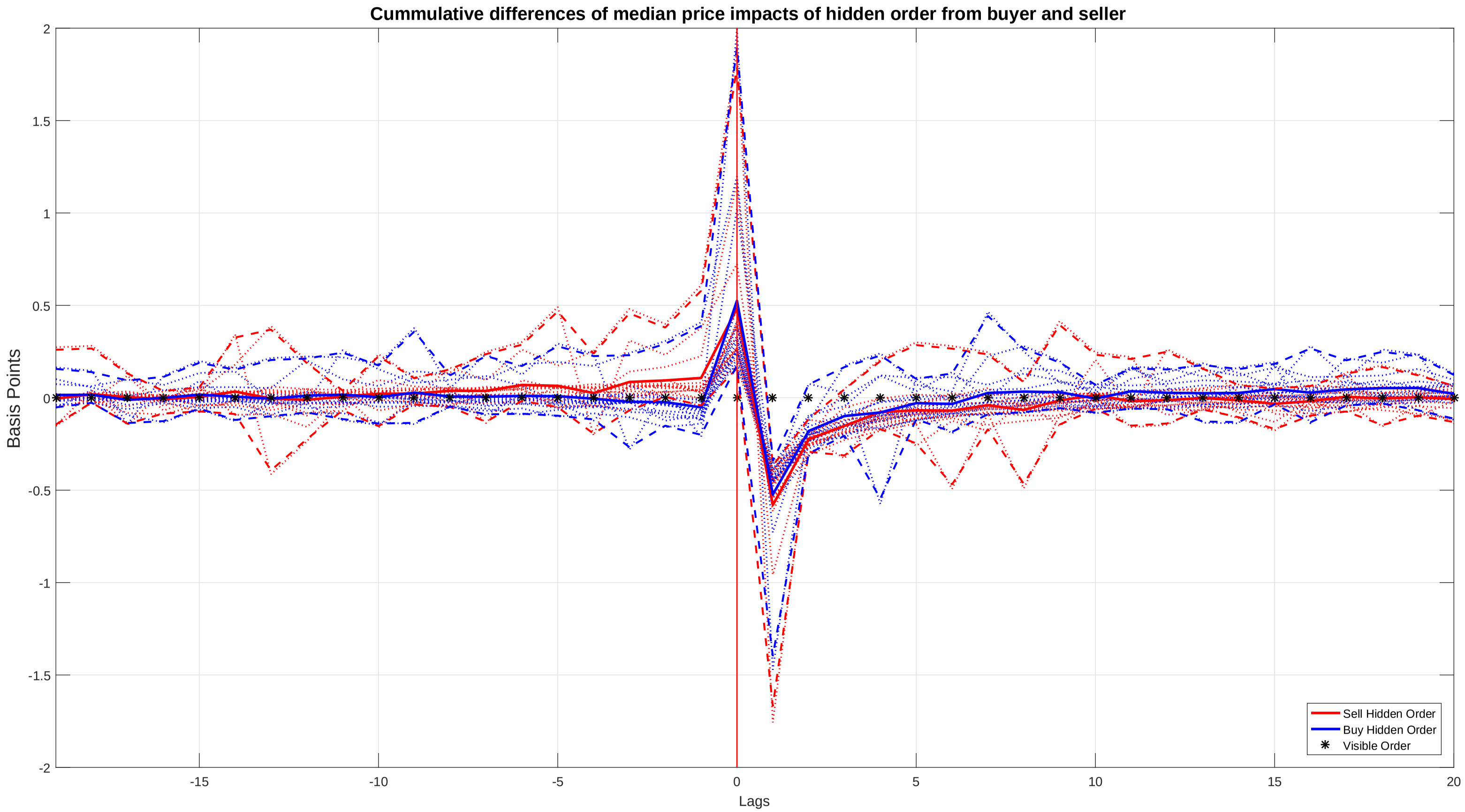



**Figure 7.2: Cumulative differences of median price impacts of hidden order from buyer and seller.**
Note: This figure presents the median percentage change in cumulative price impact beginning from 20 steps prior to selling hidden order(red thick line), buy hidden order (blue thick line) and ordinary order(* line) occur in the E-mini S&P 500 market. The figure also shows the standard errors (dashed - - line). This figure, additionally, presents overlay plots of the cumulative differences of median price impacts of hidden order from eighteen to one week's data samples from 30 E-mini S&P500 (dotted $\cdots$ line).

The median cumulative price impact measured from twenty to nine trades before the event is near to zero for buy and sell hidden orders, however from eight trades before the event, the price impact from the sell hidden order is increasing and reaches a peak when the hidden order is submitted at 0.5 basis points. Interestingly, there is a similar peak level of the average median cumulative differences price impact from both buy and sell hidden orders. After, a hidden order is submitted, the average median cumulative differences price impact from both sides plunges to the lowest level at -0.6 basis points, then bounces back from one trade after the event. Finally, the price impact remains stable from eight trades after the event at around zero basis point. This result supports the view that the hidden order, whether from a buy or sell hidden order, is more informed on the prices than normal trade.

The main interesting pattern from this figure is that the price impact of the hidden order from both buy and sell orders begins to move before the event. This movement occurs because buyers and sellers who submit hidden orders are informed about trades or these types of traders may have proceeded to the direct trade. The largest cumulative differences price impact occurs on the next trade after the hidden order submitted, which is a decrease from 0.5 to -.6 basis points for both the buy and sell hidden order. The figure also shows that there is a reversal in the cumulative differences price impact (or cumulative abnormal return (Reiss and Werner [2005])) coincident with the hidden order. This pattern explains the spike of the cumulative differences price impact on the event trade, then drops to the bottom on the next trade. Finally, the cumulative price impact increases to just around 0 basis points for a buyer (seller) from eight trades after the event. Also, the price impact from buy hidden orders is slightly higher than a sell hidden order.

Additionally, the spike of the cumulative median price impact on the occurrences of hidden order, whether sell or buy, can be explained by the immediate market reaction to a greater price impact when the order has been executed followed by a price reversal. Also, the price effect of the hidden order is short-lived and reaches a peak at the beginning of the event, then after the hidden order appears a price reversal is expected. Finally, informed or sophisticated traders can gain abnormal profits if they can detect any hidden order fast enough to respond to a shock from this type of trade; traders can benefit from the overreaction before the order is submitted with a price reversal afterwards.

### 7.4.2 Hidden order and market liquidity

In addition to examining the hidden order impact on the E-mini S&P500 market, this analysis follows OLS regression to investigate the relationship between hidden order, market liquidity, and market quality. Therefore, the model (1) is $MktQuality_{i,t} = a1MarketConcentration_{i,t} + a2HiddenOrderIntensity_{i,t} + \varepsilon_{i,t}$; the coefficient estimates are OLS, and p-value and $R^2$ are also presented. This work separates the time period into one-week blocks starting eighteen weeks before maturity, however, the table reports on the main analysis present only four periods, eighteen, twelve, six, and one week to maturity. These four include low and high-frequency trading periods. The low-frequency periods are week eighteen and week one and the average trade updated for these is 0.005 million updates for the former, and 0.158 million updates for week one. The high-frequency trading periods are week twelve and week six, and the average trade updated for these are 2.840 million updated for the former and 1.913 million for the latter.

Table 7.3 presents estimated coefficients of model (1) for week eighteen, twelve, six and one. The main interesting coefficient estimate is $a_2$, which measures the association of hidden order and the market quality. Overall, the result shows that a higher portion of hidden order is associated with a wide-spread. However, the effective spread, realized spread, and price impact is narrow. In addition, the relationship between hidden order activity and these market quality proxies are similar for all periods.

The coefficient estimates from OLS can be interpreted as a causal impact between the hidden liquidity and market quality. The result shows that while $HiddenOrderIntensity$ ($HOI$) is positively related to quoted spread ($Spread$), it is negatively related to $EffSpread$, $ReSpread$, and $Pimpt$ for all periods. The highest impact of the hidden order wide $Spread$ appears on week six at 0.500 and the lowest is in week eighteen at 0.026. In contrast, the highest negative impact of $HOI$ to $EffSpread$ appears on week twelve at -0.405, whilst the lowest is in week eighteen at -0.053. The highest negative impact of $HOI$ to $ReSpread$ is in week six at -0.380, whilst the lowest is in week eighteen at -0.039. Finally, the highest negative impact of the $HOI$ to $Pimpt$ is in week twelve at -0.579, whilst the lowest is in week eighteen at -0.062. This analysis also finds that the effect of $HOI$ on $Spread$ for week one is not statistically significant with a $p - value$ of 0.213. These results suggest that the hidden order may increase market quality as the $EffSpread$, $ReSpread$, and $Pimpt$ are narrower when the $HOI$ is increased. Nevertheless, the $Spread$ widens when the $HOI$ increases. I will discuss these relationships in the following section. Furthermore, during high-frequency trading periods the result from this OLS shows that hidden order improves the market quality with a greater impact than low-frequency trading periods.

**Table 7.3: Hidden order and market quality: OLS estimates.**

Model 1:
$MktQuality_{i,t} = a_1 MarketConcentration_{i,t} + a_2 HiddenOrderIntensity_{i,t} + \nu_{i,t}$

| | | Week_01 | | Week_06 | | Week_12 | | Week_18 | |
|---|---|---|---|---|---|---|---|---|---|
| | | $a_1$ | $a_2$ | $a_1$ | $a_2$ | $a_1$ | $a_2$ | $a_1$ | $a_2$ |
| *Spread* | $Coff.$ | 0.095 | 0.091 | 0.054 | 0.500 | 0.168 | 0.474 | 0.543 | 0.026 |
| | $(p-value)$ | (0.213) | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.022) | | (0.208) | | (0.210) | | (0.234) | |
| *EffSpread* | $Coff.$ | 0.674 | -0.260 | 0.030 | -0.393 | 0.088 | -0.405 | 0.887 | -0.053 |
| | $(p-value)$ | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.411) | | (0.166) | | (0.184) | | (0.804) | |
| *ReSread* | $Coff.$ | 0.663 | -0.214 | 0.006 | -0.380 | 0.071 | -0.374 | 0.811 | -0.039 |
| | $(p-value)$ | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.393) | | (0.146) | | (0.154) | | (0.625) | |
| *Pimpt* | $Coff.$ | 0.439 | -0.385 | 0.190 | -0.361 | 0.238 | -0.579 | 0.784 | -0.062 |
| | $(p-value)$ | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.239) | | (0.237) | | (0.465) | | (0.702) | |

Note: This table presenting pooled panel regression analyzes the relationship between hidden liquidity and market quality. The hidden liquidity measure is $HiddenOrderIntensity_{i,t}$, this is a fraction of hidden order volume and trades volume. $MktQuality_{i,t}$ is a placeholder denoting: *Spread* is quoted spread; $EffSpread$ is the effective spread; $ReSpread$ is the realized spread; $Pimpt$ is the price impact. $MarketConcentration$ is the function of a number of traders and their respective volume of the total quote volume in LOB. I standardized each variable by dividing by their standard deviation and the coefficient estimates are OLS.

**Table 7.4: Hidden buy and sell order and the market quality: OLS estimates**

Model 2: $MktQuality_{i,t} = a_1 MarketConcentration_{i,t} + a_2 AskHDCi,t + a_3 BidHDCi,t + \nu_{i,t}$

| | | Week_01 | | | Week_06 | | | Week_12 | | | Week_18 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $a_1$ | $a_2$ | $a_3$ | $a_1$ | $a_2$ | $a_3$ | $a_1$ | $a_2$ | $a_3$ | $a_1$ | $a_2$ | $a_3$ |
| *Spread* | *Coff.* | 0.092 | 0.081 | 0.155 | 0.061 | 0.330 | 0.586 | 0.263 | 0.674 | 0.778 | 0.556 | 0.067 | 0.068 |
| | $(p-value)$ | (0.378) | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (0.053) | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.022) | | | (0.208) | | | (0.400) | | | (0.236) | | |
| *EffSpread* | *Coff.* | 0.658 | -0.047 | -0.311 | 0.060 | -0.470 | -0.615 | 0.072 | -0.317 | -0.511 | 0.896 | -0.224 | -0.212 |
| | $(p-value)$ | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.420) | | | (0.188) | | | (0.186) | | | (0.811) | | |
| *Respread* | *Coff.* | 0.636 | -0.113 | -0.155 | 0.078 | -0.445 | -0.592 | 0.057 | -0.289 | -0.472 | 0.820 | -0.253 | -0.216 |
| | $(p-value)$ | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.419) | | | (0.102) | | | (0.156) | | | (0.632) | | |
| *Pimpt* | *Coff.* | 0.480 | -0.861 | -0.976 | 0.140 | -0.461 | -0.452 | 0.196 | -0.521 | -0.749 | 0.786 | -0.120 | -0.152 |
| | $(p-value)$ | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** | (<0.001)** |
| | $R^2$ | (0.332) | | | (0.280) | | | (0.484) | | | (0.705) | | |

Note: This table presenting pooled panel regression analyzes the relationship between hidden liquidity and market quality. The hidden liquidity measure is $AskHDC_{i,t}$ is a fraction of sell hidden order volume and trades volume;$BidHDC_{i,t}$ is a fraction of buy hidden order volume and trades volume. $MktQuality_{i,t}$ is a placeholder denoting: *Spread* is quoted spread; *EffSpread* is the effective spread; *ReSpread* is the realized spread; *Pimpt* is the price impact. *MarketConcentration* is the function of a number of traders and their respective volume of the total quote volume in LOB. I standardized each variable by dividing by their standard deviation, and the coefficient estimates are OLS.

To allow us to investigate the possibility of a different relationship between buy or sell hidden liquidity to market quality, I separate the $HOI$ into a buy and sell hidden order. The variable $AskHDC$ is a function of hidden sell orders and market volume, whilst $BidHDC$ is a function of the hidden buy order and market volume. I call this model 2: $MktQuality_{i,t} = a_1 MarketConcentration_{i,t} + a_2 AskHDCi, t + a_3 BidHDCi, t + \varepsilon_{i,t}$.

Table 7.4 presents the estimated coefficients of the model (2) side-by-side for week eighteen, twelve, six and one. Overall, the result shows that both $AskHDC$ and $BidHDC$ implies higher $Spread$ and lower $EffSpread$, $ReSpread$, and $Pimpt$. The result shows that the coefficient estimates from the model (2) have a similar positive effect from buying and selling hidden liquidity to market quality. This result is similar to model(1), however, there is a different degree of impact for hidden buy and sell orders as the results of the coefficient estimates are shown in table 7.3 and 7.4. Additionally, the coefficients from the model (2) suggest that the hidden buy order has a higher impact on market quality than the sell hidden order for all periods. Furthermore, the highest impact of $AskHDC$, $BidHDC$ to market quality appears on week six and twelve, when trading activity is high, and the lowest impact is in week eighteen when trading activity is low.

At this point, the contemporaneous positive association between $AskHDC$, $BidHDC$ and $Spread$ may be explained by the nature of traders who tend to hide their order when the quoted spread is wide. These results are consistent with Bessembinder et al. [2009] who claims that traders are more likely to hide their order when the arrival rate of order is low and the adverse selection is high. However, the positive relationship of hidden order and $Spread$ is statistically significant at the 0.001 level only in weeks twelve and six. Moreover, the contemporaneous negative association between hidden order and $EffSpread$ shows that the $EffSpread$ tend to be narrow when the activity of hidden order is high. This is consistent with Comerton-Forde and Tang [2009] who shows that the hidden order reduces the cost of trade. Also, the contemporaneous negative association between hidden order and $ReSpread$ indicates that the transitory price movement reduces, or the market tends to follow, the previous trades. Finally, there is a significant contemporaneous negative association between hidden order and $Pimpt$ indicating that hidden order reduces the level of information asymmetry in the market.

### 7.4.3 Vector autoregressive and serial correlation on E-mini S&P500

In this section, the E-mini S&P500 is analyzed by using the vector autoregressive (VAR) model. As in the previous section, this study estimates the VAR model for one week blocks starting from 18 weeks before maturity as a term-structure analysis. I serve the result of VAR analysis by presenting not only the coefficient estimates but also t-statistic and p-value on each block, to determine whether the value of estimated coefficients is significantly different from zero.

This study uses the first five lags because the coefficients for longer lags are small; five lags have also been used in other studies such as Hasbrouck [1991]; Dufour and Engle [2000]; Chung et al. [2005]. These previous studies focused on the $b_i$ coefficients from quote revision or $\widetilde{QR}_t$ equation which measure the price impact of trades and the $b_i$ coefficients in a trade direction or $\widetilde{TD}_t$ equation which measures a serial correlation in signed trades model. However, this analysis further modified this restricted VAR to an unrestricted VAR model. The favorite form of VAR for trade and quote revision in a dynamic model is developed by adding another two variables, such as the signed-hidden order volume and signed-trade volume. The following section serves the empirical results from four different periods from week eighteen to one. I select these four periods as a proxy of high and low trading activity sample periods. Firstly, the low-frequency period is week eighteen and one. Secondly, the high-frequency data sample period is week twelve and two. Furthermore, the empirical results from these four blocks is achieved by comparing them together instead of reporting one-by-one (See section A.1 for the result of VAR for 18-week blocks).

Panel A of table 7.5, 7.6, 7.7 and 7.8 present the estimated coefficients from the quote revision equation from VAR analysis. First, this panel shows the positive coefficient value of $b_1\widetilde{TD}$ only on lag 1, where the p-value is highly significant at 1 % level. The coefficient estimated for this parameter is 0.1360 for week eighteen, 0.1328 for week twelve, 0.1229 for week six, and 0.2530 for week one. Furthermore, there are negative coefficients from $\widetilde{TD}$ from lag 2 to lag 5. However, the coefficient for $b_2 - b_5$) are considerably smaller than $b_1$, which indicates that the lagged 2 to 5 trades are the primary cause of price movement. These results indicate that the market maker raises (lower) the quote midpoint immediately subsequent to a buy(sell) order.

While the $c_1$ for week one indicates that the $\widetilde{HDV}$ has negatively autocorrelated, the $c_1$ ($\widetilde{HDV}$) for week six, twelve and eighteen show a positive coefficient for the first lag with less than 1% significant level. The estimated coefficient is 0.0058, 0.0090 and 0.0014 for weeks six, twelve and eighteen, respectively. The positive coefficient for week six, twelve and eighteen indicates that the prior change of $\widetilde{HDV}$ leads to the change of the mid-quote. Therefore, the results of the positive impact of $\widetilde{HDV}$ indicates that after the buy-initiated hidden or-

**Table 7.5: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for one week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0283 | 37.87 | (<0.01) | b1 | 0.2530 | 267.74 | (<0.01) | c1 | -0.0068 | -7.82 | (<0.01) | d1 | 0.0156 | 20.60 | (<0.01) |
| a2 | 0.0020 | 2.70 | (<0.01) | b2 | -0.0542 | -52.31 | (<0.01) | c2 | -0.0018 | -2.05 | (<0.01) | d2 | 0.0049 | 5.60 | (<0.01) |
| a3 | -0.0013 | -1.70 | (0.08) | b3 | -0.0108 | -10.33 | (<0.01) | c3 | -0.0023 | -2.71 | (<0.01) | d3 | 0.0004 | 0.49 | (0.62) |
| a4 | 0.0003 | 0.43 | (0.66) | b4 | -0.0012 | -1.15 | (0.25) | c4 | -0.0003 | -0.36 | (0.71) | d4 | 0.0000 | 0.02 | (0.98) |
| a5 | -0.0019 | -2.67 | (<0.01) | b5 | -0.0051 | -5.29 | (<0.01) | c5 | -0.0002 | -0.25 | (0.80) | d5 | -0.0015 | -1.72 | (0.08) |
| C | | 0.0016 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0502 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0612 | -96.14 | (<0.01) | b1 | 0.3550 | 440.54 | (<0.01) | c1 | 0.0045 | 6.08 | (<0.01) | d1 | 0.0228 | 35.34 | (<0.01) |
| a2 | -0.0049 | -7.73 | (<0.01) | b2 | 0.1810 | 204.73 | (<0.01) | c2 | 0.0055 | 7.50 | (<0.01) | d2 | 0.0027 | 3.62 | (<0.01) |
| a3 | 0.0051 | 7.96 | (<0.01) | b3 | 0.0602 | 67.46 | (<0.01) | c3 | 0.0038 | 5.18 | (<0.01) | d3 | 0.0014 | 1.94 | (<0.01) |
| a4 | 0.0060 | 9.44 | (<0.01) | b4 | 0.0335 | 37.81 | (<0.01) | c4 | 0.0024 | 3.21 | (<0.01) | d4 | 0.0022 | 3.03 | (<0.01) |
| a5 | 0.0124 | 20.80 | (<0.01) | b5 | 0.0337 | 41.14 | (<0.01) | c5 | 0.0023 | 3.55 | (<0.01) | d5 | 0.0016 | 2.22 | (<0.01) |
| C | | 0.0039 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.3093 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0227 | -34.79 | ( <0.01) | b1 | -0.0310 | -37.55 | (<0.01) | c1 | 0.0144 | 19.07 | (<0.01) | d1 | 0.5015 | 759.21 | (<0.01) |
| a2 | -0.0004 | -0.62 | (<0.01) | b2 | 0.0484 | 53.42 | (<0.01) | c2 | 0.0163 | 21.55 | (<0.01) | d2 | 0.0168 | 22.10 | (<0.01) |
| a3 | 0.0016 | 2.39 | (<0.01) | b3 | 0.0136 | 14.83 | (<0.01) | c3 | 0.0127 | 16.74 | (<0.01) | d3 | 0.0030 | 3.96 | (<0.01) |
| a4 | 0.0031 | 4.70 | (<0.01) | b4 | 0.0110 | 12.07 | (<0.01) | c4 | 0.0060 | 7.94 | (<0.01) | d4 | 0.0011 | 1.50 | (0.13) |
| a5 | 0.0039 | 6.33 | (<0.01) | b5 | 0.0095 | 11.29 | (<0.01) | c5 | 0.0069 | 10.41 | (<0.01) | d5 | 0.0026 | 3.47 | (<0.01) |
| C | | 0.0011 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.2747 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0229 | -31.00 | (<0.01) | b1 | 0.0700 | 74.74 | (<0.01) | c1 | 0.0268 | 31.26 | (<0.01) | d1 | 0.1291 | 172.40 | (<0.01) |
| a2 | -0.0031 | -4.21 | (<0.01) | b2 | 0.0564 | 54.88 | (<0.01) | c2 | 0.0215 | 25.01 | (<0.01) | d2 | 0.0261 | 30.34 | (<0.01) |
| a3 | -0.0002 | -0.24 | (0.80) | b3 | 0.0181 | 17.40 | (<0.01) | c3 | 0.0107 | 12.46 | (<0.01) | d3 | 0.0162 | 18.77 | (<0.01) |
| a4 | 0.0020 | 2.75 | (<0.01) | b4 | 0.0112 | 10.86 | (<0.01) | c4 | 0.0079 | 9.18 | (<0.01) | d4 | 0.0130 | 15.13 | (<0.01) |
| a5 | 0.0046 | 6.62 | (<0.01) | b5 | 0.0093 | 9.78 | (<0.01) | c5 | 0.0143 | 19.08 | (<0.01) | d5 | 0.0138 | 16.01 | (<0.01) |
| C | | 0.0052 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0677 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QR}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

## Table 7.6: Coefficient estimates of VAR model on E-Mini S&P500 for six week to maturity

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0302 | 133.00 | (<0.01) | b1 | 0.1229 | 305.22 | (<0.01) | c1 | 0.0058 | 19.54 | (<0.01) | d1 | 0.0191 | 86.41 | (<0.01) |
| a2 | 0.0064 | 27.98 | (<0.01) | b2 | -0.0015 | -3.48 | (<0.01) | c2 | 0.0049 | 16.63 | (<0.01) | d2 | 0.0098 | 32.91 | (<0.01) |
| a3 | 0.0004 | 1.75 | (<0.01) | b3 | -0.0043 | -9.62 | (<0.01) | c3 | 0.0029 | 9.81 | (<0.01) | d3 | 0.0049 | 16.58 | (<0.01) |
| a4 | -0.0013 | -5.50 | (<0.01) | b4 | -0.0044 | -10.01 | (<0.01) | c4 | 0.0027 | 9.00 | (<0.01) | d4 | 0.0045 | 15.07 | (<0.01) |
| a5 | -0.0030 | -14.00 | (<0.01) | b5 | -0.0048 | -11.91 | (<0.01) | c5 | 0.0044 | 20.11 | (<0.01) | d5 | 0.0032 | 10.65 | (<0.01) |
| C | | 0.0001 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0170 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1070 | -809.41 | (<0.01) | b1 | 0.3833 | 1633.41 | (<0.01) | c1 | -0.0020 | -11.57 | (<0.01) | d1 | 0.0135 | 104.66 | (<0.01) |
| a2 | -0.0424 | -316.18 | (<0.01) | b2 | 0.2555 | 1006.33 | (<0.01) | c2 | -0.0015 | -8.64 | (<0.01) | d2 | 0.0057 | 32.68 | (<0.01) |
| a3 | -0.0175 | -130.28 | (<0.01) | b3 | 0.1142 | 440.93 | (<0.01) | c3 | -0.0012 | -7.16 | (<0.01) | d3 | 0.0042 | 24.35 | (<0.01) |
| a4 | -0.0032 | -24.32 | (<0.01) | b4 | 0.0723 | 285.15 | (<0.01) | c4 | -0.0010 | -6.07 | (<0.01) | d4 | 0.0030 | 17.35 | (<0.01) |
| a5 | 0.0095 | 76.91 | (<0.01) | b5 | 0.0609 | 260.27 | (<0.01) | c5 | 0.0002 | 1.85 | (0.06) | d5 | 0.0014 | 8.26 | (<0.01) |
| C | | -0.0008 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.6652 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0294 | -175.64 | (<0.01) | b1 | -0.0669 | -224.99 | (<0.01) | c1 | 0.0010 | 4.37 | (<0.01) | d1 | 0.6748 | 4145.18 | (<0.01) |
| a2 | -0.0087 | -51.18 | (<0.01) | b2 | 0.0594 | 184.60 | (<0.01) | c2 | 0.0060 | 27.36 | (<0.01) | d2 | 0.0105 | 47.95 | (<0.01) |
| a3 | -0.0027 | -16.14 | (<0.01) | b3 | 0.0238 | 72.55 | (<0.01) | c3 | 0.0034 | 15.39 | (<0.01) | d3 | 0.0017 | 7.58 | (<0.01) |
| a4 | 0.0003 | 1.76 | (0.08) | b4 | 0.0140 | 43.46 | (<0.01) | c4 | 0.0035 | 16.22 | (<0.01) | d4 | 0.0018 | 8.01 | (<0.01) |
| a5 | 0.0031 | 20.07 | (<0.01) | b5 | 0.0122 | 41.26 | (<0.01) | c5 | 0.0040 | 24.36 | (<0.01) | d5 | 0.0007 | 3.23 | (<0.01) |
| C | | 0.0000 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.4673 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0350 | -156.43 | (<0.01) | b1 | 0.0546 | 137.53 | (<0.01) | c1 | 0.0145 | 49.63 | (<0.01) | d1 | 0.0724 | 332.64 | (<0.01) |
| a2 | -0.0141 | -62.04 | (<0.01) | b2 | 0.0691 | 160.72 | (<0.01) | c2 | 0.0103 | 35.40 | (<0.01) | d2 | 0.0199 | 67.75 | (<0.01) |
| a3 | -0.0065 | -28.71 | (<0.01) | b3 | 0.0257 | 58.67 | (<0.01) | c3 | 0.0069 | 23.71 | (<0.01) | d3 | 0.0131 | 44.72 | (<0.01) |
| a4 | -0.0021 | -9.55 | (<0.01) | b4 | 0.0154 | 35.97 | (<0.01) | c4 | 0.0052 | 17.80 | (<0.01) | d4 | 0.0119 | 40.69 | (<0.01) |
| a5 | 0.0016 | 7.53 | (<0.01) | b5 | 0.0117 | 29.42 | (<0.01) | c5 | 0.0096 | 44.23 | (<0.01) | d5 | 0.0092 | 31.58 | (<0.01) |
| C | | -0.0005 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0491 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QR}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 7.7: Coefficient estimates of VAR model on E-Mini S&P500 for twelve week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0331 | 118.34 | (<0.01 ) | b1 | 0.1328 | 276.52 | (<0.01 ) | c1 | 0.0090 | 25.20 | (<0.01 ) | d1 | 0.0223 | 81.93 | (<0.01 ) |
| a2 | 0.0068 | 23.80 | (<0.01 ) | b2 | -0.0062 | -12.03 | (<0.01 ) | c2 | 0.0083 | 23.00 | (<0.01 ) | d2 | 0.0120 | 33.37 | (<0.01 ) |
| a3 | 0.0018 | 6.26 | (<0.01 ) | b3 | -0.0064 | -12.12 | (<0.01 ) | c3 | 0.0056 | 15.63 | (<0.01 ) | d3 | 0.0061 | 16.89 | (<0.01 ) |
| a4 | -0.0018 | -6.40 | (<0.01 ) | b4 | -0.0068 | -13.02 | (<0.01 ) | c4 | 0.0043 | 11.99 | (<0.01 ) | d4 | 0.0056 | 15.51 | (<0.01 ) |
| a5 | -0.0031 | -11.88 | (<0.01 ) | b5 | -0.0075 | -15.65 | (<0.01 ) | c5 | 0.0056 | 20.57 | (<0.01 ) | d5 | 0.0041 | 11.41 | (<0.01 ) |
| | C | 0.0006 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0188 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1087 | -645.31 | (<0.01 ) | b1 | 0.3777 | 1307.35 | (<0.01 ) | c1 | -0.0033 | -15.25 | (<0.01 ) | d1 | 0.0136 | 83.01 | (<0.01 ) |
| a2 | -0.0430 | -251.47 | (<0.01 ) | b2 | 0.2567 | 821.87 | (<0.01 ) | c2 | -0.0030 | -13.98 | (<0.01 ) | d2 | 0.0055 | 25.41 | (<0.01 ) |
| a3 | -0.0176 | -103.24 | (<0.01 ) | b3 | 0.1133 | 355.49 | (<0.01 ) | c3 | -0.0015 | -7.06 | (<0.01 ) | d3 | 0.0051 | 23.34 | (<0.01 ) |
| a4 | -0.0033 | -19.79 | (<0.01 ) | b4 | 0.0717 | 229.70 | (<0.01 ) | c4 | -0.0018 | -8.17 | (<0.01 ) | d4 | 0.0030 | 13.78 | (<0.01 ) |
| a5 | 0.0093 | 59.06 | (<0.01 ) | b5 | 0.0597 | 207.19 | (<0.01 ) | c5 | 0.0000 | -0.13 | (<0.01 ) | d5 | 0.0017 | 7.95 | (<0.01 ) |
| | C | -0.0009 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6441 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0283 | -134.34 | (<0.01 ) | b1 | -0.0624 | -172.64 | (<0.01 ) | c1 | -0.0032 | -11.85 | (<0.01 ) | d1 | 0.6593 | 3220.56 | (<0.01 ) |
| a2 | -0.0087 | -40.87 | (<0.01 ) | b2 | 0.0566 | 144.69 | (<0.01 ) | c2 | 0.0057 | 21.05 | (<0.01 ) | d2 | 0.0132 | 48.84 | (<0.01 ) |
| a3 | -0.0027 | -12.55 | (<0.01 ) | b3 | 0.0230 | 57.61 | (<0.01 ) | c3 | 0.0025 | 9.25 | (<0.01 ) | d3 | 0.0037 | 13.59 | (<0.01 ) |
| a4 | 0.0008 | 3.69 | (<0.01 ) | b4 | 0.0139 | 35.61 | (<0.01 ) | c4 | 0.0024 | 8.98 | (<0.01 ) | d4 | 0.0028 | 10.17 | (<0.01 ) |
| a5 | 0.0031 | 15.47 | (<0.01 ) | b5 | 0.0108 | 29.94 | (<0.01 ) | c5 | 0.0039 | 19.23 | (<0.01 ) | d5 | 0.0019 | 6.86 | (<0.01 ) |
| | C | 0.0000 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4446 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0319 | -115.02 | (<0.01 ) | b1 | 0.0478 | 100.63 | (<0.01 ) | c1 | 0.0117 | 32.81 | (<0.01 ) | d1 | 0.0761 | 283.01 | (<0.01 ) |
| a2 | -0.0127 | -45.07 | (<0.01 ) | b2 | 0.0621 | 120.99 | (<0.01 ) | c2 | 0.0099 | 27.94 | (<0.01 ) | d2 | 0.0201 | 56.57 | (<0.01 ) |
| a3 | -0.0059 | -21.08 | (<0.01 ) | b3 | 0.0234 | 44.64 | (<0.01 ) | c3 | 0.0078 | 21.94 | (<0.01 ) | d3 | 0.0133 | 37.25 | (<0.01 ) |
| a4 | -0.0012 | -4.23 | (<0.01 ) | b4 | 0.0163 | 31.84 | (<0.01 ) | c4 | 0.0036 | 10.03 | (<0.01 ) | d4 | 0.0100 | 28.21 | (<0.01 ) |
| a5 | 0.0023 | 8.94 | (<0.01 ) | b5 | 0.0087 | 18.33 | (<0.01 ) | c5 | 0.0094 | 34.87 | (<0.01 ) | d5 | 0.0107 | 30.16 | (<0.01 ) |
| | C | 0.0002 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0413 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QR}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 7.8: Coefficient estimates of VAR model on E-Mini S&P500 for eighteen week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0093 | -1.37 | (0.17) | b1 | 0.1360 | 17.82 | (<0.01) | c1 | 0.0014 | 0.18 | (0.85) | d1 | 0.0151 | 2.01 | (<0.01) |
| a2 | -0.0121 | -1.79 | (0.07) | b2 | -0.0258 | -3.35 | (<0.01) | c2 | -0.0039 | -0.51 | (0.61) | d2 | -0.0041 | -0.53 | (0.60) |
| a3 | 0.0218 | 3.23 | (<0.01) | b3 | -0.0146 | -1.88 | (0.08) | c3 | 0.0021 | 0.28 | (0.77) | d3 | 0.0045 | 0.58 | (0.62) |
| a4 | 0.0007 | 0.10 | (0.92) | b4 | -0.0108 | -1.40 | (0.16) | c4 | 0.0018 | 0.24 | (0.81) | d4 | 0.0036 | 0.46 | (0.57) |
| a5 | 0.0121 | 1.80 | (0.07) | b5 | -0.0009 | -0.12 | (0.90) | c5 | 0.0027 | 0.36 | (0.77) | d5 | -0.0039 | -0.51 | (0.60) |
| | C | -0.0045 | | | | | | | | | | | | | |
| Adj. R-squared | | 0.0202 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0195 | 2.90 | (<0.01) | b1 | 0.0482 | 6.37 | (<0.01) | c1 | -0.0045 | -0.60 | (0.63) | d1 | 0.0340 | 4.57 | (<0.01) |
| a2 | 0.0116 | 1.73 | (0.08) | b2 | 0.1138 | 14.93 | (<0.01) | c2 | 0.0040 | 0.53 | (0.60) | d2 | 0.0043 | 0.57 | (0.62) |
| a3 | 0.0104 | 1.55 | (0.12) | b3 | 0.0656 | 8.57 | (<0.01) | c3 | 0.0058 | 0.76 | (0.44) | d3 | -0.0160 | -2.09 | (<0.01) |
| a4 | 0.0074 | 1.10 | (0.27) | b4 | 0.0582 | 7.63 | (<0.01) | c4 | -0.0002 | -0.03 | (0.80) | d4 | -0.0137 | -1.79 | (0.07) |
| a5 | 0.0066 | 0.99 | (0.32) | b5 | 0.0627 | 8.25 | (<0.01) | c5 | 0.0013 | 0.18 | (0.85) | d5 | -0.0063 | -0.83 | (0.40) |
| | C | 0.0141 | | | | | | | | | | | | | |
| Adj. R-squared | | 0.0399 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0119 | 1.81 | (0.70) | b1 | -0.0428 | -5.77 | (<0.01) | c1 | 0.0149 | 2.01 | (<0.01) | d1 | 0.2456 | 33.61 | (<0.01) |
| a2 | 0.0082 | 1.24 | (0.22) | b2 | 0.0226 | 3.01 | (<0.01) | c2 | 0.0319 | 4.28 | (<0.01) | d2 | 0.0292 | 3.88 | (<0.01) |
| a3 | -0.0032 | -0.48 | (0.63) | b3 | 0.0162 | 2.15 | (<0.01) | c3 | -0.0034 | -0.46 | (0.57) | d3 | 0.0128 | 1.70 | (0.06) |
| a4 | -0.0128 | -1.94 | (0.05) | b4 | 0.0072 | 0.96 | (0.34) | c4 | 0.0205 | 2.75 | (<0.01) | d4 | -0.0013 | -0.17 | (0.86) |
| a5 | 0.0037 | 0.57 | (0.62) | b5 | 0.0190 | 2.55 | (<0.01) | c5 | 0.0064 | 0.88 | (0.37) | d5 | 0.0148 | 1.98 | (<0.01) |
| | C | -0.0022 | | | | | | | | | | | | | |
| Adj. R-squared | | 0.0715 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \nu_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0049 | -0.73 | (0.47) | b1 | -0.0293 | -3.90 | (0.10) | c1 | 0.0698 | 9.27 | (<0.01) | d1 | 0.1207 | 16.30 | (<0.01) |
| a2 | 0.0091 | 1.36 | (0.17) | b2 | 0.0101 | 1.33 | (0.10) | c2 | 0.0182 | 2.41 | (<0.01) | d2 | 0.0851 | 11.18 | (<0.01) |
| a3 | -0.0066 | -0.99 | (0.32) | b3 | 0.0069 | 0.90 | (0.10) | c3 | 0.0222 | 2.94 | (<0.01) | d3 | 0.0115 | 1.51 | (0.13) |
| a4 | -0.0066 | -0.99 | (0.32) | b4 | 0.0119 | 1.57 | (0.10) | c4 | 0.0137 | 1.82 | (0.10) | d4 | 0.0040 | 0.52 | (0.61) |
| a5 | 0.0017 | 0.25 | (0.80) | b5 | 0.0140 | 1.85 | (0.10) | c5 | 0.0351 | 4.77 | (<0.01) | d5 | 0.0061 | 0.80 | (0.42) |
| | C | -0.0045 | | | | | | | | | | | | | |
| Adj. R-squared | | 0.0470 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QR}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

der (sell-initiated hidden) leads to upward(downward) quote revision, or market maker raises(lower) the quote midpoint immediately subsequent to buy(sell) hidden order. Lastly, the initiated trading volume ($\widetilde{TV}$) shows a positive impact on the quotes revision model for all sample periods with less than a 1% significant level. The estimated coefficients are 0.0156, 0.0191, 0.0223, and 0.0151 for week one, six, twelve and eighteen. The positive coefficient estimates of $d_1$ indicate that market maker raises(lower) the quote midpoint immediately subsequent to a buy-initiated trade (sell-initiated trades) volume, the same as the association of $\widetilde{HDV}$ and $\widetilde{QR}$.

Panel B of table 7.5, 7.6, 7.7 and 7.8 present the estimated coefficients of the signed-trade direction process. The coefficients on this panel show the values of $b_i, (i = 1, 2.., 5)$ for lagged trades from all sample periods are positive with 1% significant level. This result indicates that trades are serially correlated. Next, we find the value of $c_1(\widetilde{HDV})$ for week one is the difference from week six and week twelve, however, the coefficient estimates of $c_1$ for week 18 are not statistically significant. The coefficient estimates of $c_1$ is 0.0045 and -0.0020, -0.0033, and -0.0045 for week one, six, twelve and eighteen. This result shows that the association between $\widetilde{TD}$ and $\widetilde{HDV}$ during the low activity trading period(one week to maturity) has a different impact from high trading activity period(six and twelve week to maturity). The positive coefficient estimates $c_1$ during the low-frequency trading period indicates that the prior buy(sell) initiated hidden order leads to buy(sell) initiated the trade, while during the high-frequency trading period this is vice versa. Furthermore, the impact of this parameter is very small as the estimated coefficient is the smallest compared to the estimated coefficients of other variables. Lastly, the coefficient estimates of $d_1$ for week one, six, and twelve show that buy(sell) initiated trading volume is positively correlated with the same signed-trade direction. However, the estimated coefficient of $d_1$ for week eighteen is not statistically significant.

Panel C of table 7.5, 7.6, 7.7 and 7.8 present the estimated coefficients of the hidden order process. First, the estimated coefficients of $a_i$; this work focuses only on the first lag of this variable as lag 2 to 5 has very small coefficients compared to the first lag. There are negative estimated coefficients of $a_1$ for one, six and twelve week to maturity at 1% significant level, however, for eighteen week the p-value is higher than the 10% level. The negative coefficients indicate that when the market maker lowers (rises) mid-quote, the hidden liquidity is increased(decreased). Next, the estimated coefficients of $b_1$ being also negative indicates that the hidden liquidity is increased(decreased) when the previous trade is sell(buy)-initiated. Furthermore, this panel shows the coefficient estimates of $c_i, i = 1, 2.., 5$ for lagged $\widetilde{HDV}$ are all positive with 1%significant level. This result indicates that $\widetilde{HDV}$ are positively autocorrelated. Last, the positive coefficient estimates of $d_1$ for all sample periods indicate that the prior buy(sell) initiated trade volume leads to buy(sell) initiated hidden liquidity. As this study is interested in the $\widetilde{HDV}$ system, then the magnitude of coefficient estimates for each variable is compared. However, this analysis is interested in only the lagged one because the coefficient estimates for $c_2$ to $c_5$ are considerably smaller than $c_1$. In terms of magnitude, the highest coefficient estimates on the $\widetilde{HDV}$ system is $d_1(\widetilde{TV})$ which is 25 and 18 times larger than that of the corresponding impact from $\widetilde{QR}$ and $\widetilde{TD}$, respectively. The high coefficient estimates of $\widetilde{TV}$ indicates that the association between initiated-hidden liquidity and prior initiated-trading volume has the highest degree of the impact on this $\widetilde{HDV}$ system.

Finally, panel D of table 7.5, 7.6, 7.7 and 7.8 present the estimated coefficients of the trade volume process. Firstly, the estimated coefficients of mid-quote return ($a_i$) reveal there is a negative coefficient of $a_1$ for one, six and twelve week to maturity at 1% significant level, however, for the eighteen weeks, the estimated coefficients are not significant. The negative coefficients of $a_1$ indicate that trading volume increases(decrease) when the previous mid-quote return is negative(positive). However, there is the positive coefficient of $b_1$ for week one, six, and twelve at 1% significant level; however, in week eighteen the coefficient is not significant. This positive coefficient of $b_1$ indicates that trading volume increases(decrease) when the previous trade is buy(sell)-initiated. Furthermore, the most interesting coefficient estimates in this system is the value of the signed-hidden order ($c_i$). The estimated coefficient of $c_i, i = 1, 2..., 5$ is significant at 1% level for weeks one, six, twelve and eighteen with a positive coefficient for all lags. This positive coefficient indicates that trade volume increases (decreases) when a lag of hidden order is increased(decreased). Lastly, the coefficient estimates of $d_i(i = 1, 2, .., 5)$ for lagged $\widetilde{TV}$. The result shows positive coefficient estimates for all sample periods, which indicates that the initiated-trading volume is positively autocorrelated.

The result also finds a feedback effect between each variable when I examine the result of the Granger causality test based on our unrestricted VAR system. Table 7.9 presents the results of the Granger causality test that has rejected the null hypothesis with two-way statistics, significant at 1% level for all relationships between each variable. Furthermore, when moving this test to a different sample period, there are similar results for most of E-mini S&P500 trading periods which is from sixteen to one week to maturity. However, for week eighteen and seventeen, the results of the Granger causality test are not significant in some of the feedback effects. This insignificant Granger causality test is caused by the limited number of observations which is around 5,000 and 8,000 updated per week. Conclusively, this result indicates that in our VAR system, the $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$, and $\widetilde{TV}$ have a Granger cause to each other, and also they have feedback influence from each other between sixteen and one week to maturity (see Table7.9).

### 7.4.4 Impulse response functions

This section performs an impulse response function (IRF) analysis on our VAR system. The purpose of IRF analysis is to investigate an influence on one factor giving such a level of shock impact by another factor. The aim of this IRF is to know the realized effect in each VAR system. First, quote revision equation $\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$. Second, trade direction equation $\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$. Third, initiated-hidden liquidity equation $\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$. Finally, initiated-trading volume equation $\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$. This section presents IRF plots for a week-by-week to maturity from eighteen to one week to maturity. In total in this study we have eighteen different plots of IRF, however, in this chapter there are only four plots, which is the IRF for week one, six, twelve, and eighteen. For week one and week eighteen they are presented as a proxy of low-frequency trading period, for week six and week twelve presented as a proxy of the high-frequency trading period. The following section analyzes IRF by a number of the weeks to maturity, however, I present all results of the IRF analysis in Appendix A.2.

Figure 7.3 presents the plot of Impulse response functions analysis of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$, and $\widetilde{TV}$ for one week to maturity in E-mini S&P500 trading. The upper-

**Table 7.9: Granger causality test for our VAR system.**

| Granger causality for one week to maturity | | | | | Granger causality for six weeks to maturity | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: $\widetilde{QR}$ does not Granger Cause $\widetilde{TD}$, $\widetilde{HDV}$, $\widetilde{TV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Signed-Trade Direction | 1,961.390 | 0.000 | *** | Reject | Signed-Trade Direction | 150,355.000 | 0.000 | *** | Reject |
| Signed-Hidden Order | 257.081 | 0.000 | *** | Reject | Signed-Hidden Order | 10,072.100 | 0.000 | *** | Reject |
| Signed-Trade Volume | 286.395 | 0.000 | *** | Reject | Signed-Trade Volume | 8,998.110 | 0.000 | *** | Reject |
| Panel B: $\widetilde{TD}$ does not Granger Cause $\widetilde{QR}$, $\widetilde{HDV}$, $\widetilde{TV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Mid Quote Return | 20,563.100 | 0.000 | *** | Reject | Mid Quote Return | 69,704.200 | 0.000 | *** | Reject |
| Signed-Hidden Order | 8,162.350 | 0.000 | *** | Reject | Signed-Hidden Order | 127,824.000 | 0.000 | *** | Reject |
| Signed-Trade Volume | 6,465.520 | 0.000 | *** | Reject | Signed-Trade Volume | 99,618.000 | 0.000 | *** | Reject |
| Panel C: $\widetilde{HDV}$ does not Granger Cause $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{TV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Mid Quote Return | 2,533.930 | 0.000 | *** | Reject | Mid Quote Return | 21,204.600 | 0.000 | *** | Reject |
| Signed-Trade Direction | 131.766 | 0.000 | *** | Reject | Signed-Trade Direction | 179.368 | 0.000 | *** | Reject |
| Signed-Trade Volume | 1,799.300 | 0.000 | *** | Reject | Signed-Trade Volume | 11,603.300 | 0.000 | *** | Reject |
| Panel D: $\widetilde{TV}$ does not Granger Cause $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Mid Quote Return | 3,528.550 | 0.000 | *** | Reject | Mid Quote Return | 26,268.500 | 0.000 | *** | Reject |
| Signed-Trade Direction | 336.700 | 0.000 | *** | Reject | Signed-Trade Direction | 2,014.380 | 0.000 | *** | Reject |
| Signed-Hidden Order | 125,135.000 | 0.000 | *** | Reject | Signed-Hidden Order | 3,627,085.000 | 0.000 | *** | Reject |
| Observation | 1,978,662 | | | | Observation | 22,089,526 | | | |

**Table 7.9 Continued: Granger causality test for our VAR system.**

| Granger causality for 12 weeks to maturity | | | | | Granger causality for 18 weeks to maturity | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: $\widetilde{QR}$ does not Granger Cause $\widetilde{TD}$, $\widetilde{HDV}$, $\widetilde{TV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Signed-Trade Direction | 95,767.300 | 0.000 | *** | Reject | Signed-Trade Direction | 3.021 | 0.001 | *** | Reject |
| Signed-Hidden Order | 5,016.930 | 0.000 | *** | Reject | Signed-Hidden Order | 1.365 | 0.234 | | Except |
| Signed-Trade Volume | 4,785.660 | 0.000 | *** | Reject | Signed-Trade Volume | 0.764 | 0.576 | | Except |
| Panel B: $\widetilde{TD}$ does not Granger Cause $\widetilde{QR}$, $\widetilde{HDV}$, $\widetilde{TV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Mid Quote Return | 49,209.200 | 0.000 | *** | Reject | Mid Quote Return | 88.243 | 0.001 | *** | Reject |
| Signed-Hidden Order | 71,128.900 | 0.000 | *** | Reject | Signed-Hidden Order | 8.965 | 0.001 | *** | Reject |
| Signed-Trade Volume | 53,161.800 | 0.000 | *** | Reject | Signed-Trade Volume | 4.413 | 0.001 | *** | Reject |
| Panel C: $\widetilde{HDV}$ does not Granger Cause $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{TV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Mid Quote Return | 16,029.100 | 0.000 | *** | Reject | Mid Quote Return | 13.280 | 0.001 | *** | Reject |
| Signed-Trade Direction | 101.002 | 0.000 | *** | Reject | Signed-Trade Direction | 0.173 | 0.973 | | Except |
| Signed-Trade Volume | 5,888.270 | 0.000 | *** | Reject | Signed-Trade Volume | 15.364 | 0.001 | *** | Reject |
| Panel D: $\widetilde{TV}$ does not Granger Cause $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ (Null Hypothesis) | | | | | | | | | |
| Benchmarks | F-Statistic | p-value | | Result | Benchmarks | F-Statistic | p-value | | Result |
| Mid Quote Return | 19,528.900 | 0.000 | *** | Reject | Mid Quote Return | 17.658 | 0.001 | *** | Reject |
| Signed-Trade Direction | 1,120.460 | 0.000 | *** | Reject | Signed-Trade Direction | 5.500 | 0.001 | *** | Reject |
| Signed-Hidden Order | 2,174,954.000 | 0.000 | *** | Reject | Signed-Hidden Order | 236.933 | 0.001 | *** | Reject |
| Observation | 14,428,789 | | | | Observation | 21,867 | | | |

Note: This table presents the result from the Granger causality test for mid-quote return ($\widetilde{QR}$), initiated-trade direction ($\widetilde{TD}$), initiated-hidden order volume ($\widetilde{HDV}$), and initiated-trade volume ($\widetilde{TV}$). The sample is from E-mini S&P500 for 1, 6, 12, and 18 weeks to maturity. The Granger causality is testing whether there is a Granger causality relationship between our observed variables from the VAR system. F-statistic, p-value and test results are shown in the table. *** indicates that the result is significant at 1% level.

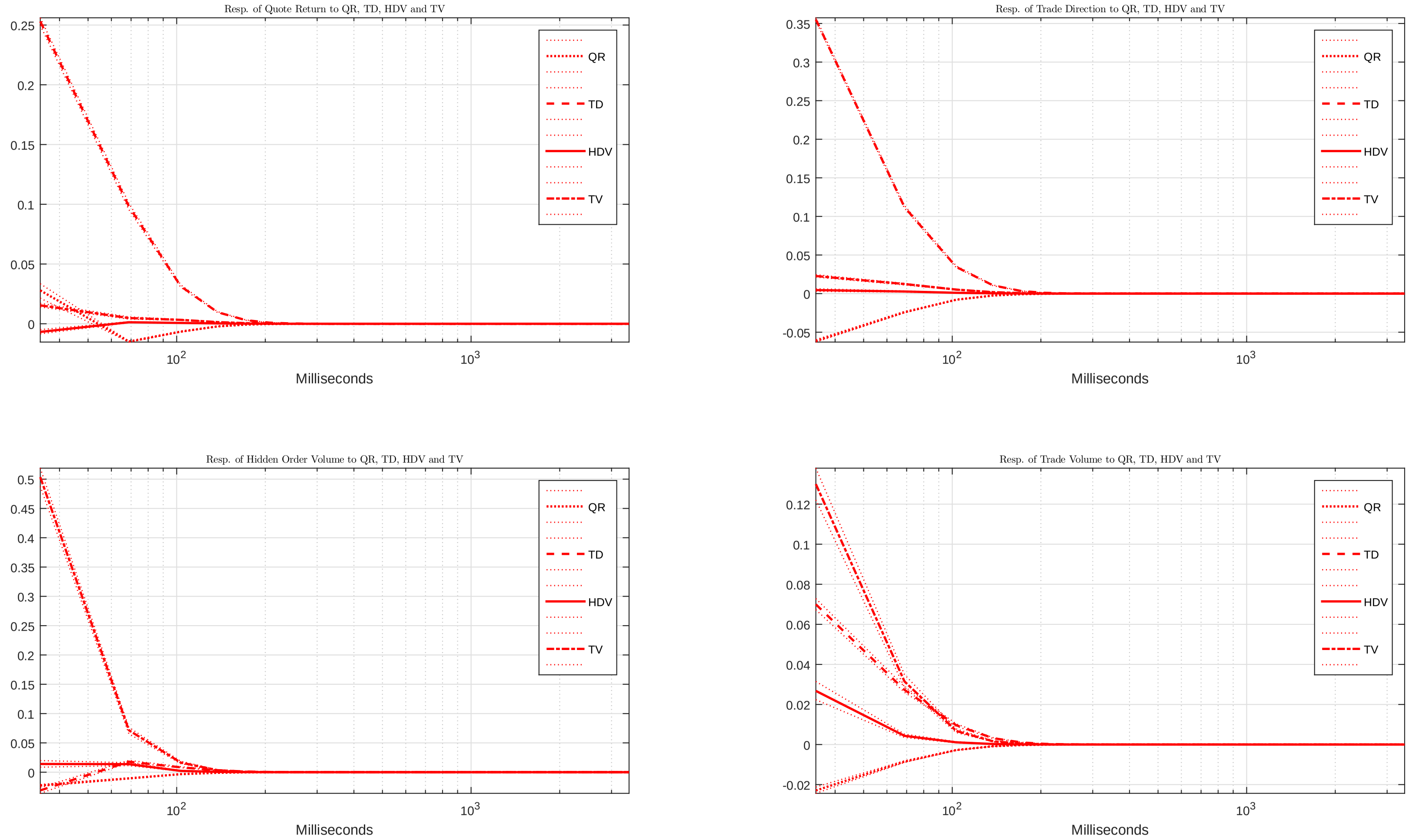



**Figure 7.3: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for one week to maturity.**

Note: This figure presents the results of impulse response analysis of $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$, and $\widetilde{TV}$ to an innovation in observed variables for one week to maturity. The horizontal axis is the time of response to a shock in millisecond time-stamped. The vertical axis, for the top left, is the response of $\widetilde{QR}$, top right is the response of $\widetilde{TD}$, bottom left is the response of $\widetilde{HDV}$, and bottom right is the response of $\widetilde{TV}$. The $\cdots$ line states the impulse response of observed variables to mid-quote return $(\widetilde{QR})$, the – line states the impulse response to initiated-trade direction $(\widetilde{TD})$, the thick line (-) states the impulse response to initiated-hidden order volume $(\widetilde{HDV})$, and -·- states the impulse response to initiated-trade volume $(\widetilde{TV})$. The sample is from E-mini S&P500 for one week to maturity.

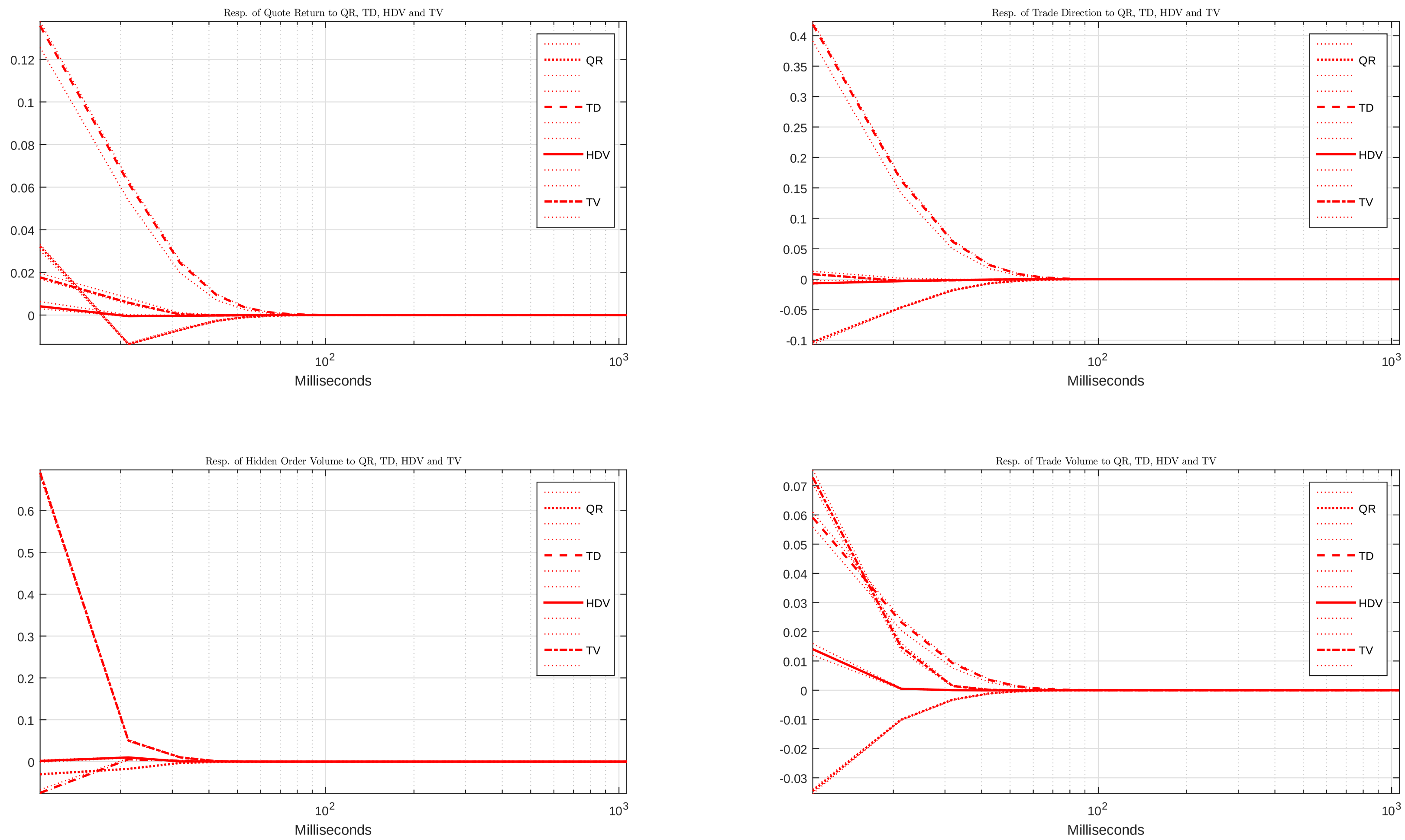



**Figure 7.4: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for six weeks to maturity.**

Note: This figure presents the result of impulse response analysis of $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$, and $\widetilde{TV}$ to an innovation in observed variables for six weeks to maturity. The horizontal axis is the time of response to a shock in millisecond time-stamped. The vertical axis, for the top left, is the response of $\widetilde{QR}$, top right is the response of $\widetilde{TD}$, bottom left is the response of $\widetilde{HDV}$, and bottom right is the response of $\widetilde{TV}$. The $\cdots$ line states the impulse response of observed variables to mid-quote return ($\widetilde{QR}$), the – line states the impulse response to initiated-trade direction ($\widetilde{TD}$), the thick line (-) states the impulse response to initiated-hidden order volume ($\widetilde{HDV}$), and -·- states the impulse response to initiated-trade volume ($\widetilde{TV}$). The sample is from E-mini S&P500 for six week to maturity.

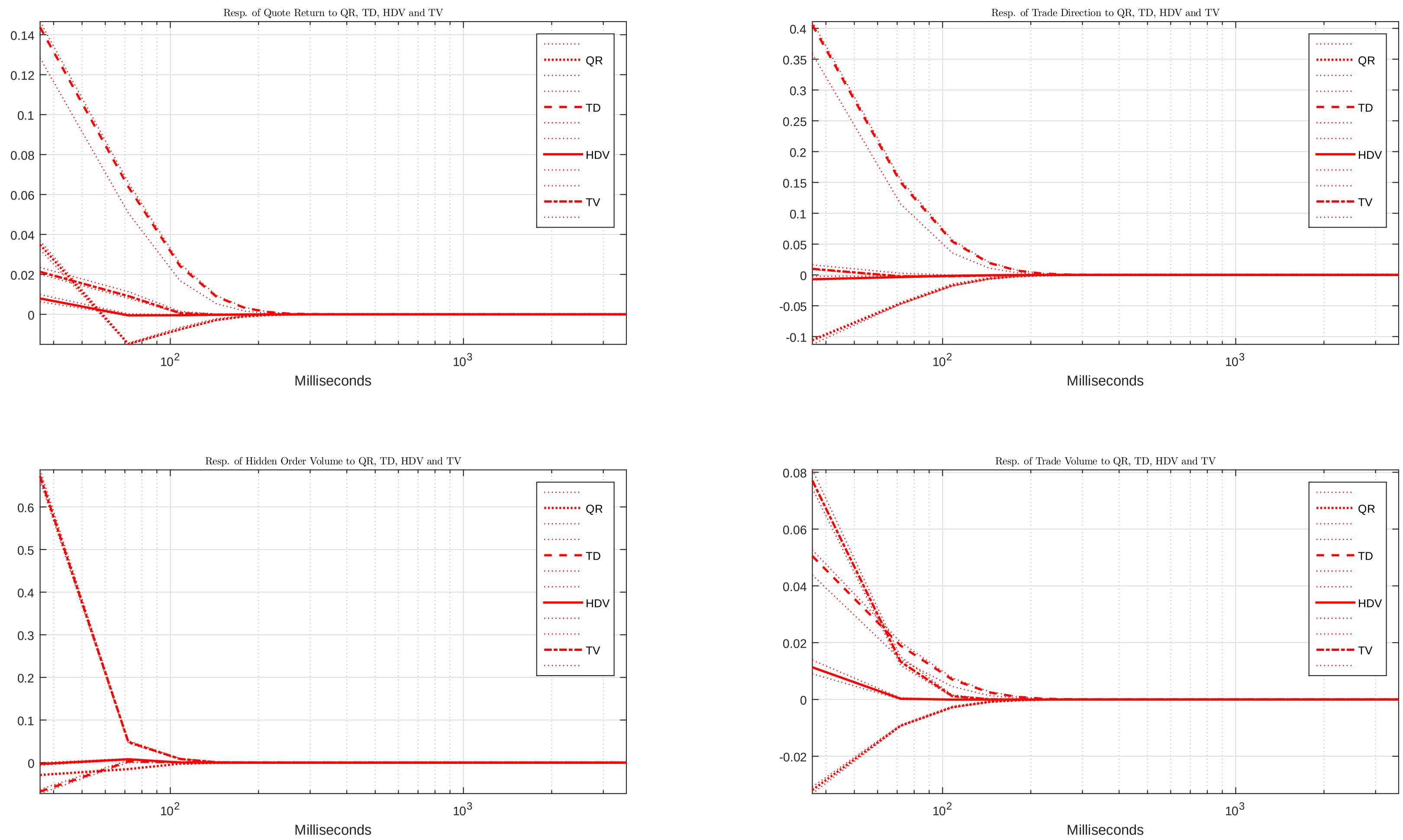



**Figure 7.5: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 12 weeks to maturity.**

Note: This figure presents the result of impulse response analysis of $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$, and $\widetilde{TV}$ to an innovation in observed variables for twelve weeks to maturity. The horizontal axis is the time of response to a shock in millisecond time-stamped. The vertical axis, for the top left, is the response of $\widetilde{QR}$, top right is the response of $\widetilde{TD}$, bottom left is the response of $\widetilde{HDV}$, and bottom right is the response of $\widetilde{TV}$. The $\cdots$ line states the impulse response of observed variables to mid-quote return ($\widetilde{QR}$), the – line states the impulse response to initiated-trade direction ($\widetilde{TD}$), the thick line (-) states the impulse response to initiated-hidden order volume ($\widetilde{HDV}$), and -·- states the impulse response to initiated-trade volume ($\widetilde{TV}$). The sample is from E-mini S&P500 for 12 weeks to maturity.

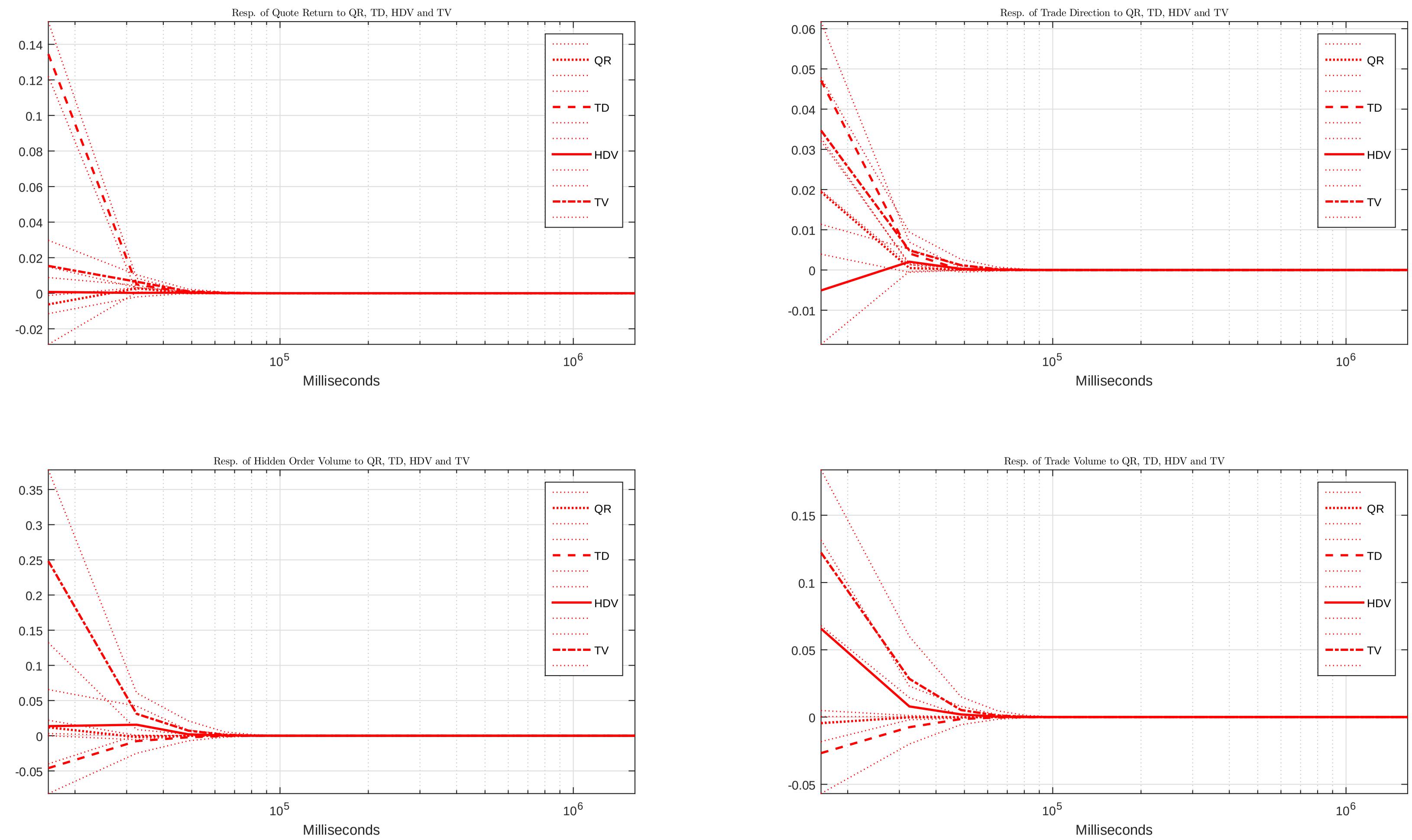



**Figure 7.6: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 18 weeks to maturity.**

Note: This figure presents the results of impulse response analysis of $\widetilde{QR}, \widetilde{TD}, \widetilde{HDV}$, and $\widetilde{TV}$ to an innovation in observed variables for eighteen weeks to maturity. The horizontal axis is the time of response to a shock in millisecond time-stamped. The vertical axis, for the top left, is the response of $\widetilde{QR}$, top right is the response of $\widetilde{TD}$, bottom left is the response of $\widetilde{HDV}$, and bottom right is the response of $\widetilde{TV}$. The $\cdots$ line states the impulse response of observed variables to mid-quote return ($\widetilde{QR}$), the – line states the impulse response to initiated-trade direction ($\widetilde{TD}$), the thick line (-) states the impulse response to initiated-hidden order volume ($\widetilde{HDV}$), and -·- states the impulse response to initiated-trade volume ($\widetilde{TV}$). The sample is from E-mini S&P500 for eighteen weeks to maturity.

left graph of figure 7.3 shows the structural IRF of an innovation in $\widetilde{TD}$, $\widetilde{HDV}$, and $\widetilde{TV}$ to mid-quote return ($\widetilde{QR}$). This subplot clearly indicates that there is a shock to $\widetilde{TD}$ causing $\widetilde{QR}$ to peak at the beginning of the shock then rapidly decrease. A one-unit shock to $\widetilde{TD}$ increases the $\widetilde{QR}$ with a value of 0.25 unit impacts at the beginning of the shock until the effect dies out after roughly 200 milliseconds. The shock in $\widetilde{TD}$ gives the highest response of $\widetilde{QR}$ compared to the shock from $\widetilde{HDV}$, $\widetilde{TV}$ and $\widetilde{QR}$. While $\widetilde{QR}$ has positively impacted caused by a shock of $\widetilde{TD}$, the shock of the $\widetilde{HDV}$ shows a small negative effect on $\widetilde{QR}$. A one unit shock to $\widetilde{HDV}$ decreases $\widetilde{QR}$ with a value of -0.01 unit impacts, then the effect dies out after 80 milliseconds.

The upper-right graph of figure 7.3 shows the structural IRF of an innovation in $\widetilde{QR}$, $\widetilde{HDV}$, and $\widetilde{TV}$ to signed-trade direction ($\widetilde{TD}$). The highest unit impacts to $\widetilde{TD}$ caused by a shock of $\widetilde{TD}$. A one-unit shock to signed-trade direction increases the $\widetilde{TD}$ 0.35 unit impacts, then vanishes until the effect dies out after roughly 300 milliseconds. This can indicate that trade direction tends to be the same direction as a previous trade. Whilst the positive shock to $\widetilde{TD}$, $\widetilde{HDV}$, and $\widetilde{TV}$ causes an increase in signed-trade direction, a one unit shock to $\widetilde{QR}$ decreases the $\widetilde{TD}$ with -0.06 unit impacts at the beginning of the shock, then increases until the effect dies out after 200 milliseconds.

The lower-left graph of figure 7.3 shows the structural IRF of an innovation in $\widetilde{TD}$, $\widetilde{QR}$, and $\widetilde{TV}$ to signed-hidden order volume ($\widetilde{HDV}$). The highest response of $\widetilde{HDV}$ is caused by a shock in $\widetilde{TV}$. A one-unit shock to the $\widetilde{TV}$ causes an increase in $\widetilde{HDV}$ around 0.5 unit impacts then it begins to decrease with an exponentially decaying function and vanishes after 150 milliseconds. Interestingly, the result indicates that a one-unit shock to the $\widetilde{TD}$ causes the decrease in $\widetilde{HDV}$

around 0.05 unit of responses. This result indicates that the change in $\widetilde{TD}$ causes a decrease in signed-hidden order volume. Lastly, these results indicate that all shocks in $\widetilde{TD}$, $\widetilde{QR}$ and $\widetilde{HDV}$ causes to a change of $\widetilde{HDV}$ and the response for the shock dies out after roughly 150 milliseconds.

The lower-right graph of figure 7.3 shows the structural IRF of an innovation in $\widetilde{TD}$, $\widetilde{QR}$, and $\widetilde{HDV}$ to signed-trade volume ($\widetilde{TV}$). The highest response of signed-trade volume ($\widetilde{TV}$) is caused by a shock in $\widetilde{TV}$. A one-unit shock to the $\widetilde{TV}$ causes an increase in the $\widetilde{TV}$ around 0.13 unit of responses, followed by a decrease until the effect dies out after roughly 200 milliseconds. Next, the second highest response of $\widetilde{TV}$ is to the shock of $\widetilde{TD}$ and $\widetilde{HDV}$ with the unit responses of 0.08 and 0.03 respectively. However, a one-unit shock to $\widetilde{HDV}$ causes a decrease in $\widetilde{QR}$ with -0.02 unit responses followed by increase until the effect dies out after 200 milliseconds. Additionally, during this sample period, the responses of $\widetilde{HDV}$ to the shock of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ vanishes after 200 milliseconds.

The following section examines the IRF analysis for the period of 6 and 12 weeks to maturity. During this period, the E-mini trading is in a state of high-frequency trading. The observed trade updated in this week is about 22 million, which is around ten times greater compared to trade updated for one week to maturity (see Table 2.2). However, there is a similar pattern of impulse response between this figure and the figure 7.3 but the unit impacts and the vanishing time of the response is different. Furthermore, for week six and week twelve, the result shows a similar pattern of an impulse response with roughly the same unit impacts and the vanishing time period of responses. Using the same analysis format as explained in the previous section for Figure7.3, I will illustrate the IRF analysis of six weeks to maturity (Figure 7.4) and twelve weeks to maturity

(Figure 7.5) together.

Figure 7.4 and figure 7.5 present the graph of Impulse response functions of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$, and $\widetilde{TV}$ on E-mini S&P500 for six and twelve weeks to maturity, respectively. The upper-left of these figures presents the structural IRF of an intervention in $\widetilde{TD}$, $\widetilde{HDV}$, and $\widetilde{TV}$ to mid-quote return ($\widetilde{QR}$). This graph clearly indicates that the highest response of $\widetilde{QR}$ is caused by a shock of $\widetilde{TD}$. The response reaches a peak at the beginning, a one-unit shock to $\widetilde{TD}$ that causes an increase of $\widetilde{QR}$ with roughly about 0.14 unit responses for both periods. The response of $\widetilde{QR}$ to $\widetilde{TD}$ vanishes after roughly 80 and 300 milliseconds for week six and week twelve respectively. For an innovation in $\widetilde{QR}$, there is a shock to quote return cases an increase in $\widetilde{QR}$, followed by a decrease, and followed by an increase until the effect dies out after 70 and 250 milliseconds for the week six and week twelve respectively. However, there is a small effect from innovation in $\widetilde{HDV}$ and $\widetilde{TV}$ on $\widetilde{QR}$, when the impact vanishes within 100 milliseconds.

The upper-right graph of Figure 7.4 and Figure7.5 present the plot of Impulse response functions of $\widetilde{QR}$, $\widetilde{HDV}$, and $\widetilde{TV}$ to signed-trade direction ($\widetilde{TD}$) for six and twelve weeks to maturity. The highest unit response of $\widetilde{TD}$ is caused by an innovation in $\widetilde{TD}$. A one unit shock to signed-trade direction increases the $\widetilde{TD}$ about 0.40 unit responses, followed by a decrease and vanishes after roughly 80 and 250 milliseconds for week six and week twelve respectively. However, the shock in $\widetilde{QR}$ causes a decrease in $\widetilde{TD}$, followed by increase until the effect dies out after roughly 60 and 200 milliseconds for week six and week twelve respectively. A one-unit shock to the $\widetilde{QR}$ decreases the $\widetilde{TD}$ -0.1 unit responses for both periods. There is a small unit response of $\widetilde{TD}$ to a shock of $\widetilde{HDV}$ and $\widetilde{TV}$ which is near to zero.

The lower-left graph of Figure 7.4 and Figure 7.5 present the structural IRF of an innovation in $\widetilde{TD}$, $\widetilde{QR}$, and $\widetilde{TV}$ to signed-hidden order volume ($\widetilde{HDV}$) for six and twelve weeks to maturity respectively. The highest unit response of signed-hidden order ($\widetilde{HDV}$) volume is caused by an innovation in signed-trade volume ($\widetilde{TV}$). A one unit shock to $\widetilde{TV}$ increases $\widetilde{HDV}$ roughly 0.70 unit responses, followed by a decrease until the effect dies out after 50 and 150 milliseconds for week six and week twelve respectively. Lastly, there is a small unit response of $\widetilde{HDV}$ to a shock of $\widetilde{QR}$, $\widetilde{TD}$, and $\widetilde{HDV}$.

The lower-right graph of Figure 7.4 and Figure 7.5 present the structural IRF of an innovation in $\widetilde{TD}$, $\widetilde{QR}$, and $\widetilde{HDV}$ to signed-trade volume ($\widetilde{TV}$) for six and twelve weeks to maturity respectively. The highest positive unit response of $\widetilde{TV}$ is caused by an innovation in $\widetilde{TV}$. A one unit shock to $\widetilde{TV}$ causes the $\widetilde{TV}$ to reach a peak at the beginning ,at about 0.08 unit responses, followed by a decrease until the impact dies out roughly 50 and 150 milliseconds for week six and week twelve respectively. Next, an innovation in $\widetilde{TD}$ causes an increase in $\widetilde{TV}$ 0.06 and 0.05 unit responses for week six and week twelve respectively. The lowest negative unit responses of $\widetilde{TV}$ is caused by an innovation in $\widetilde{QR}$, a one unit shock to the $\widetilde{QR}$ causes a decrease in $\widetilde{TV}$ around -0.03, followed by an increase until the effect dies out roughly after 60 and 200 milliseconds for six and twelve weeks to maturity respectively. The innovation in $\widetilde{HDV}$ on $\widetilde{TV}$ indicates that the shock to $\widetilde{HDV}$ causes an increase in $\widetilde{TV}$ and dies out very quickly roughly after 30 and 80 milliseconds for week six and week twelve respectively.

Lastly, Figure 7.6 presents the plot of Impulse response functions of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$, and $\widetilde{TV}$ on E-mini S&P500 for eighteen week to maturity. Interestingly, this plot shows a similar pattern of impulse response between this period and a shorter period. However, the figure shows that the unit impact of response and its standard deviation (dashed line – ) are larger than other periods, as a wider range of dotted line is present. For instance, the highest unit response of $\widetilde{QR}$ is caused by an innovation in $\widetilde{TV}$ at about 0.14 unit, followed by a decrease until the effect vanishes after five minutes. A one-unit shock to $\widetilde{TD}$ causes the $\widetilde{TD}$ to peak at 0.05 unit responses from the beginning, followed by a decrease until the impact dies out after seven minutes. For the $\widetilde{HDV}$, an innovation in $\widetilde{TV}$ causes the $\widetilde{HDV}$ to increase 0.25 unit responses, followed by a decrease, then the effect disappears after seven minutes. Finally, the highest unit responses of $\widetilde{TV}$ is caused by a one unit shock to $\widetilde{TV}$ and $\widetilde{HDV}$ with 0.13 and 0.06 unit responses respectively, then the effect dies out after eight minutes.

Overall, these results are consistent with the previous section. For instance, the positive impact from the shock of $\widetilde{TD}$ indicates that the market maker rises (low) their quote when trade appears as a buy(sell)-initiated. Furthermore, our expectation is that the hidden order should have a minor effect on the E-mini S&P500 market, thus, a one unit shock to $\widetilde{HDV}$ causes a minor impact on other variables for all periods compared to the other, and this effect dies out in a very short period. While the unit response to the shock of $\widetilde{HDV}$ shows a minor effect on the change of $\widetilde{QR}, \widetilde{TD}$, and $\widetilde{TV}$, the highest unit responses of the $\widetilde{HDV}$ are caused by a one unit shock in the $\widetilde{TV}$. This pattern is also repeated for all periods.

## 7.5 Summary Chapter

This analysis has undertaken an empirical analysis of the impact of hidden order on the E-mini S&P 500 futures market, based on publicly available limit order book data, which is the most liquid equity index futures trading traded on CME's electronic or Globex trading platform. On this platform, traders can choose to make visible, partially invisible or invisible all of their order called hidden, iceberg order or max show order. The research has documented what motivates those traders to make their order invisible and how this type of order impacts on the market. At first, I believe that they are liquidity-motivated traders by submitting hidden order because this kind of trader tries to reduce the degree of pre-trade transparency. Second, they are price-informativeness motivated traders who try to minimize the price impact of their trades (Buti and Rindi [2008, 2013]). Therefore, I have analyzed the price impact and the effect from buying and selling hidden order to market quality and the structural IRF analysis of hidden order to other market proxies.

This chapter makes two significant contributions. First, I have developed a detection algorithm to detect hidden order in the limit order book using publicly available data. Second, this chapter has identified the impact of hidden order on several dimensions of market quality: spread, effective spread, realized spread, and price impact. Also, I separate this study into high and low trading activity period. The study also provides insights into the role of the hidden order in the financial market. The main finding from the E-mini futures market indicates that an increase in hidden order activity improves the market quality and decreases the level of information asymmetry which is measured via the price impact. Specifically, from the result of price impact, trade prices start moving before hidden order occurs in the market. This may indicate that a trader using an invisible order strategy is a strategic trader who trades based on privileged information and he/she knows when to submit their hidden order. Additionally, the price impact drops to the lowest level after a hidden order is submitted, then bounces back to equilibrium at 20 trades after the order is submitted.

One main innovation in this study is to analyze the impact of the hidden order on E-mini S&P500 index futures via a term-structure. An empirical exercise in this study is set by separating the whole sample starting from 2008 until 2015 into 18 different periods, which is from 18 to 1 weeks before maturity. I have performed a direct test to investigate the association between hidden order and the market quality for all periods. The empirical results show that hidden liquidity positively improves the traditional yardsticks of market quality except inside quote spread. The results of wide-spread may be explained by the behavior of traders who want to submit a non-display order, when they try to reduce the risk of being undercut from a price war by posting their price away from the midpoint to convince the new entering traders to join the queue at prices away from the best bid or ask (Parlour [1998], Buti and Rindi [2008]). Furthermore, the buy-initiated hidden order shows a greater impact than sell-initiated hidden order. The findings also indicate that during high-frequency trading period a hidden order improves market quality more than a low-frequency trading one.

The analysis of the hidden order in a system of the dynamic equations of the VAR model to capture the unique impact of hidden order activity suggests that during high and low-frequency trading periods the shock of $\widetilde{HDV}$ has a similar impact on another variable. Also, the greatest response of $\widetilde{HDV}$ is caused by a one unit shock to the signed-trade volume. However, the impulse response of $\widetilde{HDV}$ to the shock of $\widetilde{TV}$ starts at the beginning and vanishes in a very short time. Lastly, the result also shows that the response of initiated-trade volume $\widetilde{TV}$ increases with the shock of the hidden order ($\widetilde{HDV}$). Therefore, with this result, the impulse response shows a further support for our findings on the two-way relationship between hidden liquidity and trading volume.

Considering the big picture, the evidence in this chapter is more favorable to the notion that hidden order improves the market quality whether in the high or low-frequency trading condition for E-mini S&P500 index future market. The hidden order activity favours trading volume as the $\widetilde{HDV}$ is increased when the $\widetilde{TV}$ increase.

# Chapter 8

# Conclusion

This thesis comprises topics on market transparency and market liquidity in market microstructure study.

The chapter 4 examines the effectiveness of probability of informed trading - PIN proposed by Easley et al. [1996]. PIN is an information asymmetry measure in the market based on the arrival rate of informed and uninformed traders. The model assumes the arrival rate which follows a Poisson process with the Bernoulli probability on the occurrence of two type of news, good or bad news. This chapter provides a primary results of PIN for the Eurodollar futures market. The result indicates that the pattern of the PIN shows a systematic movement regarding the toxicity events, such as, the introduction of CME Globex, the collapse of Lehman Brothers. Interestingly, the PIN results in this futures market is away high and higher than the equity market (circa 0.4 to 0.9 versus 0.1 to 0.5; see Easley et al. [1996]; Aslan et al. [2011]; Abad and Yagüe [2012]).However, other PIN studies of interest rate futures have found something similarly high (see Kim et al. [2014]).

The chapter 5 provides the evidence of LIBOR manipulation and evaluates the PIN around this toxicity event. To detect informed trading, my experiment is somewhat different from other studies in the literature. I catalogue a specific date on the LIBOR manipulation identified by the CFTC and FSA, then the PIN is estimated using a data set of Eurodollar futures market. The result shows that the PIN could have been used as an early warning of unusual activity in the LIBOR reference rate. Extending this, the PIN is computed around the maturity date as a normal event in the futures contract and I investigate the variation of PIN around these events. The result shows that the PIN reacts strongly to certain types of events and the variation is statistically significant relative to the control group. In contrast, for the general long-run variation of the PIN was not statistically significantly different relative to both LIBOR manipulation and the maturity event.

The Chapter 6 compares the effectiveness of PIN and the Volume Synchronize-PIN (VPIN) in determining changes in the information structure and order flow of a Eurodollar futures market around LIBOR manipulation. Unlike chapter 4, this chapter examines the PIN and VPIN as a term-structure analysis instead of calendar time. In addition, I adapt the original VPIN upon Easley et al. [2011] and Easley et al. [2012] by considering volume buckets 20, 50, 100, 200 and I compute the bucket size as a fraction of daily trading volume to avoid the bias of activity or inactivity of the trading period. Also, tick-by-tick data is used instead of one-minute interval, as in original empirical work, because I want to implement this approach with a real HFT environment. Finally, a new bootstrap approach and asymptotic standard error are created on the VPIN and PIN respectively around documented LIBOR manipulation.

The results shows a very strong connection between PIN, VPIN and time to maturity. Both PIN and VPIN vary systematically and in a statistically significant pattern in respect to the term structure of the futures contracts. PIN varies in a v-shaped pattern, with long (2000 to 3500 days) and short maturity (0 to 500 days) contracts having significantly higher PIN and VPIN measurements than intermediate contracts (which are actually the most heavily traded). When I move to the manipulation event, PIN and VPIN shift systematically around a relevant documented LIBOR manipulation event. However, when I build cross sectional averages across events, the result indicates that there is no significant evidence of systematic shifts in either the PIN or VPIN metric.

Lastly, chapter 7 examines how hidden order impacts on the E-Mini S&P 500. The result shows that the price impact relates to hidden order, when the price impact increases when hidden order is submitted and dropped to bottom followed by an increase. Finally, the impact of hidden order to the price declined 20 trades afterwards. Also, I examine the time series relationship between market quality and hidden order. The result shows that the presence of hidden liquidity improves the quality and reduces the level of information asymmetry in the E-Mini market.

# Appendix A

## .1 The VAR analysis on the term-structure of E-Mini S&P500 (Supplement for Chapter 7)

This section provides a supplement information of the VAR results using E-Mini S&P 500 data. We analyze the VAR week-by-week as a term-structure analysis, from eighteen to one week to maturity. Here we show the coefficients for all periods, however, the result for week eighteen, twelve, six and one is showed in chapter 7.

**Table 1: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Two week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0279 | 109.57 | (<0.01) | b1 | 0.1340 | 296.51 | (<0.01) | c1 | 0.0047 | 14.14 | (<0.01) | d1 | 0.0148 | 59.46 | (<0.01) |
| a2 | 0.0028 | 10.86 | (<0.01) | b2 | -0.0085 | -17.28 | (<0.01) | c2 | 0.0032 | 9.56 | (<0.01) | d2 | 0.0061 | 18.55 | (<0.01) |
| a3 | -0.0035 | -13.65 | (<0.01) | b3 | -0.0053 | -10.47 | (<0.01) | c3 | 0.0014 | 4.36 | (<0.01) | d3 | 0.0023 | 7.08 | (<0.01) |
| a4 | -0.0040 | -15.71 | (<0.01) | b4 | -0.0034 | -6.81 | (<0.01) | c4 | 0.0022 | 6.77 | (<0.01) | d4 | 0.0020 | 6.07 | (<0.01) |
| a5 | -0.0042 | -17.76 | (<0.01) | b5 | -0.0027 | -5.99 | (<0.01) | c5 | 0.0019 | 7.51 | (<0.01) | d5 | 0.0007 | 1.98 | (0.04) |
| | C | -0.0006 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0173 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1018 | -684.50 | (<0.01) | b1 | 0.1340 | 296.51 | (<0.01) | c1 | 0.0047 | 14.14 | (<0.01) | d1 | 0.0148 | 59.46 | (<0.01) |
| a2 | -0.0371 | -246.40 | (<0.01) | b2 | -0.0085 | -17.28 | (<0.01) | c2 | 0.0032 | 9.56 | (<0.01) | d2 | 0.0061 | 18.55 | (<0.01) |
| a3 | -0.0136 | -90.47 | (<0.01) | b3 | -0.0053 | -10.47 | (<0.01) | c3 | 0.0014 | 4.36 | (<0.01) | d3 | 0.0023 | 7.08 | (<0.01) |
| a4 | -0.0012 | -8.09 | (<0.01) | b4 | -0.0034 | -6.81 | (<0.01) | c4 | 0.0022 | 6.77 | (<0.01) | d4 | 0.0020 | 6.07 | (<0.01) |
| a5 | 0.0103 | 74.46 | (<0.01) | b5 | -0.0027 | -5.99 | (<0.01) | c5 | 0.0019 | 7.51 | (<0.01) | d5 | 0.0007 | 1.98 | (0.04) |
| | C | -0.0004 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6644 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0279 | -147.67 | (<0.01) | b1 | -0.0595 | -177.51 | (<0.01) | c1 | -0.0041 | -16.79 | (<0.01) | d1 | 0.6670 | 3620.15 | (<0.01) |
| a2 | -0.0074 | -38.84 | (<0.01) | b2 | 0.0595 | 162.75 | (<0.01) | c2 | 0.0059 | 24.05 | (<0.01) | d2 | 0.0159 | 64.74 | (<0.01) |
| a3 | -0.0016 | -8.54 | (<0.01) | b3 | 0.0225 | 60.45 | (<0.01) | c3 | 0.0045 | 18.29 | (<0.01) | d3 | 0.0028 | 11.21 | (<0.01) |
| a4 | 0.0008 | 4.15 | (<0.01) | b4 | 0.0137 | 37.60 | (<0.01) | c4 | 0.0039 | 15.90 | (<0.01) | d4 | 0.0019 | 7.86 | (<0.01) |
| a5 | 0.0037 | 20.80 | (<0.01) | b5 | 0.0110 | 32.85 | (<0.01) | c5 | 0.0048 | 26.18 | (<0.01) | d5 | 0.0015 | 6.07 | (<0.01) |
| | C | -0.0002 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4613 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0338 | -135.23 | (<0.01) | b1 | 0.0657 | 148.04 | (<0.01) | c1 | 0.0156 | 48.15 | (<0.01) | d1 | 0.0746 | 305.96 | (<0.01) |
| a2 | -0.0111 | -43.96 | (<0.01) | b2 | 0.0699 | 144.52 | (<0.01) | c2 | 0.0121 | 37.31 | (<0.01) | d2 | 0.0223 | 68.61 | (<0.01) |
| a3 | -0.0040 | -15.79 | (<0.01) | b3 | 0.0272 | 55.11 | (<0.01) | c3 | 0.0080 | 24.74 | (<0.01) | d3 | 0.0144 | 44.36 | (<0.01) |
| a4 | -0.0007 | -2.62 | (<0.01) | b4 | 0.0130 | 26.94 | (<0.01) | c4 | 0.0057 | 17.71 | (<0.01) | d4 | 0.0130 | 39.86 | (<0.01) |
| a5 | 0.0030 | 12.82 | (<0.01) | b5 | 0.0108 | 24.28 | (<0.01) | c5 | 0.0126 | 51.89 | (<0.01) | d5 | 0.0116 | 35.54 | (<0.01) |
| | C | 0.0009 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0569 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

## Table 2: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Three week to maturity

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0215 | 95.93 | (<0.01) | b1 | 0.1275 | 318.71 | (<0.01) | c1 | 0.0048 | 16.03 | (<0.01) | d1 | 0.0129 | 59.18 | (<0.01) |
| a2 | 0.0026 | 11.52 | (<0.01) | b2 | -0.0026 | -6.08 | (<0.01) | c2 | 0.0059 | 20.04 | (<0.01) | d2 | 0.0095 | 31.87 | (<0.01) |
| a3 | -0.0017 | -7.54 | (<0.01) | b3 | -0.0037 | -8.44 | (<0.01) | c3 | 0.0043 | 14.60 | (<0.01) | d3 | 0.0050 | 16.87 | (<0.01) |
| a4 | -0.0037 | -16.32 | (<0.01) | b4 | -0.0043 | -9.97 | (<0.01) | c4 | 0.0043 | 14.52 | (<0.01) | d4 | 0.0038 | 12.65 | (<0.01) |
| a5 | -0.0051 | -24.33 | (<0.01) | b5 | -0.0046 | -11.40 | (<0.01) | c5 | 0.0051 | 23.71 | (<0.01) | d5 | 0.0030 | 10.18 | (<0.01) |
| | C | 0.0005 | | | | | | | | | | | | | |
| A | dj.$R^2$ | 0.0176 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1067 | -823.70 | (<0.01) | b1 | 0.3908 | 1687.99 | (<0.01) | c1 | -0.0019 | -10.95 | (<0.01) | d1 | 0.0145 | 114.98 | (<0.01) |
| a2 | -0.0432 | -328.64 | (<0.01) | b2 | 0.2519 | 1002.41 | (<0.01) | c2 | -0.0019 | -10.82 | (<0.01) | d2 | 0.0058 | 33.76 | (<0.01) |
| a3 | -0.0176 | -133.99 | (<0.01) | b3 | 0.1128 | 440.01 | (<0.01) | c3 | -0.0018 | -10.54 | (<0.01) | d3 | 0.0048 | 28.06 | (<0.01) |
| a4 | -0.0036 | -28.01 | (<0.01) | b4 | 0.0716 | 285.04 | (<0.01) | c4 | -0.0012 | -6.90 | (<0.01) | d4 | 0.0032 | 18.72 | (<0.01) |
| a5 | 0.0092 | 75.99 | (<0.01) | b5 | 0.0603 | 261.12 | (<0.01) | c5 | 0.0001 | 0.53 | (¡0.01) | d5 | 0.0013 | 7.60 | (<0.01) |
| | C | -0.0012 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6700 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0287 | -177.17 | (<0.01) | b1 | -0.0662 | -228.51 | (<0.01) | c1 | -0.0013 | -5.88 | (<0.01) | d1 | 0.6889 | 4377.46 | (<0.01) |
| a2 | -0.0088 | -53.69 | (<0.01) | b2 | 0.0593 | 188.47 | (<0.01) | c2 | 0.0044 | 20.52 | (<0.01) | d2 | 0.0136 | 63.10 | (<0.01) |
| a3 | -0.0026 | -15.57 | (<0.01) | b3 | 0.0225 | 70.24 | (<0.01) | c3 | 0.0040 | 18.39 | (<0.01) | d3 | 0.0034 | 15.85 | (<0.01) |
| a4 | 0.0006 | 3.70 | (<0.01) | b4 | 0.0144 | 45.64 | (<0.01) | c4 | 0.0035 | 16.06 | (<0.01) | d4 | 0.0017 | 7.78 | (<0.01) |
| a5 | 0.0032 | 20.90 | (<0.01) | b5 | 0.0111 | 38.53 | (<0.01) | c5 | 0.0035 | 22.53 | (<0.01) | d5 | 0.0004 | 1.78 | (<0.01) |
| | C | 0.0000 | | | | | | | | | | | | | |
| | Adj.$R^2$ | 0.4870 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0340 | -154.46 | (<0.01) | b1 | 0.0593 | 150.54 | (<0.01) | c1 | 0.0145 | 49.64 | (<0.01) | d1 | 0.0814 | 380.43 | (<0.01) |
| a2 | -0.0135 | 60.38 | (<0.01) | b2 | 0.0657 | 153.69 | (<0.01) | c2 | 0.0116 | 39.68 | (<0.01) | d2 | 0.0232 | 79.36 | (<0.01) |
| a3 | -0.0059 | -26.33 | (<0.01) | b3 | 0.0247 | 56.72 | (<0.01) | c3 | 0.0066 | 22.50 | (<0.01) | d3 | 0.0151 | 51.48 | (<0.01) |
| a4 | -0.0017 | -7.48 | (<0.01) | b4 | 0.0136 | 31.93 | (<0.01) | c4 | 0.0061 | 20.80 | (<0.01) | d4 | 0.0132 | 44.96 | (<0.01) |
| a5 | 0.0030 | 14.48 | (<0.01) | b5 | 0.0108 | 27.60 | (<0.01) | c5 | 0.0115 | 53.81 | (<0.01) | d5 | 0.0086 | 29.33 | (<0.01) |
| | C | -0.0010 | | | | | | | | | | | | | |
| | Adj.$R^2$ | 0.0526 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 3: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Four week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0297 | 123.63 | (<0.01) | b1 | 0.1232 | 292.23 | (<0.01) | c1 | 0.0075 | 24.54 | (<0.01) | d1 | 0.0222 | 94.82 | (<0.01) |
| a2 | 0.0057 | 23.22 | (<0.01) | b2 | -0.0007 | -1.63 | (0.10) | c2 | 0.0054 | 17.52 | (<0.01) | d2 | 0.0108 | 35.33 | (<0.01) |
| a3 | 0.0014 | 5.91 | (<0.01) | b3 | -0.0048 | -10.25 | (<0.01) | c3 | 0.0037 | 11.99 | (<0.01) | d3 | 0.0069 | 22.51 | (<0.01) |
| a4 | -0.0024 | -10.14 | (<0.01) | b4 | -0.0053 | -11.54 | (<0.01) | c4 | 0.0036 | 11.76 | (<0.01) | d4 | 0.0055 | 17.92 | (<0.01) |
| a5 | -0.0034 | -15.24 | (<0.01) | b5 | -0.0057 | -13.48 | (<0.01) | c5 | 0.0047 | 20.07 | (<0.01) | d5 | 0.0037 | 11.98 | (<0.01) |
| | C | 0.0003 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0179 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1062 | -750.67 | (<0.01) | b1 | 0.3821 | 1538.48 | (<0.01) | c1 | -0.0020 | -10.94 | (<0.01) | d1 | 0.0138 | 99.86 | (<0.01) |
| a2 | -0.0419 | -292.27 | (<0.01) | b2 | 0.2532 | 943.37 | (<0.01) | c2 | -0.0018 | -9.75 | (<0.01) | d2 | 0.0053 | 29.15 | (<0.01) |
| a3 | -0.0173 | -120.69 | (<0.01) | b3 | 0.1146 | 418.73 | (<0.01) | c3 | -0.0017 | -9.56 | (<0.01) | d3 | 0.0040 | 21.98 | (<0.01) |
| a4 | -0.0027 | -19.19 | (<0.01) | b4 | 0.0734 | 273.70 | (<0.01) | c4 | -0.0016 | -9.13 | (<0.01) | d4 | 0.0028 | 15.48 | (<0.01) |
| a5 | 0.0099 | 74.77 | (<0.01) | b5 | 0.0619 | 249.90 | (<0.01) | c5 | -0.0002 | -1.19 | 0.23 | d5 | 0.0016 | 8.66 | (<0.01) |
| | C | -0.0011 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6586 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0294 | -161.58 | (<0.01) | b1 | -0.0587 | -183.89 | (<0.01) | c1 | -0.0006 | -2.51 | (<0.01) | d1 | 0.6517 | 3676.22 | (<0.01) |
| a2 | -0.0087 | -47.05 | (<0.01) | b2 | 0.0573 | 165.94 | (<0.01) | c2 | 0.0053 | 22.91 | (<0.01) | d2 | 0.0137 | 58.83 | (<0.01) |
| a3 | -0.0027 | -14.51 | (<0.01) | b3 | 0.0233 | 66.17 | (<0.01) | c3 | 0.0031 | 13.53 | (<0.01) | d3 | 0.0033 | 14.35 | (<0.01) |
| a4 | 0.0004 | 2.22 | (<0.01) | b4 | 0.0143 | 41.34 | (<0.01)) | c4 | 0.0017 | 7.31 | (<0.01) | d4 | 0.0026 | 11.02 | (<0.01) |
| a5 | 0.0031 | 17.96 | (<0.01) | b5 | 0.0115 | 36.16 | (<0.01) | c5 | 0.0030 | 17.14 | (<0.01) | d5 | 0.0022 | 9.53 | (<0.01) |
| | C | -0.0006 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4390 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0345 | -145.57 | (<0.01) | b1 | 0.0567 | 136.22 | (<0.01) | c1 | 0.0138 | 45.57 | (<0.01) | d1 | 0.0753 | 326.00 | (<0.01) |
| a2 | -0.0137 | -56.82 | (<0.01) | b2 | 0.0667 | 148.42 | (<0.01) | c2 | 0.0100 | 33.12 | (<0.01) | d2 | 0.0199 | 65.58 | (<0.01) |
| a3 | -0.0064 | -26.75 | (<0.01) | b3 | 0.0237 | 51.78 | (<0.01) | c3 | 0.0077 | 25.46 | (<0.01) | d3 | 0.0135 | 44.56 | (<0.01) |
| a4 | -0.0019 | -7.90 | (<0.01) | b4 | 0.0147 | 32.64 | (<0.01) | c4 | 0.0045 | 15.05 | (<0.01) | d4 | 0.0110 | 36.32 | (<0.01) |
| a5 | 0.0020 | 8.98 | (<0.01) | b5 | 0.0112 | 27.02 | (<0.01) | c5 | 0.0099 | 43.14 | (<0.01) | d5 | 0.0105 | 34.79 | (<0.01) |
| | C | -0.0002 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0484 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

## Table 4: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Five week to maturity

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0254 | 104.72 | (<0.01) | b1 | 0.1275 | 298.52 | (<0.01) | c1 | 0.0071 | 22.78 | (<0.01) | d1 | 0.0195 | 83.12 | (<0.01) |
| a2 | 0.0039 | 15.65 | (<0.01) | b2 | -0.0025 | -5.41 | (<0.01) | c2 | 0.0052 | 16.90 | (<0.01) | d2 | 0.0105 | 33.71 | (<0.01) |
| a3 | -0.0002 | -0.92 | (<0.01) | b3 | -0.0054 | -11.48 | (<0.01) | c3 | 0.0037 | 12.05 | (<0.01) | d3 | 0.0068 | 21.82 | (<0.01) |
| a4 | -0.0034 | -14.07 | (<0.01) | b4 | -0.0058 | -12.57 | (<0.01) | c4 | 0.0028 | 9.03 | (<0.01) | d4 | 0.0047 | 15.27 | (<0.01) |
| a5 | -0.0040 | -17.88 | (<0.01) | b5 | -0.0058 | -13.60 | (<0.01) | c5 | 0.0045 | 19.20 | (<0.01) | d5 | 0.0037 | 11.82 | (<0.01) |
| | C | 0.0006 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0177 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1098 | -771.89 | (<0.01) | b1 | 0.3795 | 1514.45 | (<0.01) | c1 | -0.0023 | -12.87 | (<0.01) | d1 | 0.0136 | 98.54 | (<0.01) |
| a2 | -0.0440 | -304.50 | (<0.01) | b2 | 0.2558 | 944.54 | (<0.01) | c2 | -0.0023 | -12.40 | (<0.01) | d2 | 0.0055 | 29.94 | (<0.01) |
| a3 | -0.0185 | -128.32 | (<0.01) | b3 | 0.1156 | 418.32 | (<0.01) | c3 | -0.0017 | -9.12 | (<0.01) | d3 | 0.0045 | 24.80 | (<0.01) |
| a4 | -0.0034 | -23.93 | (<0.01) | b4 | 0.0731 | 270.06 | (<0.01) | c4 | -0.0014 | -7.74 | (<0.01) | d4 | 0.0030 | 16.60 | (<0.01) |
| a5 | 0.0092 | 68.92 | (<0.01) | b5 | 0.0613 | 245.11 | (¡0.01) | c5 | 0.0002 | 1.21 | (0.27) | d5 | 0.0014 | 7.68 | (<0.01) |
| | C | -0.0011 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6612 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0314 | -171.95 | (<0.01) | b1 | -0.0629 | -195.59 | (<0.01) | c1 | 0.0002 | 0.81 | (0.44) | d1 | 0.6584 | 3723.76 | (<0.01) |
| a2 | -0.0093 | -50.17 | (<0.01) | b2 | 0.0586 | 168.68 | (<0.01) | c2 | 0.0067 | 28.55 | (<0.01) | d2 | 0.0109 | 46.85 | (<0.01) |
| a3 | -0.0034 | -18.40 | (<0.01) | b3 | 0.0236 | 66.52 | (<0.01) | c3 | 0.0035 | 14.93 | (<0.01) | d3 | 0.0009 | 3.79 | (<0.01) |
| a4 | 0.0003 | 1.39 | (0.16) | b4 | 0.0140 | 40.23 | (<0.01) | c4 | 0.0027 | 11.65 | (<0.01) | d4 | 0.0019 | 8.22 | (<0.01) |
| a5 | 0.0031 | 17.97 | (<0.01) | b5 | 0.0122 | 38.03 | (<0.01) | c5 | 0.0034 | 19.19 | (<0.01)) | d5 | 0.0014 | 6.12 | (<0.01) |
| | C | -0.0006 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4461 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0364 | -151.78 | (<0.01) | b1 | 0.0536 | 126.97 | (<0.01) | c1 | 0.0122 | 39.77 | (<0.01) | d1 | 0.0692 | 298.27 | (<0.01) |
| a2 | -0.0143 | -58.81 | (<0.01) | b2 | 0.0678 | 148.70 | (<0.01) | c2 | 0.0106 | 34.78 | (<0.01) | d2 | 0.0213 | 69.49 | (<0.01) |
| a3 | -0.0066 | -27.03 | (<0.01) | b3 | 0.0257 | 55.25 | (<0.01) | c3 | 0.0068 | 22.31 | (<0.01) | d3 | 0.0129 | 42.00 | (<0.01) |
| a4 | -0.0018 | -7.52 | (<0.01) | b4 | 0.0140 | 30.74 | (<0.01) | c4 | 0.0063 | 20.73 | (<0.01) | d4 | 0.0123 | 40.05 | (<0.01) |
| a5 | 0.0020 | 8.95 | (<0.01) | b5 | 0.0112 | 26.65 | (<0.01) | c5 | 0.0099 | 42.60 | (<0.01) | d5 | 0.0083 | 27.21 | (<0.01) |
| | C | -0.0002 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0465 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 5: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Seven week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0275 | 126.00 | (<0.01) | b1 | 0.1288 | 332.89 | (<0.01) | c1 | 0.0056 | 19.88 | (<0.01) | d1 | 0.0169 | 79.54 | (<0.01) |
| a2 | 0.0060 | 27.26 | (<0.01) | b2 | -0.0048 | -11.46 | (<0.01) | c2 | 0.0047 | 16.76 | (<0.01) | d2 | 0.0095 | 33.72 | (<0.01) |
| a3 | 0.0011 | 5.16 | (<0.01) | b3 | -0.0055 | -12.85 | (<0.01) | c3 | 0.0025 | 9.02 | <0.01(<0.01) | d3 | 0.0053 | 18.90 | (<0.01) |
| a4 | -0.0026 | -11.87 | (<0.01) | b4 | -0.0056 | -13.24 | (<0.01) | c4 | 0.0021 | 7.59 | (<0.01) | d4 | 0.0049 | 17.49 | (<0.01) |
| a5 | -0.0030 | -14.67 | (<0.01) | b5 | -0.0053 | -13.75 | (0.07) | c5 | 0.0038 | 17.97 | (<0.01) | d5 | 0.0033 | 11.63 | (<0.01) |
| C | 0.0008 | | | | | | | | | | | | | | |
| Adj. $R^2$ | 0.0168 | | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1045 | -822.04 | (<0.01) | b1 | 0.3901 | 1730.70 | (<0.01) | c1 | -0.0023 | -13.76 | (<0.01) | d1 | 0.0122 | 98.56 | (<0.01) |
| a2 | -0.0415 | -321.81 | (<0.01) | b2 | 0.2537 | 1037.66 | (<0.01) | c2 | -0.0012 | -7.49 | (<0.01) | d2 | 0.0054 | 32.96 | (<0.01) |
| a3 | -0.0168 | -130.38 | (<0.01) | b3 | 0.1130 | 452.85 | (<0.01) | c3 | -0.0014 | -8.77 | (<0.01) | d3 | 0.0035 | 21.34 | (<0.01) |
| a4 | -0.0030 | -23.50 | (<0.01) | b4 | 0.0704 | 288.25 | (<0.01) | c4 | -0.0007 | -4.21 | (<0.01) | d4 | 0.0030 | 18.03 | (<0.01) |
| a5 | 0.0091 | 76.13 | (<0.01) | b5 | 0.0593 | 263.75 | (<0.01) | c5 | 0.0002 | 1.79 | (<0.01) | d5 | 0.0009 | 5.60 | (<0.01) |
| C | -0.0014 | | | | | | | | | | | | | | |
| Adj. $R^2$ | 0.6655 | | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0288 | -176.73 | (<0.01) | b1 | 0.0579 | 151.50 | (<0.01) | c1 | 0.0123 | 44.35 | (<0.01) | d1 | 0.0718 | 343.25 | (<0.01) |
| a2 | -0.0082 | -49.61 | (<0.01) | b2 | 0.0656 | 158.42 | (<0.01) | c2 | 0.0099 | 35.57 | (<0.01) | d2 | 0.0225 | 80.98 | (<0.01) |
| a3 | -0.0029 | -17.83 | (<0.01) | b3 | 0.0255 | 60.34 | (<0.01) | c3 | 0.0066 | 23.85 | (<0.01) | d3 | 0.0138 | 49.44 | (<0.01) |
| a4 | 0.0005 | 2.82 | (<0.01) | b4 | 0.0151 | 36.45 | (<0.01) | c4 | 0.0059 | 21.40 | (<0.01) | d4 | 0.0121 | 43.58 | (<0.01) |
| a5 | 0.0032 | 20.62 | (<0.01) | b5 | 0.0100 | 26.20 | (<0.01) | c5 | 0.0104 | 50.01 | (<0.01) | d5 | 0.0086 | 31.10 | (<0.01) |
| C | -0.0005 | | | | | | | | | | | | | | |
| Adj. $R^2$ | 0.4538 | | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0338 | -156.92 | (<0.01) | b1 | 0.0579 | 151.50 | (<0.01) | c1 | 0.0123 | 44.35 | (<0.01) | d1 | 0.0718 | 343.25 | (<0.01) |
| a2 | -0.0135 | -61.88 | (<0.01) | b2 | 0.0656 | 158.42 | (<0.01) | c2 | 0.0099 | 35.57 | (<0.01) | d2 | 0.0225 | 80.98 | (<0.01) |
| a3 | -0.0061 | -27.77 | (<0.01) | b3 | 0.0255 | 60.34 | (<0.01) | c3 | 0.0066 | 23.85 | (<0.01) | d3 | 0.0138 | 49.44 | (<0.01) |
| a4 | -0.0011 | -5.14 | (<0.01) | b4 | 0.0151 | 36.45 | (<0.01) | c4 | 0.0059 | 21.40 | (<0.01) | d4 | 0.0121 | 43.58 | (<0.01) |
| a5 | 0.0024 | 11.74 | (<0.01) | b5 | 0.0100 | 26.20 | (<0.01) | c5 | 0.0104 | 50.01 | (<0.01) | d5 | 0.0086 | 31.10 | (<0.01) |
| C | 0.0004 | | | | | | | | | | | | | | |
| Adj. $R^2$ | 0.0481 | | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 6: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Eight week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0261 | 118.58 | (<0.01) | b1 | 0.1369 | 350.13 | (<0.01) | c1 | 0.0041 | 14.11 | (<0.01) | d1 | 0.0140 | 65.90 | (<0.01) |
| a2 | 0.0060 | 26.65 | (<0.01) | b2 | -0.0048 | -11.20 | (<0.01) | c2 | 0.0045 | 15.35 | (<0.01) | d2 | 0.0102 | 34.80 | (<0.01) |
| a3 | 0.0002 | 0.68 | (<0.01) | b3 | -0.0055 | -12.74 | (<0.01) | c3 | 0.0030 | 10.38 | (<0.01) | d3 | 0.0046 | 15.76 | (<0.01) |
| a4 | -0.0013 | -5.93 | (<0.01) | b4 | -0.0054 | -12.77 | (<0.01) | c4 | 0.0026 | 8.91 | (<0.01) | d4 | 0.0039 | 13.36 | (<0.01) |
| a5 | -0.0042 | -20.27 | (<0.01) | b5 | -0.0065 | -16.57 | (<0.01) | c5 | 0.0039 | 18.44 | (<0.01) | d5 | 0.0032 | 10.89 | (<0.01) |
| C | | 0.0009 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0183 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1068 | -832.75 | (<0.01) | b1 | 0.3983 | 1749.54 | (<0.01) | c1 | -0.0018 | -10.65 | (<0.01) | d1 | 0.0136 | 109.99 | (<0.01) |
| a2 | -0.0416 | -319.67 | (<0.01) | b2 | 0.2525 | 1017.50 | (<0.01) | c2 | -0.0015 | -8.88 | (<0.01) | d2 | 0.0050 | 29.64 | (<0.01) |
| a3 | -0.0163 | -125.27 | (<0.01) | b3 | 0.1097 | 433.59 | (<0.01) | c3 | -0.0008 | -4.58 | (<0.01) | d3 | 0.0046 | 26.79 | (<0.01) |
| a4 | -0.0030 | -23.19 | (<0.01) | b4 | 0.0674 | 271.67 | (<0.01) | c4 | -0.0004 | -2.43 | (<0.01) | d4 | 0.0028 | 16.21 | (<0.01) |
| a5 | 0.0092 | 76.96 | (<0.01) | b5 | 0.0562 | 247.15 | (<0.01) | c5 | 0.0009 | 7.42 | (<0.01) | d5 | 0.0012 | 7.07 | (<0.01) |
| C | | -0.0013 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.6660 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0290 | -181.52 | (<0.01) | b1 | -0.0644 | -227.02 | (<0.01) | c1 | -0.0039 | -18.50 | (<0.01) | d1 | 0.6890 | 4460.47 | (<0.01) |
| a2 | -0.0078 | -48.26 | (<0.01) | b2 | 0.0592 | 191.53 | (<0.01) | c2 | 0.0065 | 30.85 | (<0.01) | d2 | 0.0165 | 78.07 | (<0.01) |
| a3 | -0.0021 | -13.24 | (<0.01) | b3 | 0.0221 | 70.06 | (<0.01) | c3 | 0.0039 | 18.30 | (<0.01) | d3 | 0.0013 | 6.06 | (<0.01) |
| a4 | 0.0004 | 2.29 | (<0.01) | b4 | 0.0129 | 41.63 | (<0.01) | c4 | 0.0038 | 17.99 | (<0.01) | d4 | 0.0004 | 2.02 | (<0.01) |
| a5 | 0.0032 | 21.25 | (<0.01) | b5 | 0.0105 | 37.06 | (<0.01) | c5 | 0.0037 | 24.15 | (<0.01) | d5 | 0.0009 | 4.02 | (<0.01) |
| C | | -0.0006 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.4867 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0339 | -156.28 | (<0.01) | b1 | 0.0600 | 155.61 | (<0.01) | c1 | 0.0124 | 43.04 | (<0.01) | d1 | 0.0813 | 387.20 | (<0.01) |
| a2 | -0.0133 | -60.43 | (<0.01) | b2 | 0.0651 | 154.87 | (<0.01) | c2 | 0.0095 | 33.00 | (<0.01) | d2 | 0.0226 | 78.44 | (<0.01) |
| a3 | -0.0052 | -23.68 | (<0.01) | b3 | 0.0250 | 58.31 | (<0.01) | c3 | 0.0077 | 26.64 | (<0.01) | d3 | 0.0138 | 47.83 | (<0.01) |
| a4 | -0.0021 | -9.72 | (<0.01) | b4 | 0.0133 | 31.62 | (<0.01) | c4 | 0.0053 | 18.45 | (<0.01) | d4 | 0.0102 | 35.50 | (<0.01) |
| a5 | 0.0021 | 10.30 | (<0.01) | b5 | 0.0102 | 26.60 | (<0.01) | c5 | 0.0114 | 54.43 | (<0.01) | d5 | 0.0083 | 29.00 | (<0.01) |
| C | | -0.0004 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0506 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 7: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Nine week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0309 | 133.71 | (<0.01) | b1 | 0.1307 | 320.63 | (<0.01) | c1 | 0.0068 | 22.68 | (<0.01) | d1 | 0.0174 | 77.41 | (<0.01) |
| a2 | 0.0073 | 31.35 | (<0.01) | b2 | -0.0030 | -6.84 | (<0.01) | c2 | 0.0053 | 17.67 | (<0.01) | d2 | 0.0093 | 31.11 | (<0.01) |
| a3 | 0.0020 | 8.76 | (<0.01) | b3 | -0.0050 | -11.05 | (<0.01) | c3 | 0.0034 | 11.37 | (<0.01) | d3 | 0.0054 | 18.09 | (<0.01) |
| a4 | -0.0016 | -6.79 | (<0.01) | b4 | -0.0061 | -13.75 | (<0.01) | c4 | 0.0035 | 11.75 | (<0.01) | d4 | 0.0044 | 14.76 | (<0.01) |
| a5 | -0.0032 | -14.98 | (<0.01) | b5 | -0.0089 | -21.82 | (<0.01) | c5 | 0.0044 | 19.82 | (<0.01) | d5 | 0.0027 | 9.11 | (<0.01) |
| | C | 0.0004 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0175 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1059 | -784.67 | (<0.01) | b1 | 0.3825 | 1604.19 | (<0.01) | c1 | -0.0015 | -8.56 | (<0.01) | d1 | 0.0138 | 105.23 | (<0.01) |
| a2 | -0.0411 | -300.49 | (<0.01) | b2 | 0.2562 | 993.00 | (<0.01) | c2 | -0.0012 | -6.87 | (<0.01) | d2 | 0.0054 | 30.68 | (<0.01) |
| a3 | -0.0165 | -120.59 | (<0.01) | b3 | 0.1145 | 434.84 | (<0.01) | c3 | -0.0012 | -6.82 | (<0.01) | d3 | 0.0042 | 23.66 | (<0.01) |
| a4 | -0.0026 | -19.20 | (<0.01) | b4 | 0.0722 | 279.87 | (<0.01) | c4 | -0.0008 | -4.43 | (<0.01) | d4 | 0.0032 | 18.45 | (<0.01) |
| a5 | 0.0095 | 75.48 | (<0.01) | b5 | 0.0603 | 253.37 | (<0.01) | c5 | 0.0005 | 4.15 | (<0.01) | d5 | 0.0015 | 8.61 | (<0.01) |
| | C | -0.0012 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6630 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0274 | -159.93 | (<0.01) | b1 | -0.0615 | -203.07 | (<0.01) | c1 | -0.0001 | -0.38 | (<0.01) | d1 | 0.6703 | 4023.73 | (<0.01) |
| a2 | -0.0072 | -41.47 | (<0.01) | b2 | 0.0585 | 178.62 | (<0.01) | c2 | 0.0058 | 26.25 | (<0.01) | d2 | 0.0128 | 57.53 | (<0.01) |
| a3 | -0.0023 | -13.00 | (<0.01) | b3 | 0.0216 | 64.61 | (<0.01) | c3 | 0.0033 | 14.86 | (<0.01) | d3 | 0.0027 | 12.19 | (<0.01) |
| a4 | 0.0006 | 3.62 | (<0.01) | b4 | 0.0140 | 42.63 | (<0.01) | c4 | 0.0029 | 12.95 | (<0.01) | d4 | 0.0027 | 11.89 | (<0.01) |
| a5 | 0.0031 | 19.09 | (<0.01) | b5 | 0.0100 | 33.07 | (<0.01) | c5 | 0.0037 | 22.37 | (<0.01) | d5 | 0.0025 | 11.06 | (<0.01) |
| | C | -0.0004 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4619 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0340 | -149.22 | (<0.01) | b1 | 0.0555 | 137.84 | (<0.01) | c1 | 0.0141 | 47.52 | (<0.01) | d1 | 0.0760 | 343.34 | (<0.01) |
| a2 | -0.0136 | -59.06 | (<0.01) | b2 | 0.0646 | 148.39 | (<0.01) | c2 | 0.0103 | 34.73 | (<0.01) | d2 | 0.0219 | 73.99 | (<0.01) |
| a3 | -0.0061 | -26.27 | (<0.01) | b3 | 0.0254 | 57.05 | (<0.01) | c3 | 0.0087 | 29.32 | (<0.01) | d3 | 0.0152 | 51.11 | (<0.01) |
| a4 | -0.0015 | -6.68 | (<0.01) | b4 | 0.0140 | 32.28 | (<0.01) | c4 | 0.0062 | 20.79 | (<0.01) | d4 | 0.0109 | 36.80 | (<0.01) |
| a5 | 0.0021 | 9.98 | (<0.01) | b5 | 0.0109 | 27.15 | (<0.01) | c5 | 0.0119 | 54.08 | (<0.01) | d5 | 0.0095 | 32.03 | (<0.01) |
| | C | -0.0005 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0486 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 8: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for Ten week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0271 | 116.30 | (<0.01) | b1 | 0.1296 | 314.76 | (<0.01) | c1 | 0.0055 | 18.21 | (<0.01) | d1 | 0.0201 | 89.01 | (<0.01) |
| a2 | 0.0061 | 25.99 | (<0.01) | b2 | -0.0030 | -6.80 | (<0.01) | c2 | 0.0053 | 17.75 | (<0.01) | d2 | 0.0117 | 38.88 | (<0.01) |
| a3 | 0.0001 | 0.61 | (0.54) | b3 | -0.0059 | -12.94 | (<0.01) | c3 | 0.0037 | 12.28 | (<0.01) | d3 | 0.0054 | 18.06 | (<0.01) |
| a4 | -0.0019 | -7.95 | (<0.01) | b4 | -0.0057 | -12.75 | (<0.01) | c4 | 0.0026 | 8.64 | (<0.01) | d4 | 0.0046 | 15.41 | (<0.01) |
| a5 | -0.0038 | -17.65 | (<0.01) | b5 | -0.0075 | -18.17 | (<0.01) | c5 | 0.0041 | 18.21 | (<0.01) | d5 | 0.0035 | 11.73 | (<0.01) |
| | C | 0.0002 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0176 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1079 | -795.36 | (<0.01) | b1 | 0.3813 | 1587.78 | (<0.01) | c1 | -0.0023 | -13.33 | (<0.01) | d1 | 0.0131 | 99.94 | (<0.01) |
| a2 | -0.0428 | -310.80 | (<0.01) | b2 | 0.2565 | 987.00 | (<0.01) | c2 | -0.0021 | -11.97 | (<0.01) | d2 | 0.0056 | 31.78 | (<0.01) |
| a3 | -0.0172 | -124.85 | (<0.01) | b3 | 0.1156 | 435.77 | (<0.01) | c3 | -0.0017 | -9.56 | (<0.01) | d3 | 0.0042 | 23.94 | (<0.01) |
| a4 | -0.0030 | -22.01 | (<0.01) | b4 | 0.0729 | 280.74 | (<0.01) | c4 | -0.0008 | -4.46 | (<0.01) | d4 | 0.0029 | 16.72 | (<0.01) |
| a5 | 0.0094 | 74.27 | (<0.01) | b5 | 0.0608 | 253.58 | (<0.01) | c5 | 0.0003 | 2.06 | (0.04) | d5 | 0.0009 | 5.31 | (<0.01) |
| | C | -0.0008 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6649 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0303 | -174.82 | (<0.01) | b1 | -0.0657 | -213.99 | (<0.01) | c1 | 0.0004 | 1.76 | () | d1 | 0.6665 | 3965.80 | (<0.01) |
| a2 | -0.0088 | -50.19 | (<0.01) | b2 | 0.0585 | 176.17 | (<0.01) | c2 | 0.0085 | 38.08 | (<0.01) | d2 | 0.0151 | 67.55 | (<0.01) |
| a3 | -0.0026 | -14.59 | (<0.01) | b3 | 0.0231 | 68.01 | (<0.01) | c3 | 0.0055 | 24.81 | (<0.01) | d3 | 0.0023 | 10.17 | (<0.01) |
| a4 | 0.0009 | 4.91 | (<0.01) | b4 | 0.0135 | 40.70 | (<0.01) | c4 | 0.0037 | 16.55 | (<0.01) | d4 | 0.0027 | 11.83 | (<0.01) |
| a5 | 0.0032 | 19.64 | (<0.01) | b5 | 0.0110 | 35.79 | (<0.01) | c5 | 0.0058 | 34.89 | (<0.01) | d5 | 0.0026 | 11.77 | (<0.01) |
| | C | -0.0003 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4576 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0353 | -153.59 | (<0.01) | b1 | 0.0492 | 120.99 | (<0.01) | c1 | 0.0169 | 57.03 | (<0.01) | d1 | 0.0753 | 338.41 | (<0.01) |
| a2 | -0.0146 | -62.85 | (<0.01) | b2 | 0.0631 | 143.56 | (<0.01) | c2 | 0.0140 | 47.35 | (<0.01) | d2 | 0.0257 | 86.48 | (<0.01) |
| a3 | -0.0067 | -28.61 | (<0.01) | b3 | 0.0226 | 50.43 | (<0.01) | c3 | 0.0101 | 34.08 | (<0.01) | d3 | 0.0185 | 62.26 | (<0.01) |
| a4 | -0.0020 | -8.90 | (<0.01) | b4 | 0.0140 | 31.97 | (<0.01) | c4 | 0.0093 | 31.32 | (<0.01) | d4 | 0.0147 | 49.67 | (<0.01) |
| a5 | 0.0025 | 11.49 | (<0.01) | b5 | 0.0099 | 24.41 | (<0.01) | c5 | 0.0154 | 69.50 | (<0.01) | d5 | 0.0142 | 47.94 | (<0.01) |
| | C | -0.0001 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0487 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 9: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for 11 week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0308 | 127.37 | (<0.01) | b1 | 0.1329 | 316.19 | (<0.01) | c1 | 0.0099 | 34.69 | (<0.01) | d1 | 0.0288 | 123.23 | (<0.01) |
| a2 | 0.0082 | 33.33 | (<0.01) | b2 | -0.0076 | -16.67 | (<0.01) | c2 | 0.0067 | 23.30 | (<0.01) | d2 | 0.0135 | 46.80 | (<0.01) |
| a3 | 0.0014 | 5.55 | (<0.01) | b3 | -0.0069 | -14.90 | (<0.01) | c3 | 0.0035 | 12.11 | (<0.01) | d3 | 0.0074 | 25.62 | (<0.01) |
| a4 | -0.0014 | -5.85 | (<0.01) | b4 | -0.0065 | -14.37 | (<0.01) | c4 | 0.0027 | 9.37 | (<0.01) | d4 | 0.0058 | 20.14 | (<0.01) |
| a5 | -0.0038 | -16.67 | (<0.01) | b5 | -0.0092 | -21.88 | (<0.01) | c5 | 0.0045 | 19.14 | (<0.01) | d5 | 0.0047 | 16.46 | (<0.01) |
| | C | 0.0000 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0188 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1095 | -762.75 | (<0.01) | b1 | 0.3803 | 1526.59 | (<0.01) | c1 | -0.0027 | -16.10 | (<0.01) | d1 | 0.0109 | 78.55 | (<0.01) |
| a2 | -0.0437 | -300.14 | (<0.01) | b2 | 0.2585 | 960.50 | (<0.01) | c2 | -0.0029 | -16.96 | (<0.01) | d2 | 0.0038 | 22.57 | (<0.01) |
| a3 | -0.0176 | -120.99 | (<0.01) | b3 | 0.1150 | 418.56 | (<0.01) | c3 | -0.0023 | -13.70 | (<0.01) | d3 | 0.0035 | 20.40 | (<0.01) |
| a4 | -0.0035 | -24.14 | (<0.01) | b4 | 0.0716 | 266.30 | (<0.01) | c4 | -0.0015 | -9.05 | (<0.01) | d4 | 0.0025 | 14.65 | (<0.01) |
| a5 | 0.0094 | 70.10 | (<0.01) | b5 | 0.0591 | 237.81 | (<0.01) | c5 | -0.0009 | -6.82 | (<0.01) | d5 | 0.0006 | 3.55 | (<0.01) |
| | C | -0.0009 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6547 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0323 | -164.30 | (<0.01) | b1 | -0.0453 | -133.05 | (<0.01) | c1 | 0.0016 | 6.70 | (<0.01) | d1 | 0.5868 | 3101.99 | (<0.01) |
| a2 | -0.0082 | -41.26 | (<0.01) | b2 | 0.0560 | 152.12 | (<0.01) | c2 | 0.0055 | 23.78 | (<0.01) | d2 | 0.0127 | 54.42 | (<0.01) |
| a3 | -0.0027 | -13.50 | (<0.01) | b3 | 0.0219 | 58.28 | (<0.01) | c3 | 0.0027 | 11.69 | (<0.01) | d3 | 0.0035 | 15.21 | (<0.01) |
| a4 | 0.0008 | 4.02 | (<0.01) | b4 | 0.0132 | 35.88 | (<0.01) | c4 | 0.0022 | 9.51 | (<0.01) | d4 | 0.0038 | 16.11 | (<0.01) |
| a5 | 0.0034 | 18.31 | (<0.01) | b5 | 0.0100 | 29.48 | (<0.01) | c5 | 0.0033 | 17.51 | (<0.01) | d5 | 0.0015 | 6.45 | (<0.01) |
| | C | 0.0001 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.3603 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0347 | -144.42 | (<0.01) | b1 | 0.0463 | 110.99 | (<0.01) | c1 | 0.0128 | 44.92 | (<0.01) | d1 | 0.0659 | 284.56 | (<0.01) |
| a2 | -0.0129 | -52.95 | (<0.01) | b2 | 0.0673 | 149.34 | (<0.01) | c2 | 0.0092 | 32.25 | (<0.01) | d2 | 0.0191 | 67.02 | (<0.01) |
| a3 | -0.0064 | -26.36 | (<0.01) | b3 | 0.0240 | 52.03 | (<0.01) | c3 | 0.0058 | 20.37 | (<0.01) | d3 | 0.0132 | 46.18 | (<0.01) |
| a4 | -0.0017 | -7.08 | (<0.01) | b4 | 0.0138 | 30.54 | (<0.01) | c4 | 0.0051 | 17.93 | (<0.01) | d4 | 0.0112 | 39.13 | (<0.01) |
| a5 | 0.0021 | 9.52 | (<0.01) | b5 | 0.0103 | 24.66 | (<0.01) | c5 | 0.0085 | 36.80 | (<0.01) | d5 | 0.0084 | 29.55 | (<0.01) |
| | C | -0.0008 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0399 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 10: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for 13 week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0333 | 136.44 | (<0.01) | b1 | 0.1307 | 303.92 | (<0.01) | c1 | 0.0097 | 30.60 | (<0.01) | d1 | 0.0196 | 82.47 | (<0.01) |
| a2 | 0.0070 | 28.16 | (<0.01) | b2 | -0.0055 | -11.77 | (<0.01) | c2 | 0.0080 | 25.21 | (<0.01) | d2 | 0.0099 | 30.93 | (<0.01) |
| a3 | 0.0010 | 3.88 | (<0.01) | b3 | -0.0066 | -13.88 | (<0.01) | c3 | 0.0048 | 15.09 | (<0.01) | d3 | 0.0050 | 15.64 | (<0.01) |
| a4 | -0.0005 | -2.16 | (<0.01) | b4 | -0.0064 | -13.80 | (<0.01) | c4 | 0.0043 | 13.37 | (<0.01) | d4 | 0.0047 | 14.69 | (<0.01) |
| a5 | -0.0038 | -16.52 | (<0.01) | b5 | -0.0083 | -19.43 | (<0.01) | c5 | 0.0056 | 23.46 | (<0.01) | d5 | 0.0036 | 11.22 | (<0.01) |
| | C | -0.0002 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0178 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1038 | -725.09 | (<0.01) | b1 | 0.3865 | 1533.42 | (<0.01) | c1 | -0.0028 | -15.03 | (<0.01) | d1 | 0.0136 | 97.39 | (<0.01) |
| a2 | -0.0402 | -276.84 | (<0.01) | b2 | 0.2570 | 940.87 | (<0.01) | c2 | -0.0020 | -10.69 | (<0.01) | d2 | 0.0058 | 31.16 | (<0.01) |
| a3 | -0.0155 | -107.22 | (<0.01) | b3 | 0.1128 | 404.67 | (<0.01) | c3 | -0.0019 | -10.32 | (<0.01) | d3 | 0.0045 | 23.93 | (<0.01) |
| a4 | -0.0025 | -17.22 | (<0.01) | b4 | 0.0707 | 259.00 | (<0.01) | c4 | -0.0010 | -5.50 | (<0.01) | d4 | 0.0033 | 17.51 | (<0.01) |
| a5 | 0.0094 | 69.96 | (<0.01) | b5 | 0.0588 | 233.73 | (<0.01) | c5 | 0.0003 | 2.45 | (<0.01) | d5 | 0.0012 | 6.61 | (<0.01) |
| | C | -0.0004 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6618 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0264 | -146.05 | (<0.01) | b1 | -0.0619 | -194.22 | (<0.01) | c1 | -0.0025 | -10.46 | (<0.01) | d1 | 0.6721 | 3814.64 | (<0.01) |
| a2 | -0.0076 | -41.61 | (<0.01) | b2 | 0.0567 | 164.03 | (<0.01) | c2 | 0.0060 | 25.42 | (<0.01) | d2 | 0.0153 | 64.82 | (<0.01) |
| a3 | -0.0020 | -10.80 | (<0.01) | b3 | 0.0228 | 64.71 | (<0.01) | c3 | 0.0041 | 17.23 | (<0.01) | d3 | 0.0026 | 10.83 | (<0.01) |
| a4 | 0.0005 | 2.85 | (<0.01) | b4 | 0.0131 | 38.00 | (<0.01) | c4 | 0.0038 | 15.98 | (<0.01) | d4 | 0.0018 | 7.81 | (<0.01) |
| a5 | 0.0031 | 18.25 | (<0.01) | b5 | 0.0111 | 35.03 | (<0.01) | c5 | 0.0040 | 22.56 | (<0.01) | d5 | -0.0001 | -0.53 | (<0.01) |
| | C | 0.0000 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4630 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0312 | -129.39 | (<0.01) | b1 | 0.0562 | 132.25 | (<0.01) | c1 | 0.0111 | 35.41 | (<0.01) | d1 | 0.0761 | 324.06 | (<0.01) |
| a2 | -0.0126 | -51.37 | (<0.01) | b2 | 0.0645 | 140.04 | (<0.01) | c2 | 0.0087 | 27.59 | (<0.01) | d2 | 0.0214 | 67.99 | (<0.01) |
| a3 | -0.0052 | -21.46 | (<0.01) | b3 | 0.0232 | 49.33 | (<0.01) | c3 | 0.0064 | 20.43 | (<0.01) | d3 | 0.0156 | 49.36 | (<0.01) |
| a4 | -0.0018 | -7.51 | (<0.01) | b4 | 0.0138 | 30.08 | (<0.01) | c4 | 0.0057 | 18.18 | (<0.01) | d4 | 0.0115 | 36.60 | (<0.01) |
| a5 | 0.0021 | 9.31 | (<0.01) | b5 | 0.010149 | 23.95 | (<0.01) | c5 | 0.0093 | 39.55 | (<0.01) | d5 | 0.0088 | 27.93 | (<0.01) |
| | C | 0.0004 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0458 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 11: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for 14 week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0244 | 103.41 | (<0.01) | b1 | 0.1484 | 362.37 | (<0.01) | c1 | 0.0088 | 29.96 | (<0.01) | d1 | 0.0223 | 97.79 | (<0.01) |
| a2 | 0.0087 | 36.44 | (<0.01) | b2 | -0.0050 | -11.25 | (<0.01) | c2 | 0.0069 | 23.56 | (<0.01) | d2 | 0.0147 | 50.22 | (<0.01) |
| a3 | 0.0020 | 8.24 | (<0.01) | b3 | -0.0094 | -20.60 | (<0.01) | c3 | 0.0053 | 18.05 | (<0.01) | d3 | 0.0083 | 28.44 | (<0.01) |
| a4 | -0.0009 | -4.00 | (<0.01) | b4 | -0.0102 | -23.04 | (<0.01) | c4 | 0.0042 | 14.25 | (<0.01) | d4 | 0.0063 | 21.44 | (<0.01) |
| a5 | -0.0025 | -11.30 | (<0.01) | b5 | -0.0133 | -32.55 | (<0.01) | c5 | 0.0066 | 28.86 | (<0.01) | d5 | 0.0050 | 16.97 | (<0.01) |
| C | | 0.0005 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0214 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.1049 | -744.56 | (<0.01) | b1 | 0.3914 | 1602.66 | (<0.01) | c1 | -0.0034 | -19.65 | (<0.01) | d1 | 0.0137 | 100.72 | (<0.01) |
| a2 | -0.0396 | -277.17 | (<0.01) | b2 | 0.2550 | 960.12 | (<0.01) | c2 | -0.0027 | -15.32 | (<0.01) | d2 | 0.0052 | 30.00 | (<0.01) |
| a3 | -0.0144 | -101.06 | (<0.01) | b3 | 0.1109 | 409.24 | (<0.01) | c3 | -0.0014 | -8.16 | (<0.01) | d3 | 0.0042 | 24.05 | (<0.01) |
| a4 | -0.0015 | -10.50 | (<0.01) | b4 | 0.0675 | 254.19 | (<0.01) | c4 | -0.0015 | -8.51 | (<0.01) | d4 | 0.0025 | 14.11 | (<0.01) |
| a5 | 0.0098 | 75.04 | (<0.01) | b5 | 0.0576 | 236.16 | (<0.01) | c5 | -0.0003 | -2.55 | (<0.01) | d5 | 0.0009 | 5.07 | (<0.01) |
| C | | -0.0010 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.6511 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0278 | -151.63 | (<0.01) | b1 | -0.0528 | -166.23 | | c1 | -0.0006 | -2.64 | | d1 | 0.6336 | 3583.30() | |
| a2 | -0.0079 | -42.51 | | b2 | 0.0563 | 163.09 | (<0.01) | c2 | 0.0063 | 27.71 | (<0.01) | d2 | 0.0130 | 57.27 | (<0.01) |
| a3 | -0.0018 | -9.46 | (<0.01) | b3 | 0.0228 | 64.53 | (<0.01) | c3 | 0.0049 | 21.53 | (<0.01) | d3 | 0.0018 | 8.10 | (<0.01) |
| a4 | 0.0010 | 5.63 | (<0.01) | b4 | 0.0125 | 36.26 | (<0.01) | c4 | 0.0023 | 9.95 | (<0.01) | d4 | 0.0004 | 1.92 | (0.60) |
| a5 | 0.0028 | 16.56 | (<0.01) | b5 | 0.0095 | 29.88 | (<0.01) | c5 | 0.0035 | 19.92 | (<0.01) | d5 | 0.0022 | 9.81 | (<0.01) |
| C | | -0.0004 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.4147 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0328 | -140.18 | (<0.01) | b1 | 0.0531 | 130.84 | (<0.01) | c1 | 0.0132 | 45.40 | (<0.01) | d1 | 0.0791 | 350.08 | (<0.01) |
| a2 | -0.0129 | -54.43 | (<0.01) | b2 | 0.0631 | 142.78 | (<0.01) | c2 | 0.0079 | 27.18 | (<0.01) | d2 | 0.0195 | 66.92 | (<0.01) |
| a3 | -0.0054 | -22.64 | (<0.01) | b3 | 0.0236 | 52.30 | (<0.01) | c3 | 0.0080 | 27.49 | (<0.01) | d3 | 0.0150 | 51.50 | (<0.01) |
| a4 | -0.0015 | -6.55 | (<0.01) | b4 | 0.0128 | 29.09 | (<0.01) | c4 | 0.0044 | 15.25 | (<0.01) | d4 | 0.0097 | 33.31 | (<0.01) |
| a5 | 0.0016 | 7.50 | (<0.01) | b5 | 0.0090 | 22.14 | (<0.01) | c5 | 0.0099 | 43.72 | (<0.01) | d5 | 0.0083 | 28.72 | (<0.01) |
| C | | -0.0004 | | | | | | | | | | | | | |
| Adj. $R^2$ | | 0.0436 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 12: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for 15 week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0238 | 58.87 | (<0.01) | b1 | 0.1840 | 276.25 | (<0.01) | c1 | 0.0021 | 4.07 | (<0.01) | d1 | 0.0117 | 29.02 | (<0.01) |
| a2 | 0.0080 | 19.65 | (<0.01) | b2 | -0.0459 | -62.00 | (<0.01) | c2 | 0.0040 | 7.69 | (<0.01) | d2 | 0.0111 | 21.07 | (<0.01) |
| a3 | 0.0031 | 7.67 | (<0.01) | b3 | -0.0150 | -19.88 | (<0.01) | c3 | 0.0038 | 7.32 | (<0.01) | d3 | 0.0053 | 10.03 | (<0.01) |
| a4 | -0.0026 | -6.48 | (<0.01) | b4 | -0.0096 | -12.98 | (<0.01) | c4 | 0.0023 | 4.46 | (<0.01) | d4 | 0.0038 | 7.27 | (<0.01) |
| a5 | -0.0011 | -2.79 | (<0.01) | b5 | -0.0143 | -21.42 | (<0.01) | c5 | 0.0041 | 10.14 | (<0.01) | d5 | 0.0041 | 7.75 | (<0.01) |
| | C | -0.0001 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0203 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0660 | -258.16 | (<0.01) | b1 | 0.4408 | 1044.97 | (<0.01) | c1 | 0.0014 | 4.27 | (<0.01) | d1 | 0.0139 | 54.61 | (<0.01) |
| a2 | -0.0170 | -66.01 | (<0.01) | b2 | 0.2341 | 499.48 | (<0.01) | c2 | -0.0007 | -2.18 | (<0.01) | d2 | 0.0001 | 0.29 | (<0.01) |
| a3 | -0.0042 | -16.41 | (<0.01) | b3 | 0.0856 | 179.74 | (<0.01) | c3 | -0.0021 | -6.41 | (<0.01) | d3 | 0.0014 | 4.16 | (<0.01) |
| a4 | 0.0010 | 3.95 | (<0.01) | b4 | 0.0500 | 106.73 | (<0.01) | c4 | -0.0015 | -4.56 | (<0.01) | d4 | 0.0013 | 3.84 | (<0.01) |
| a5 | 0.0094 | 38.69 | (<0.01) | b5 | 0.0492 | 116.42 | (<0.01) | c5 | -0.0010 | -3.96 | (<0.01) | d5 | 0.0000 | 0.04 | (<0.01) |
| | C | -0.0010 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.6067 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0178 | -57.97 | (<0.01) | b1 | -0.0471 | -93.24 | (<0.01) | c1 | 0.0021 | 5.39 | (<0.01) | d1 | 0.6511 | 2132.39 | (<0.01) |
| a2 | -0.0025 | -8.22 | (<0.01) | b2 | 0.0527 | 93.84 | (<0.01) | c2 | 0.0082 | 20.64 | (<0.01) | d2 | 0.0144 | 36.13 | (<0.01) |
| a3 | 0.0000 | 0.16 | (<0.01) | b3 | 0.0172 | 30.05 | (<0.01) | c3 | 0.0051 | 12.71 | (<0.01) | d3 | -0.0010 | -2.40 | (<0.01) |
| a4 | 0.0014 | 4.71 | (<0.01) | b4 | 0.0104 | 18.55 | (<0.01) | c4 | 0.0033 | 8.35 | (<0.01) | d4 | -0.0016 | -3.99 | (<0.01) |
| a5 | 0.0025 | 8.60 | (<0.01) | b5 | 0.0090 | 17.82 | (<0.01) | c5 | 0.0037 | 12.23 | (<0.01) | d5 | 0.0008 | 1.88 | (<0.01) |
| | C | -0.0006 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4367 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0212 | -53.35 | (<0.01) | b1 | 0.0796 | 121.23 | (<0.01) | c1 | 0.0150 | 28.97 | (<0.01) | d1 | 0.0889 | 224.21 | (<0.01) |
| a2 | -0.0056 | -13.88 | (<0.01) | b2 | 0.0585 | 80.20 | (<0.01) | c2 | 0.0118 | 22.78 | (<0.01) | d2 | 0.0185 | 35.66 | (<0.01) |
| a3 | -0.0025 | -6.33 | (<0.01) | b3 | 0.0175 | 23.55 | (<0.01) | c3 | 0.0072 | 13.95 | (<0.01) | d3 | 0.0114 | 22.04 | (<0.01) |
| a4 | -0.0006 | -1.50 | (<0.01) | b4 | 0.0092 | 12.67 | (<0.01) | c4 | 0.0060 | 11.53 | (<0.01) | d4 | 0.0093 | 17.92 | (<0.01) |
| a5 | 0.0015 | 4.07 | (<0.01) | b5 | 0.0062 | 9.48 | (<0.01) | c5 | 0.0097 | 24.42 | (<0.01) | d5 | 0.0079 | 15.30 | (<0.01) |
| | C | 0.0004 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0507 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 13: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for 16 week to maturity**

*Panel A: Quote revision equation*

$\widetilde{QR}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{qr}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0057 | -3.99 | (<0.01) | b1 | 0.1863 | 90.933 | (<0.01) | c1 | -0.0086 | -4.415 | (<0.01) | d1 | 0.0023 | 1.457 | (<0.01) |
| a2 | -0.0055 | -3.84 | (<0.01) | b2 | -0.0773 | -35.066 | (<0.01) | c2 | -0.0036 | -1.850 | (<0.01) | d2 | 0.0077 | 3.920 | (<0.01) |
| a3 | -0.0024 | -1.64 | (<0.01) | b3 | -0.0165 | -7.396 | (<0.01) | c3 | 0.0002 | 0.095 | (0.34) | d3 | 0.0032 | 1.644 | (<0.01) |
| a4 | 0.0140 | 9.77 | (<0.01) | b4 | -0.0130 | -5.894 | (<0.01) | c4 | 0.0030 | 1.525 | (<0.01) | d4 | -0.0005 | -0.259 | (<0.01) |
| a5 | -0.0080 | -5.69 | (<0.01) | b5 | -0.0144 | -6.964 | (<0.01) | c5 | 0.0001 | 0.083 | (<0.01) | d5 | -0.0033 | -1.677 | (<0.01) |
| | C | 0.0025 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0206 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$\widetilde{TD}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{td}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0113 | -10.270 | (<0.01) | b1 | 0.3496 | 223.990 | (<0.01) | c1 | 0.0193 | 12.948 | (<0.01) | d1 | 0.0191 | 16.209 | (<0.01) |
| a2 | -0.0049 | -4.431 | (<0.01) | b2 | 0.2203 | 131.176 | (<0.01) | c2 | 0.0058 | 3.912 | (<0.01) | d2 | -0.0094 | -6.317 | (<0.01) |
| a3 | -0.0012 | -1.068 | (<0.01) | b3 | 0.0931 | 54.659 | (<0.01) | c3 | 0.0028 | 1.875 | (0.11) | d3 | -0.0018 | -1.199 | (0.27) |
| a4 | -0.0042 | -3.843 | (<0.01) | b4 | 0.0538 | 31.975 | (<0.01) | c4 | 0.0022 | 1.489 | (0.14) | d4 | 0.0011 | 0.745 | (0.46) |
| a5 | 0.0053 | 4.900 | (<0.01) | b5 | 0.0543 | 34.579 | (<0.01) | c5 | 0.0035 | 2.999 | (<0.01) | d5 | -0.0003 | -0.211 | (<0.01) |
| | C | 0.0042 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4318 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0020 | -1.781 | (<0.01) | b1 | -0.0622 | -39.156 | (<0.01) | c1 | 0.0098 | 6.451 | (<0.01) | d1 | 0.6239 | 520.205 | (<0.01) |
| a2 | 0.0009 | 0.827 | (<0.01) | b2 | 0.0543 | 31.759 | (<0.01) | c2 | 0.0062 | 4.066 | (<0.01) | d2 | 0.0129 | 8.500 | (<0.01) |
| a3 | 0.0010 | 0.918 | (0.36) | b3 | 0.0210 | 12.099 | (<0.01) | c3 | 0.0102 | 6.719 | (<0.01) | d3 | 0.0085 | 5.586 | (<0.01) |
| a4 | -0.0015 | -1.313 | (0.19) | b4 | 0.0142 | 8.308 | (<0.01) | c4 | 0.0085 | 5.588 | (<0.01) | d4 | 0.0029 | 1.929 | (0.05) |
| a5 | 0.0010 | 0.887 | (0.38) | b5 | 0.0139 | 8.675 | (<0.01) | c5 | 0.0094 | 7.871 | (<0.01) | d5 | 0.0015 | 0.995 | (0.37) |
| | C | -0.0013 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.4105 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$\widetilde{TV}_t = \sum_{i=1}^{k} a_i\widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i\widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i\widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i\widetilde{TV}_{t-i} + \varepsilon_t^{tv}$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0079 | -5.676 | (<0.01) | b1 | 0.0968 | 49.026 | (<0.01) | c1 | 0.0289 | 15.319 | (<0.01) | d1 | 0.1150 | 77.137 | (<0.01) |
| a2 | -0.0012 | -0.853 | (0.29) | b2 | 0.0680 | 31.967 | (<0.01) | c2 | 0.0172 | 9.144 | (<0.01) | d2 | 0.0250 | 13.245 | (<0.01) |
| a3 | -0.0016 | -1.142 | (0.29) | b3 | 0.0234 | 10.868 | (<0.01) | c3 | 0.0124 | 6.556 | (<0.01) | d3 | 0.0192 | 10.182 | (<0.01) |
| a4 | -0.0026 | -1.889 | (0.29) | b4 | 0.0115 | 5.389 | (<0.01) | c4 | 0.0055 | 2.912 | (<0.01) | d4 | 0.0196 | 10.398 | (<0.01) |
| a5 | 0.0010 | 0.722 | (0.29) | b5 | 0.0089 | 4.457 | (<0.01) | c5 | 0.0142 | 9.556 | (<0.01) | d5 | 0.0183 | 9.672 | (<0.01) |
| | C | -0.0011 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0898 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

**Table 14: Coefficient estimates of the vector autoregressive(VAR) model for E-Mini S&P500 for 17 week to maturity**

*Panel A: Quote revision equation*

$$\widetilde{QR}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{qr}$$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0079 | 1.82 | (0.07) | b1 | 0.1416 | 29.59 | (<0.01) | c1 | 0.0011 | 0.22 | (0.83) | d1 | 0.0016 | 0.34 | (0.73) |
| a2 | -0.0060 | -1.37 | (0.17) | b2 | -0.0369 | -7.63 | (<0.01) | c2 | 0.0039 | 0.78 | (0.44) | d2 | 0.0014 | 0.27 | (0.78) |
| a3 | 0.0027 | 0.63 | (0.53) | b3 | -0.0143 | -2.94 | (<0.01) | c3 | 0.0031 | 0.63 | (0.53) | d3 | 0.0028 | 0.55 | (0.58) |
| a4 | 0.0196 | 4.50 | (<0.01) | b4 | -0.0065 | -1.34 | (0.18) | c4 | -0.0003 | -0.06 | (0.95) | d4 | 0.0003 | 0.06 | (0.95) |
| a5 | -0.0033 | -0.77 | (0.44) | b5 | -0.0032 | -0.66 | (0.53) | c5 | 0.0026 | 0.57 | (0.58) | d5 | -0.0073 | -1.46 | (0.14) |
| | C | 0.0035 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0204 | | | | | | | | | | | | | |

*Panel B: Trade direction equation*

$$\widetilde{TD}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{td}$$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0203 | 4.72 | (<0.01) | b1 | 0.0395 | 8.36 | (<0.01) | c1 | 0.0116 | 2.37 | (<0.01) | d1 | 0.0327 | 7.19 | (<0.01) |
| a2 | 0.0131 | 3.05 | (<0.01) | b2 | 0.1209 | 25.25 | (<0.01) | c2 | -0.0018 | -0.36 | (0.71) | d2 | 0.0023 | 0.46 | (0.71) |
| a3 | 0.0108 | 2.51 | (<0.01) | b3 | 0.0795 | 16.54 | (<0.01) | c3 | 0.0061 | 1.25 | (0.21) | d3 | -0.0068 | -1.37 | (0.14) |
| a4 | 0.0084 | 1.96 | (<0.01) | b4 | 0.0568 | 11.87 | (<0.01) | c4 | -0.0035 | -0.72 | (0.47) | d4 | -0.0083 | -1.68 | (0.09) |
| a5 | 0.0067 | 1.57 | (0.11) | b5 | 0.0468 | 9.80 | (<0.01) | c5 | -0.0016 | -0.35 | (<0.01) | d5 | 0.0058 | 1.19 | (0.23) |
| | C | -0.0115 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0431 | | | | | | | | | | | | | |

*Panel C: Hidden order equation*

$$\widetilde{HDV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{hdv}$$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | 0.0075 | 1.86 | (0.06) | b1 | -0.0779 | -17.50 | (<0.01) | c1 | -0.0104 | -2.25 | (<0.01) | d1 | 0.4023 | 93.95 | (<0.01) |
| a2 | 0.0038 | 0.94 | (0.34) | b2 | 0.0201 | 4.46 | (<0.01) | c2 | 0.0158 | 3.43 | (<0.01) | d2 | 0.0221 | 4.75 | (<0.01) |
| a3 | 0.0038 | 0.95 | (0.34) | b3 | 0.0155 | 3.43 | (<0.01) | c3 | 0.0001 | 0.02 | (0.98) | d3 | 0.0009 | 0.19 | (0.84) |
| a4 | 0.0017 | 0.43 | (0.73) | b4 | 0.0181 | 4.02 | (<0.01) | c4 | -0.0148 | -3.21 | (¡0.01) | d4 | -0.0127 | -2.72 | (<0.01) |
| a5 | 0.0034 | 0.86 | (0.39) | b5 | 0.0017 | 0.39 | (0.78) | c5 | 0.0096 | 2.26 | (¡0.01) | d5 | 0.0167 | 3.59 | (<0.01) |
| | C | 0.0022 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.1534 | | | | | | | | | | | | | |

*Panel D: Trade volume equation*

$$\widetilde{TV}_t = \sum_{i=1}^{k} a_i \widetilde{QR}_{t-i} + \sum_{i=1}^{k} b_i \widetilde{TD}_{t-i} + \sum_{i=1}^{k} c_i \widetilde{HDV}_{t-i} + \sum_{i=1}^{k} d_i \widetilde{TV}_{t-i} + \varepsilon_t^{tv}$$

| | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value | | Coeff. | t-stat | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a1 | -0.0013 | -0.30 | (0.76) | b1 | 0.0166 | 3.47 | (<0.01) | c1 | 0.0232 | 4.68 | (<0.01) | d1 | 0.0970 | 21.09 | (<0.01) |
| a2 | -0.0012 | -0.28 | (0.77) | b2 | 0.0372 | 7.68 | (<0.01) | c2 | 0.0118 | 2.39 | (<0.01) | d2 | 0.0344 | 6.89 | (<0.01) |
| a3 | -0.0025 | -0.57 | (0.79) | b3 | 0.0244 | 5.02 | (<0.01) | c3 | 0.0084 | 1.70 | (0.09) | d3 | 0.0103 | 2.06 | (<0.01) |
| a4 | -0.0005 | -0.12 | (0.94) | b4 | 0.0189 | 3.90 | (<0.01) | c4 | 0.0047 | 0.94 | (0.37) | d4 | -0.0133 | -2.67 | (<0.01) |
| a5 | 0.0063 | 1.47 | (0.14) | b5 | 0.0073 | 1.51 | (0.13) | c5 | 0.0157 | 3.44 | (<0.01) | d5 | 0.0119 | 2.38 | (<0.01) |
| | C | 0.0078 | | | | | | | | | | | | | |
| | Adj. $R^2$ | 0.0233 | | | | | | | | | | | | | |

Note: This table shows the results of our vector autoregressive(VAR) model, where $\widetilde{QT}$ is quote return, $\widetilde{TD}$ is signed trade direction, $\widetilde{HDV}$ is signed hidden order volume and $\widetilde{TV}$ is signed trading volume at time $t$.

## .2 Impulse response analysis on the term-structure of E-mini S&P500 (Supplement for chapter 7)

This section provides a supplement information of the impulse response function analysis using E-Mini S&P 500 data. We analyze the IRF week-by-week as a term-structure analysis, from eighteen to one week to maturity. Here we show the coefficients for all periods, however, the result for week eighteen, twelve, six and one is showed in chapter 7.

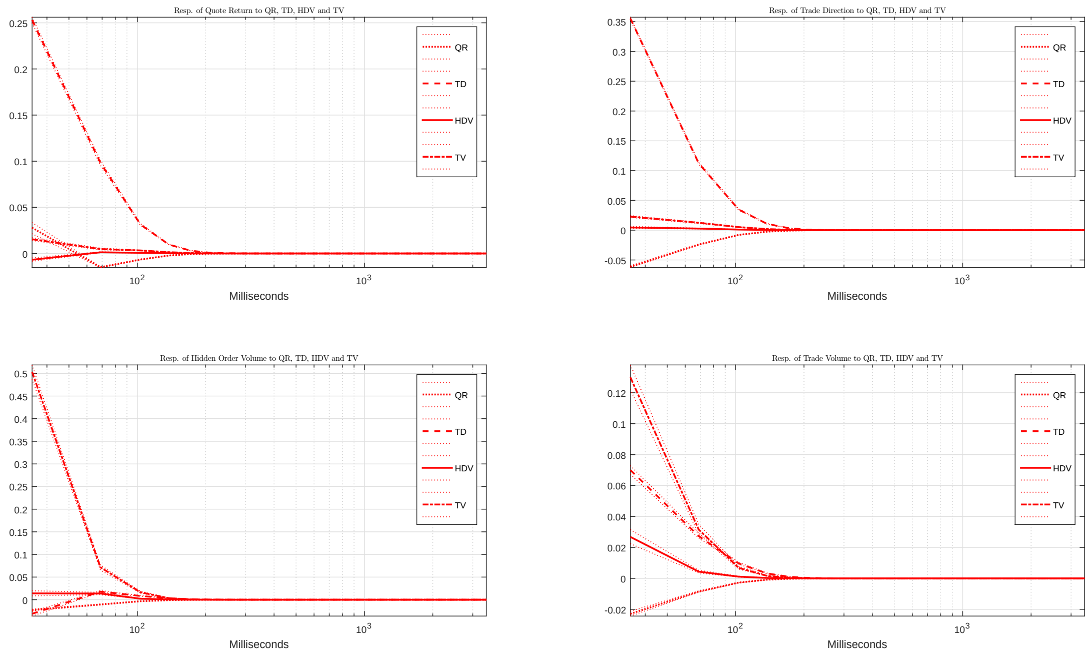



**Figure 1: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for one week to maturity.**

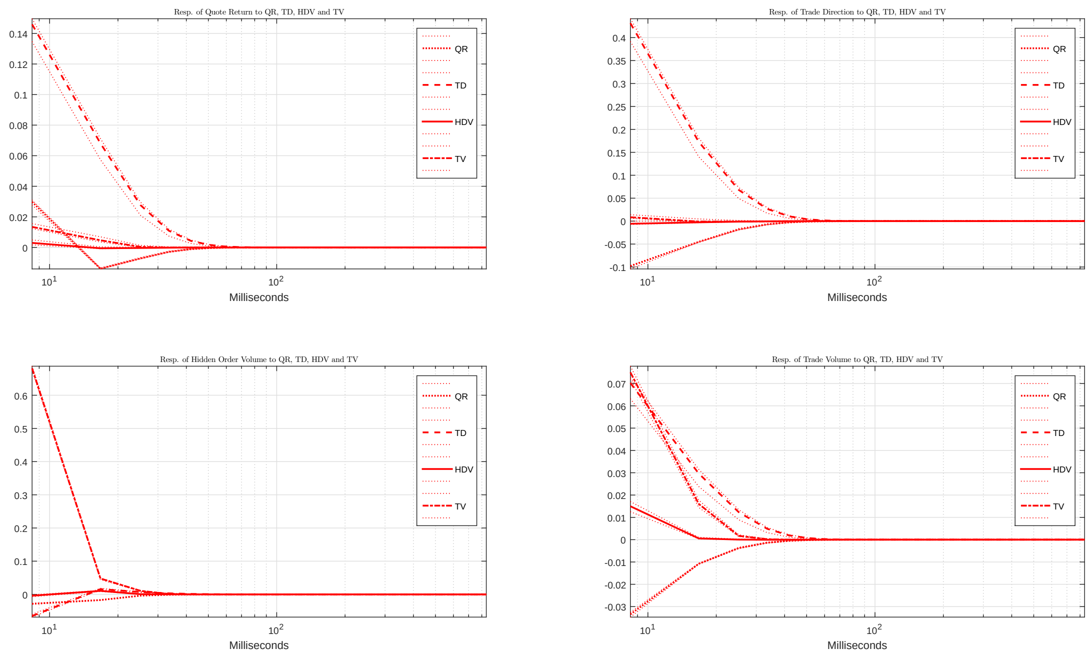



**Figure 2: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for two week to maturity.**

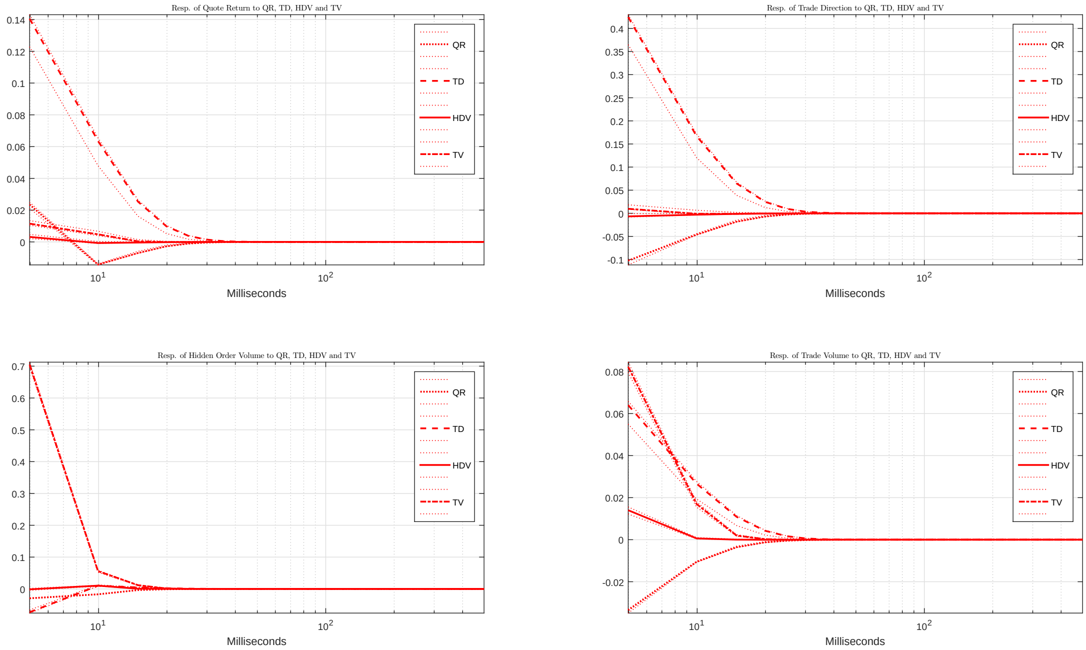



**Figure 3: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for three week to maturity.**

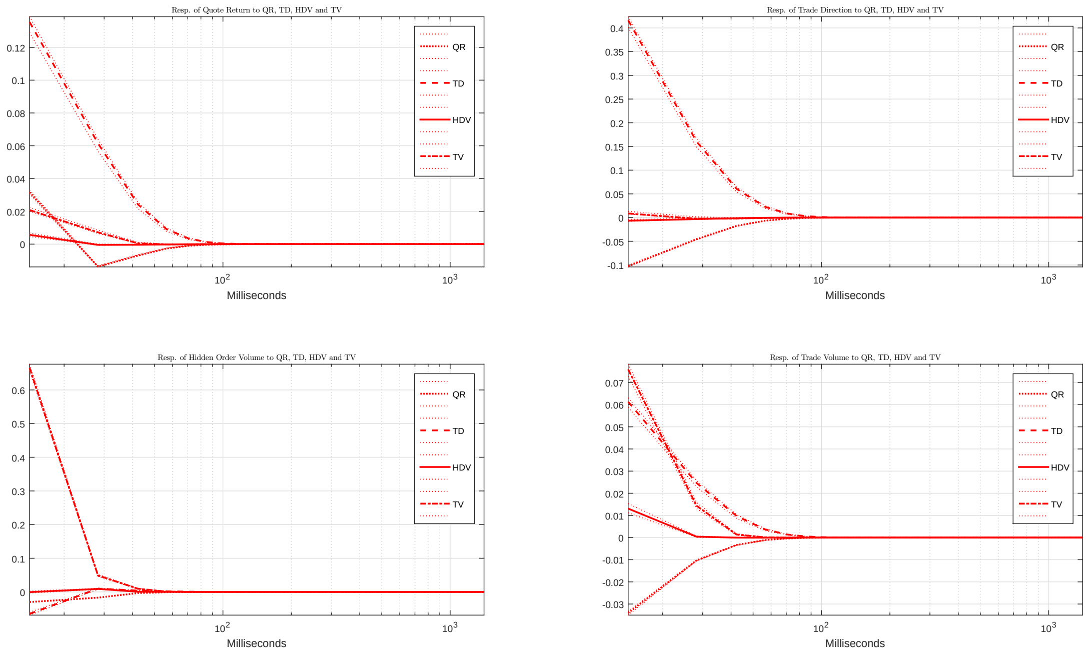



**Figure 4: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for four week to maturity.**

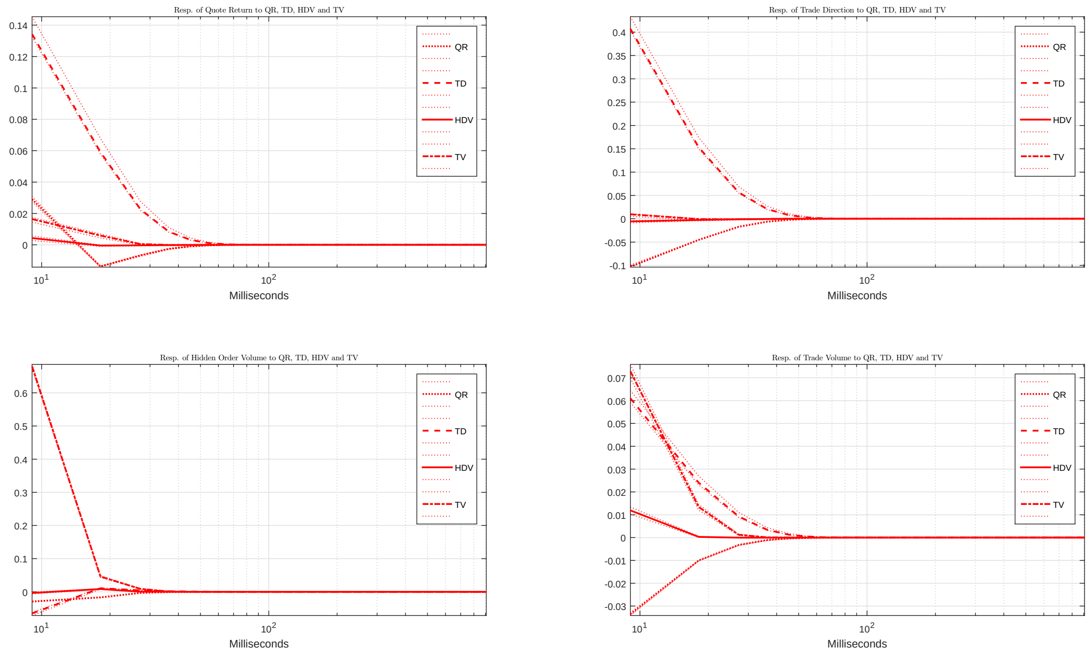



**Figure 5: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for seven week to maturity.**

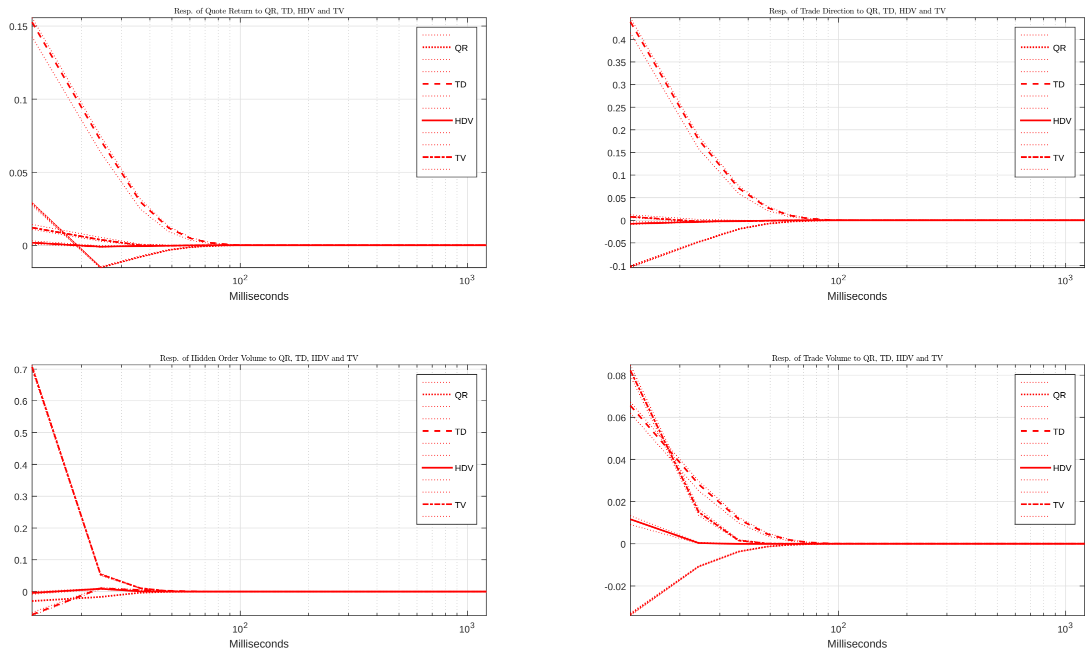



**Figure 6: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for eight week to maturity.**

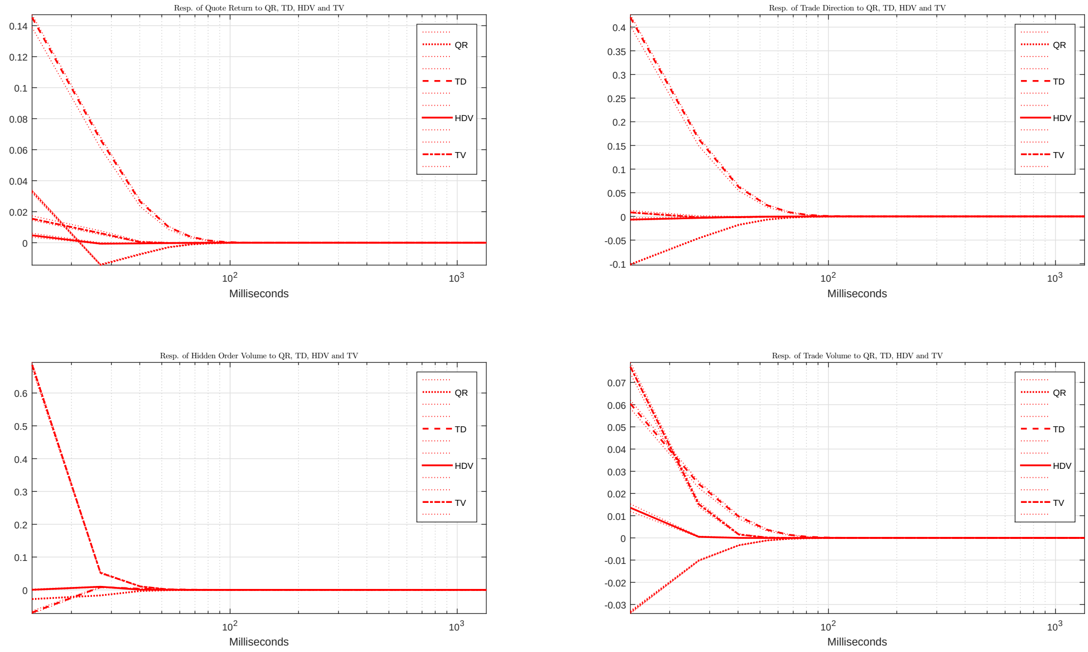



**Figure 7: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for nine week to maturity.**

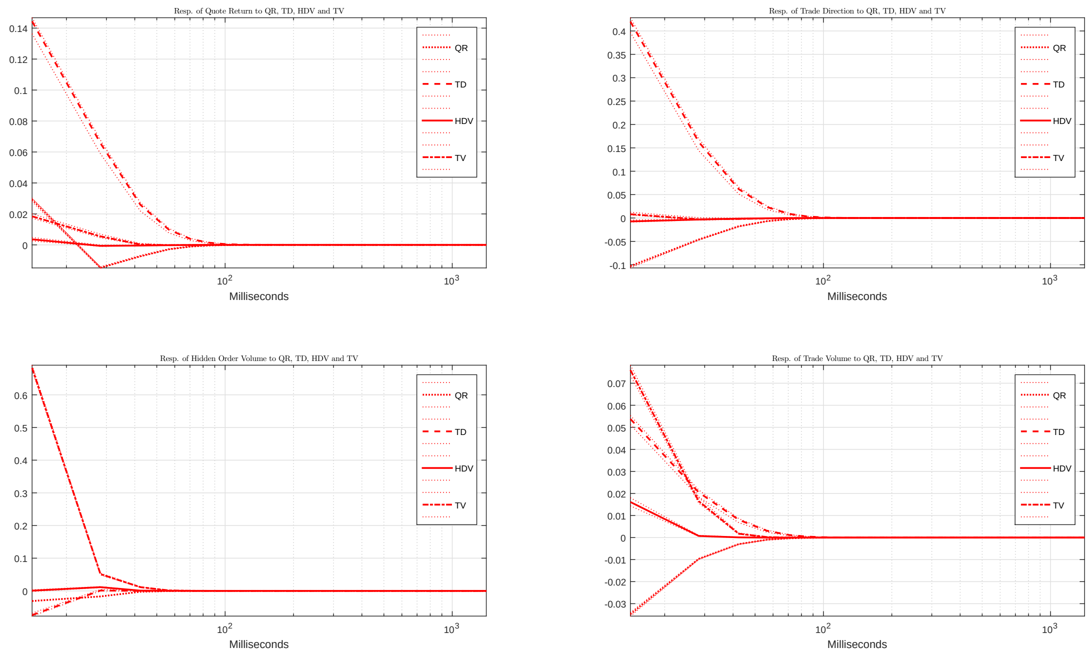



**Figure 8: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for ten week to maturity.**

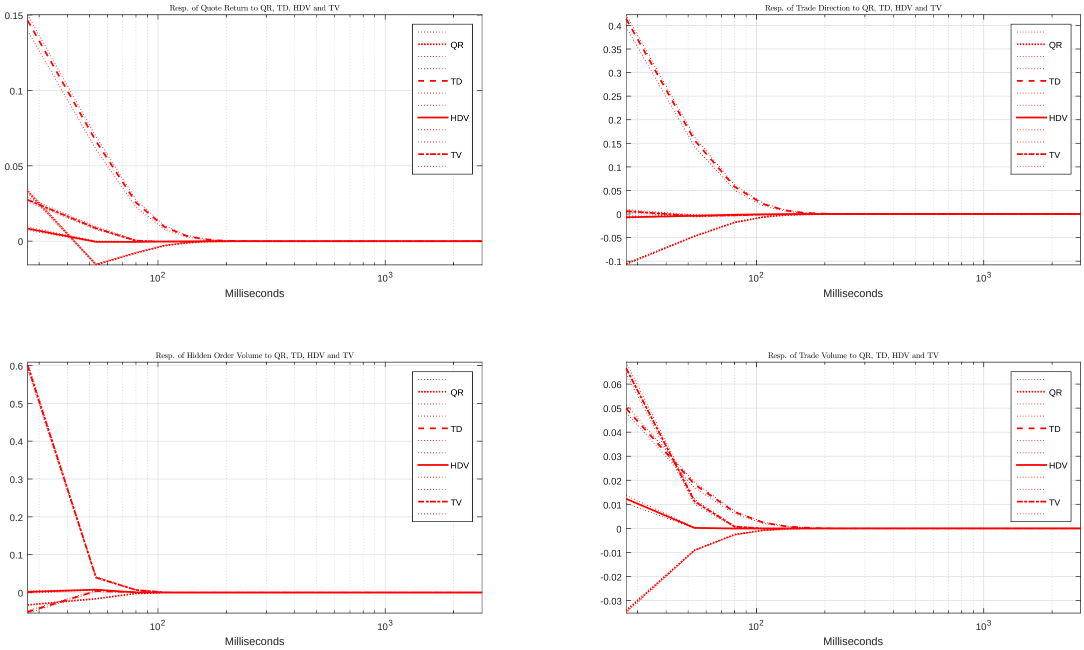



**Figure 9: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 11 weeks to maturity.**

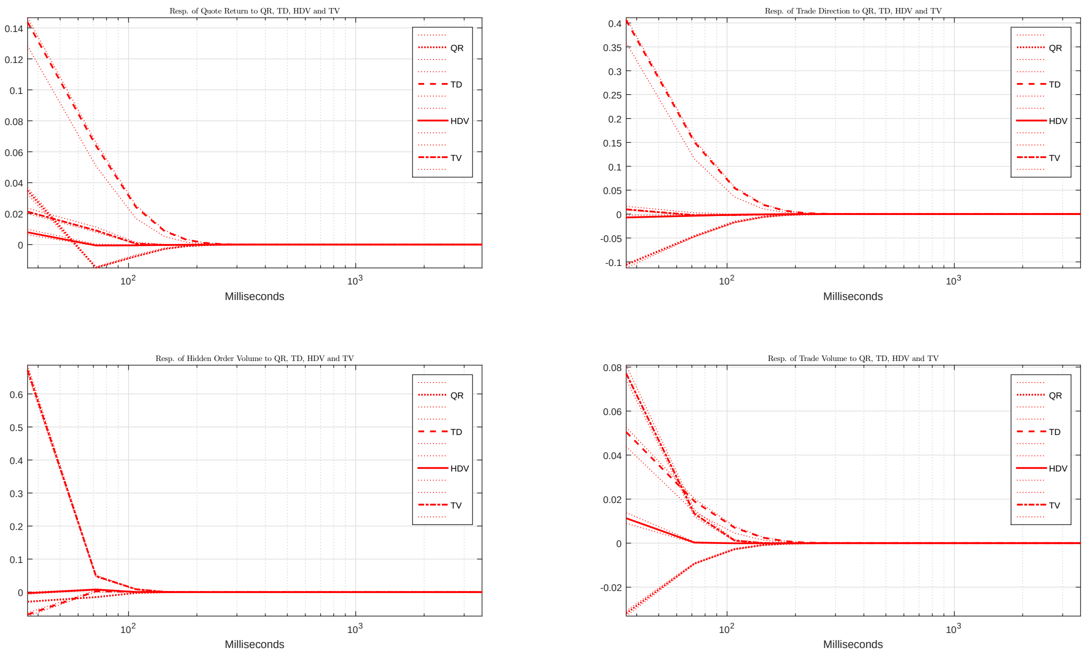



**Figure 10: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 12 weeks to maturity.**

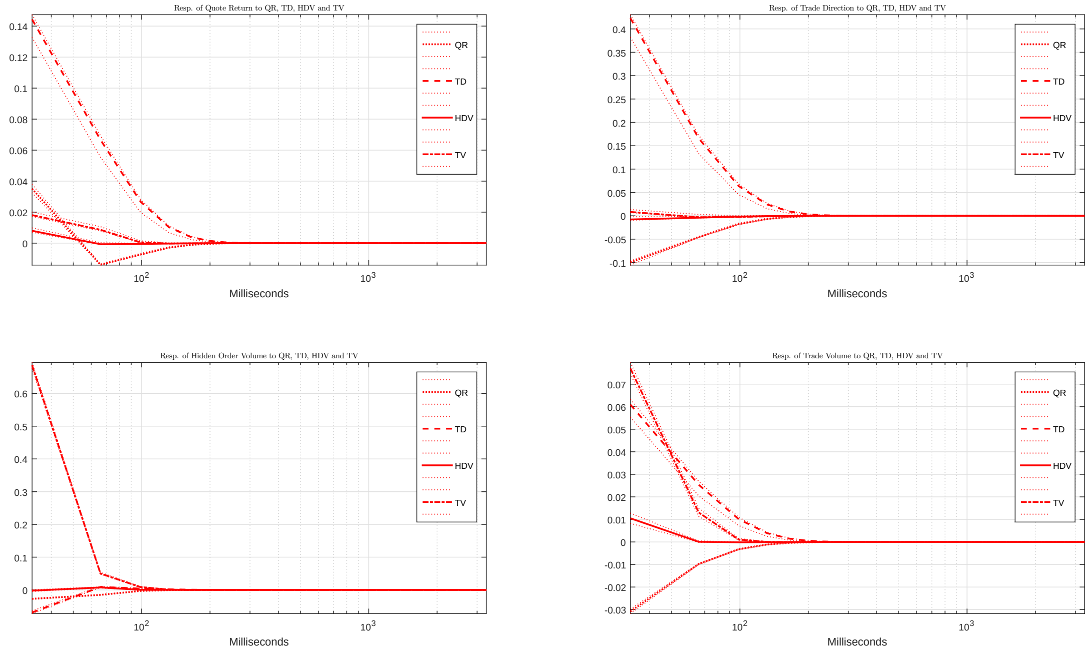



**Figure 11: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 13 weeks to maturity.**

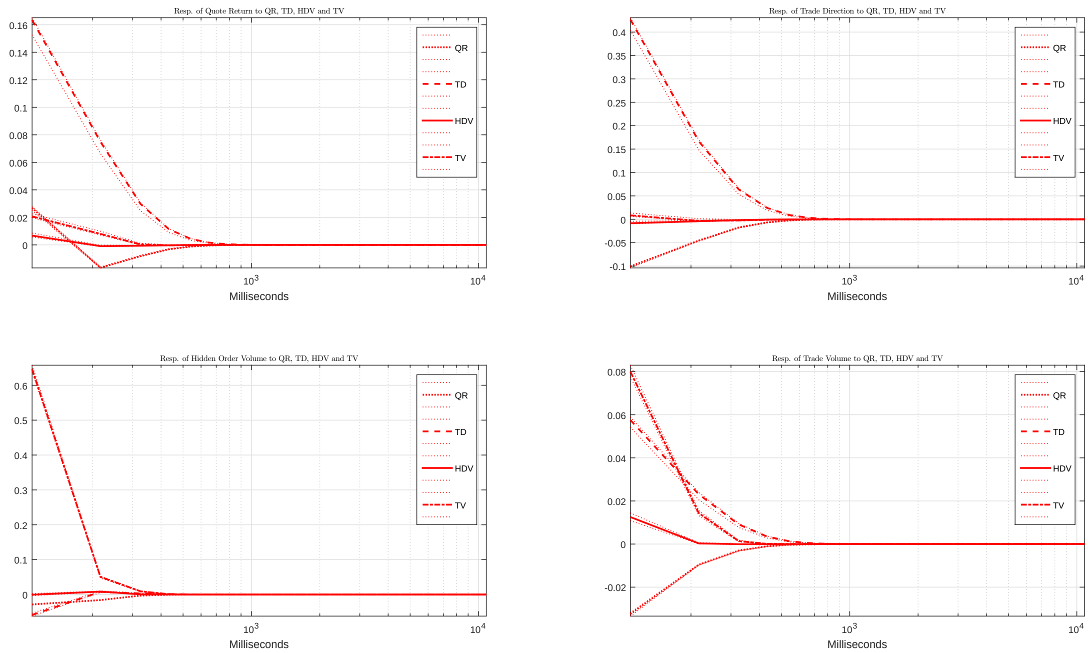



**Figure 12: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 14 weeks to maturity.**

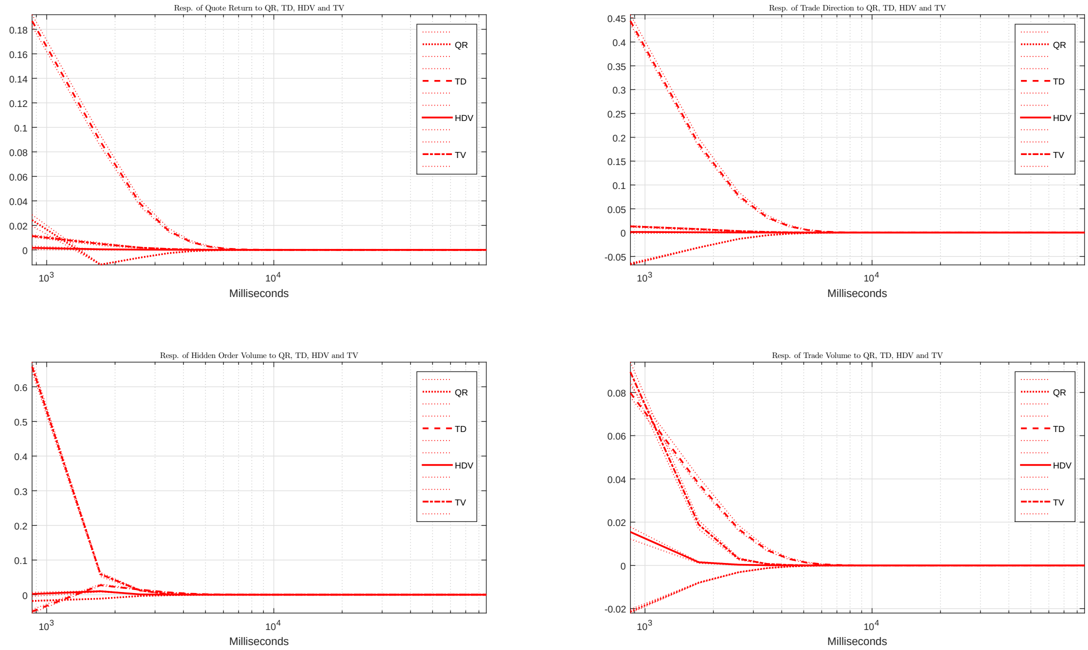



**Figure 13: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 15 weeks to maturity.**

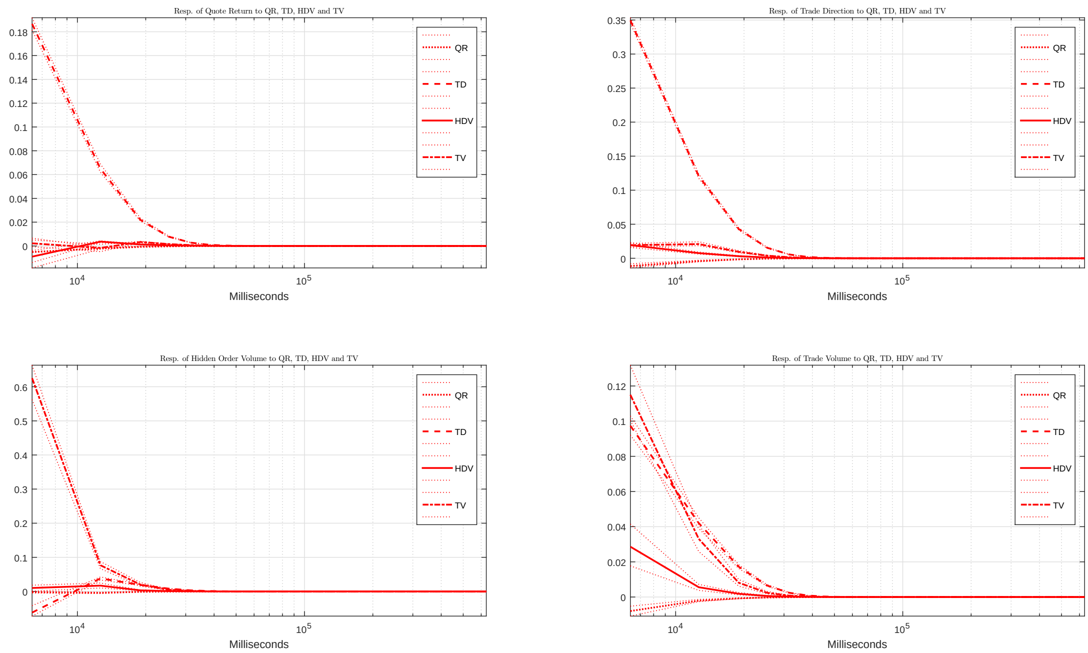



**Figure 14: Impulse response function of $\widetilde{QR}$, $\widetilde{TD}$, $\widetilde{HDV}$ and $\widetilde{TV}$ on E-mini S&P500 for 16 weeks to maturity.**

# Bibliography


Abad, D., M. Massot, and R. Pascual (2015). Evaluating vpin as a trigger for single-stock circuit breakers. *Available at SSRN 2584346*. 161, 196

Abad, D. and J. Yagüe (2012). From pin to vpin: An introduction to order flow toxicity. *The Spanish Review of Financial Economics 10*(2), 74–83. 9, 44, 56, 68, 69, 81, 160, 261

Abrantes-Metz, R. M., M. Kraten, A. D. Metz, and G. S. Seow (2012). Libor manipulation? *Journal of Banking & Finance 36*(1), 136–150. 88, 91

Admati, A. R. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica 53*(3), pp. 629–658. 36, 58, 131

Admati, A. R. and P. Pfleiderer (1986). A monopolistic market for information. *Journal of Economic Theory 39*(2), 400–438. 36

Admati, A. R. and P. Pfleiderer (1987). Viable allocations of information in financial markets. *Journal of Economic Theory 43*(1), 76–115. 36

Admati, A. R. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies 1*(1), pp. 3–40. 58, 131

Admati, A. R. and P. Pfleiderer (1990). Direct and indirect sale of information. *Econometrica: Journal of the Econometric Society*, 901–928. 36

Aitken, M. J., H. Berkman, and D. Mak (2001). The use of undisclosed limit orders on the australian stock exchange. *Journal of Banking & Finance 25*(8), 1589–1603. 201, 205, 207, 208, 214, 228

Aitken, M. J., A. Frino, A. M. Hill, and E. Jarnecic (2004). The impact of electronic trading on bid-ask spreads: Evidence from futures markets in hong kong, london, and sydney. *Journal of Futures Markets 24*(7), 675–696. 164

Almgren, R., C. Thum, E. Hauptmann, and H. Li (2005). Direct estimation of equity market impact. *Risk 18*(7), 58–62. 208

Anand, A. and D. G. Weaver (2004). Can order exposure be mandated? *Journal of Financial Markets 7*(4), 405–426. 205, 228

Andersen, T. and O. Bondarenko (2013). Assessing vpin measurement of order flow toxicity via perfect trade classification. *Available at SSRN 2292602*. 11, 133, 161

Andersen, T. G. and O. Bondarenko (2014a). Assessing measures of order flow toxicity and early warning signals for market turbulence. *Review of Finance, Forthcoming*. 11, 129, 133, 196

Andersen, T. G. and O. Bondarenko (2014b). Reflecting on the vpin dispute. *Journal of Financial Markets 17*, 53–64. 11, 129, 133

Andersen, T. G. and O. Bondarenko (2014c). Vpin and the flash crash. *Journal of Financial Markets 17*, 1–46. 11, 129, 133, 161

Ascioglu, A., S. P. Hegde, and J. B. McDermott (2008). Information asymmetry and investmentcash flow sensitivity. *Journal of Banking & Finance 32*(6), 1036 – 1048. 62

Aslan, H., D. Easley, S. Hvidkjaer, and M. O'Hara (2011). The characteristics of informed trading: Implications for asset pricing. *Journal of Empirical Finance 18*(5), 782 – 801. 9, 56, 61, 62, 81, 261

Ates, A. and G. H. Wang (2005). Information transmission in electronic versus open-outcry trading systems: An analysis of us equity index futures markets. *Journal of Futures Markets 25*(7), 679–715. 164

Ballocchi, G., M. Dacorogna, R. Gençay, and B. Piccinato (2001). Time-to-expiry seasonalities in eurofutures. *Studies in Nonlinear Dynamics & Econometrics 4*(4). 169, 170

Baruch, S. (2005). Who benefits from an open limit-order book?*. *The Journal of Business 78*(4), 1267–1306. 206

Beber, A. and C. Caglio (2005). Order submission strategies and information: Empirical evidence from the nyse. In *EFA 2003 Annual Conference Paper*, Number 875. 209

Berkowitz, S. A., D. E. Logue, and E. A. Noser (1988). The total cost of transactions on the nyse. *The Journal of Finance 43*(1), 97–112. 217

Bernaola-Galván, P., P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, and H. E. Stanley (2000). Scale invariance in the nonstationarity of physiological signals. *arXiv preprint cond-mat/0005284*. 209, 211

Bessembinder, H., M. Panayides, and K. Venkataraman (2009). Hidden liquidity: an analysis of order exposure strategies in electronic stock markets. *Journal of Financial Economics 94*(3), 361–383. 200, 201, 207, 208, 215, 236

Biais, B., P. Hillion, and C. Spatt (1995). An empirical analysis of the limit order book and the order flow in the paris bourse. *the Journal of Finance 50*(5), 1655–1689. 209

Bisière, C. and T. Kamionka (2000). Timing of orders, orders aggressiveness and the order book at the paris bourse. *Annales d'Economie et de Statistique*, 43–72. 209

Bloomfield, R. and M. O'Hara (2000). Can transparent markets survive? *Journal of financial Economics 55*(3), 425–459. 202

Bloomfield, R., M. O'Hara, and G. Saar (2015). Hidden liquidity: Some new light on dark trading. *The Journal of Finance 70*(5), 2227–2274. 200, 209

Blume, M. E. and M. A. Goldstein (1997). Quotes, order flow, and price discovery. *The Journal of Finance 52*(1), 221–244. 202

Boehmer, E., J. Grammig, and E. Theissen (2007). Estimating the probability of informed tradingdoes trade misclassification matter? *Journal of Financial Markets 10*(1), 26–47. 55, 65, 132

Boulatov, A. and T. J. George (2013). Hidden and displayed liquidity in securities markets with informed liquidity providers. *Review of Financial Studies 26*(8), 2096–2137. 206

Breon-Drish, B. (2015). On existence and uniqueness of equilibrium in a class of noisy rational expectations models. *The Review of Economic Studies*, rdv012. 36, 37

Brockman, P. and X. S. Yan (2009). Block ownership and firm-specific information. *Journal of Banking & Finance 33*(2), 308 – 316. 63

Brown, P., N. Thomson, and D. Walsh (1999). Characteristics of the order flow through an electronic open limit order book. *Journal of International Financial Markets, Institutions and Money 9*(4), 335 – 357. 60

Buti, S. and B. Rindi (2008). Hidden orders and optimal submission strategies in a dynamic limit order market. *SSRN eLibrary*. 205, 258, 259

Buti, S. and B. Rindi (2013). Undisclosed orders and optimal submission strategies in a limit order market. *Journal of Financial Economics 109*(3), 797–812. 200, 201, 207, 258

CFTC (2012). Cftc docket no. 12-25: Order institution proceedings pursuant to sections 6(c) and 6(d) of the commodity exchange act, as amended, making findings and imposing remedial sanctions: In the matter of barclays plc, barclays bank plc and barclays capital inc. Technical report, Commodity Futures Trading Commision. 97, 103, 104, 106

CFTC (2013). Cftc docket no. 12-25: Order institution proceedings pursuant to sections 6(c) and 6(d) of the commodity exchange act, as amended, making findings and imposing remedial sanctions: In the matter of coöperative central raiffeisen-boerenleenbank b.a. Technical report, Commodity Futures Trading Commission. 118

Chakrabarty, B., P. K. Jain, A. Shkilko, and K. Sokolov (2014). Speed of market access and market quality: Evidence from the sec naked access ban. *Western Finance Association*. 220

Chakrabarty, B., P. C. Moulton, and A. Shkilko (2012). Short sales, long sales, and the lee–ready trade classification algorithm revisited. *Journal of Financial Markets 15*(4), 467–491. 67

Chakrabarty, B., R. Pascual, and A. Shkilko (2012). Trade classification algorithms: A horse race between the bulk-based and the tick-based rules. *Available at SSRN 2182819*. 65

Chen, Q., I. Goldstein, and W. Jiang (2007). Price informativeness and investment sensitivity to stock price. *Review of Financial Studies 20*(3), 619–650. 63

Chen, Y. and H. Zhao (2012). Informed trading, information uncertainty, and price momentum. *Journal of Banking & Finance 36*(7), 2095 – 2109. 63

Chordia, T., R. Roll, and A. Subrahmanyam (2008). Liquidity and market efficiency. *Journal of Financial Economics 87*(2), 249–268. 220

Chung, K. H., M. Li, and T. H. McInish (2005). Information-based trading, price impact of trades, and trade autocorrelation. *Journal of banking & finance 29*(7), 1645–1669. 12, 224, 225, 237

CME (2005). *An Introduction to Futures And Options*. CME Market Education Department in Chicago. 10, 16, 18, 19

Comerton-Forde, C. and K. M. Tang (2009). Anonymity, liquidity and fragmentation. *Journal of Financial Markets 12*(3), 337–367. 237

Copeland, T. E. and D. Galai (1983). Information effects on the bid-ask spread. *the Journal of Finance 38*(5), 1457–1469. 217

De Prado, M. M. L. (2011). *Advances in High Frequency Strategies*. Complutense University. 14, 42, 43

De Winne, R. and C. D'hondt (2007a). Hide-and-seek in the market: placing and detecting hidden orders. *Review of Finance 11*(4), 663–692. 200, 201, 205, 207, 209, 215, 228

De Winne, R. and C. D'hondt (2007b). Hide-and-seek in the market: placing and detecting hidden orders. *Review of Finance 11*(4), 663–692. 200

Dufour, A. and R. F. Engle (2000). Time and the price impact of a trade. *Journal of Finance*, 2467–2498. 3, 12, 224, 225, 237

Easley, D., M. L. De Prado, and M. OHara (2011). The microstructure of the flash crash: Flow toxicity, liquidity crashes and the probability of informed trading. *Journal of Portfolio Management 37*(2), 118–128. 69, 149, 196, 262

Easley, D., M. M. L. de Prado, and M. O'Hara (2012). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies 25*(5), 1457–1493. 6, 10, 52, 69, 132, 159, 160, 196, 197, 262

Easley, D., M. M. L. de Prado, and M. O'Hara (2014). Vpin and the flash crash: A rejoinder. *Journal of Financial Markets 17*(0), 47 – 52. 11, 133, 161

Easley, D., R. F. Engle, M. O'Hara, and L. Wu (2008). Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Econometrics 6*(2), 171–207. 3, 52

Easley, D., S. Hvidkjaer, and M. OHara (2002). Is information risk a determinant of asset returns? *The Journal of Finance 57*(5), 2185–2221. 61

Easley, D., N. M. Kiefer, and M. O'Hara (1997a). The information content of the trading process. *Journal of Empirical Finance 4*(23), 159 – 186. High Frequency Data in Finance, Part 1. 60

Easley, D., N. M. Kiefer, and M. O'Hara (1997b). One day in the life of a very common stock. *The Review of Financial Studies 10*(3), pp. 805–835. 8, 60

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *The Journal of Finance 51*(4), pp. 1405–1436. 6, 10, 1, 2, 8, 9, 44, 45, 46, 47, 51, 55, 56, 59, 60, 61, 63, 64, 65, 68, 69, 75, 81, 84, 85, 124, 130, 131, 137, 139, 142, 159, 160, 196, 261

Easley, D., M. Lopez de Prado, and M. OHara (2010). Measuring flow toxicity in a high frequency world. Technical report, unpublished Cornell University working paper (October). 3, 10, 52, 54

Easley, D., M. López de Prado, and M. O'Hara (2012). Bulk classification of trading activity. *SSRN, March*. 65, 66

Easley, D. and M. O'Hara (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics 19*(1), 69 – 90. 8, 58, 59

Easley, D. and M. O'Hara (1992). Time and the process of security price adjustment. *The Journal of finance 47*(2), 577–605. 1, 44, 59, 68, 124, 131, 196

Easley, D., M. O'hara, and G. Saar (2001). How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative Analysis 36*(01), 25–51. 61

Ellis, K., R. Michaely, and M. O'Hara (2000). The accuracy of trade classification rules: evidence from nasdaq. *Journal of Financial and Quantitative Analysis 35*(04), 529–551. 65, 67

Esser, A. and B. Mönch (2007). The navigation of an iceberg: The optimal use of hidden orders. *Finance Research Letters 4*(2), 68–81. 207

FCA (2012). Journey to the fca, the financial conduct authority. Technical report, FCA. 10, 35

Finucane, T. J. (2000). A direct test of methods for inferring trade direction from intra-day data. *Journal of Financial and Quantitative Analysis 35*(04), 553–576. 67

Frey, S. and P. Sandås (2009). The impact of iceberg orders in limit order books. In *AFA 2009 San Francisco Meetings Paper*. 200, 201, 205, 207, 209, 214, 228

FSA (2012). Fsa enforcement and financial crime division: Final notice to barckays bank plc (27 june 2012). Technical report, Financial Services Authority. 101, 106

Gao, F., F. Song, and J. Wang (2013). Rational expectations equilibrium with uncertain proportion of informed traders. *Journal of Financial Markets 16*(3), 387–413. 36, 37, 38

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71 – 100. 6, 8, 42, 58, 59, 60, 130, 131, 132, 134, 136

Gozluklu, A. E. et al. (2009). Pre-trade transparency and informed trading: An experimental approach to hidden liquidity. In *ESA European Meeting*. Citeseer. 209

Grammig, J., D. Schiereck, and E. Theissen (2001). Knowing me, knowing you:: Trader anonymity and informed trading in parallel markets. *Journal of Financial Markets 4*(4), 385 – 412. 61

Griffiths, M. D., B. F. Smith, D. A. S. Turnbull, and R. W. White (2000). The costs and determinants of order aggressiveness. *Journal of Financial Economics 56*(1), 65–88. 209

Grossman, S. J. (1988). An analysis of the implications for stock and futures price volatility of program trading and dynamic hedging strategies. *The Journal of Business 61*(3), pp. 275–298. 37

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American Economic Review 70*(3), pp. 393–408. 36, 57, 131

Hamilton, J. D. (1994). *Time series analysis*, Volume 2. 223

Harris, L. (1996). *Does a large minimum price variation encourage order exposure?*, Volume 96. New York Stock Exchange. 205, 207, 208, 228

Harris, L. E. and V. Panchapagesan (2005). The information content of the limit order book: evidence from nyse specialist trading decisions. *Journal of Financial Markets 8*(1), 25–67. 217

Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance 46*(1), 179–207. 3, 12, 60, 131, 224, 225, 237

Hasbrouck, J. (2003). Intraday price formation in us equity index markets. *The Journal of Finance 58*(6), 2375–2400. 27

Hasbrouck, J. and G. Saar (2013). Low-latency trading. *Journal of Financial Markets 16*(4), 646–679. 3, 4, 12, 215, 220

Hautsch, N. and R. Huang (2012a). The market impact of a limit order. *Journal of Economic Dynamics and Control 36*(4), 501–522. 207

Hautsch, N. and R. Huang (2012b). On the dark side of the market: identifying and analyzing hidden order placements. *Available at SSRN 2004231*. 200, 201, 207, 208

Heidle, H. G. and R. D. Huang (2002). Information-based trading in dealer and auction markets: An analysis of exchange listings. *The Journal of Financial and Quantitative Analysis 37*(3), pp. 391–424. 61

Hellerstein, J. M. (2008). Quantitative data cleaning for large databases. *United Nations Economic Commission for Europe (UNECE)*. 25

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory 22*(3), 477 – 498. 36, 57, 131

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *The Journal of Finance 66*(1), 1–33. 219

Hull, J. C. (2012). *Options, futures, and other derivatives*. Eighth Edition, Global Edition. Pearson Education. 33

Idier, J. and S. Nardelli (2011). Probability of informed trading on the euro overnight market rate. *International Journal of Finance & Economics 16*(2), 131–145. 63, 69, 160

Jiang, J. (2015). *Volume-Synchronized Probability of Informed Trading (VPIN), Market Volatility, and High-Frequency Liquidity*. Ph. D. thesis, School of Business, Brock University. 222

Kang, M. (2010). Probability of information-based trading and the january effect. *Journal of Banking & Finance 34*(12), 2985 – 2994. International Financial Integration. 63

Karyampas, D. and P. Paiardini (2011). Probability of informed trading and volatility for an etf. 51, 142

Ke, W.-C. (2014). The sensitivity to trade classification algorithms for estimating the probability of informed trading. *International Journal of Trade, Economics and Finance 5*(5). 55, 132

Kim, C. W., T. T. Perry, and M. Dhatt (2014). Informed trading and price discovery around the clock. *The Journal of Alternative Investments 17*(2), 68–81. 9, 56, 69, 70, 81, 135, 159, 160, 261

Kirilenko, A. A., A. S. Kyle, M. Samadi, and T. Tuzun (2015). The flash crash: The impact of high frequency trading on an electronic market. *Available at SSRN 1686004*. 26

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335. 6, 8, 39, 58, 60, 131

Labuszewski, J. W. (2011). INTEREST RATES understanding eurodollar futures. `http://www.cmegroup.com/trading/interest-rates/files/understanding-eurodollar-futures.pdf`. 98

Lee, C. and B. Radhakrishna (2000). Inferring investor behavior: Evidence from torq data. *Journal of Financial Markets 3*(2), 83–111. 67

Lee, C. and M. J. Ready (1991). Inferring trade direction from intraday data. *The Journal of Finance 46*(2), 733–746. 55, 65, 67, 132

Leland, H. E. (1992). Insider trading: Should it be prohibited? *Journal of Political Economy*, 859–887. 36

Leung, M. T., H. Daouk, and A.-S. Chen (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting 16*(2), 173–190. 222

Li, H. and X. Ye (2013). A dynamic, volume-weighted average price approach based on the fast fourier transform algorithm. *Asia-Pacific Journal of Financial Studies 42*(6), 969–991. 217

Lütkepohl, H. and H.-E. Reimers (1992). Impulse response analysis of cointegrated systems. *Journal of economic dynamics and control 16*(1), 53–78. 227

Madhavan, A. (2000). Market microstructure: A survey. *Journal of financial markets 3*(3), 205–258. 1, 202

Madhavan, A., D. Porter, and D. Weaver (2005). Should securities markets be transparent? *Journal of Financial Markets 8*(3), 265–287. 206

Madhavan, A. N. (2002). Vwap strategies. *The Journal of Trading 2002*(1), 32–39. 217

Moinas, S. (2010). Hidden limit orders and liquidity in order driven markets. *Toulouse School of Economics Working Paper*. 207

Moro, E., J. Vicente, L. G. Moyano, A. Gerig, J. D. Farmer, G. Vaglica, F. Lillo, and R. N. Mantegna (2009). Market impact and trading profile of hidden orders in stock markets. *Physical Review E 80*(6), 066102. 208, 211

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets 3*(3), 259–286. 65, 67

O'Hara, M. (1995). *Market microstructure theory*, Volume 108. Blackwell Cambridge. 1

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics 116*(2), 257–270. 4

Pardo, A. and R. Pascual (2012). On the hidden side of liquidity. *The European Journal of Finance 18*(10), 949–967. 201, 208, 209, 211, 214, 217

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies 11*(4), 789–816. 205, 259

Phylaktis, K. (1999). Capital market integration in the pacific basin region: an impulse response analysis. *Journal of International Money and Finance 18*(2), 267–287. 227

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets 7*(1), 53–74. 209

Reiss, P. C. and I. M. Werner (2005). Anonymity, adverse selection, and the sorting of interdealer trades. *Review of Financial Studies 18*(2), 599–636. 229, 231

Shah, S. and B. W. Brorsen (2011). Electronic vs. open outcry: Side-by-side trading of kcbt wheat futures. *Journal of Agricultural and Resource Economics 36*(1). 164

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48. 222

Sturm, F. and P. Barker (2011). Eurodollar futures: The basics. 98

Torre, N. (1997). Barra market impact model handbook. *BARRA Inc., Berkeley*. 208

Tse, Y. and T. V. Zabotina (2001). Transaction costs and market quality: Open outcry versus electronic trading. *Journal of Futures Markets 21*(8), 713–735. 164

Tuttle, L. A. (2003). Hidden orders, trading costs and information. *Available at SSRN: http://ssrn.com/abstract=676019*. 208

Vaglica, G., F. Lillo, E. Moro, and R. N. Mantegna (2008). Scaling laws of strategic behavior and size heterogeneity in agent dynamics. *Physical Review E 77*(3), 036110. 208, 211

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics 82*(1), 103 – 133. 62

Veredas, D. and R. Pascual (2004). What pieces of limit order book information are informative?: An empirical analysis of a pure order driven market. *CORE Discussion Paper, 2004/33*. 209

Verrecchia, R. E. (1982). Information acquisition in a noisy rational expectations economy. *Econometrica: Journal of the Econometric Society*, 1415–1430. 36

Viljoen, T., P. J. Westerholm, and H. Zheng (2014). Algorithmic trading, liquidity, and price discovery: An intraday analysis of the spi 200 futures. *Financial Review 49*(2), 245–270. 12, 224

Wei, W. C. (2013). *Essays on Information Asymmetry and Price Impact in Market Microstructure*. Ph. D. thesis. 2

Yan, Y. and S. Zhang (2012). An improved estimation method and empirical properties of the probability of informed trading. *Journal of Banking & Finance 36*(2), 454–467. 132

Yan, Y. and S. Zhang (2014). Quality of pin estimates and the pin-return relationship. *Journal of Banking & Finance 43*, 137–149. 69, 160

Zebedee, A. A. (2001). The impact of a trade on national best bid and offer quotes: a new approach to modeling irregularly spaced data. *Journal of Multinational Financial Management 11*(4), 363–383. 12, 224